

How do Humans Determine Reflectance Properties under Unknown Illumination?*

Roland W. Fleming, Ron O. Dror, and Edward H. Adelson
Massachusetts Institute of Technology
roland@psyche.mit.edu, rondror@ai.mit.edu, adelson@psyche.mit.edu

Abstract

Under normal viewing conditions, humans find it easy to distinguish between objects made out of different materials such as plastic, metal, or paper. Untextured materials such as these have different surface reflectance properties, including lightness and gloss. With single isolated images and unknown illumination conditions, the task of estimating surface reflectance is highly underconstrained, because many combinations of reflection and illumination are consistent with a given image. In order to work out how humans estimate surface reflectance properties, we asked subjects to match the appearance of isolated spheres taken out of their original contexts. We found that subjects were able to perform the task accurately and reliably without contextual information to specify the illumination. The spheres were rendered under a variety of artificial illuminations, such as a single point light source, and a number of photographically-captured real-world illuminations from both indoor and outdoor scenes. Subjects performed more accurately for stimuli viewed under real-world patterns of illumination than under artificial illuminations, suggesting that subjects use stored assumptions about the regularities of real-world illuminations to solve the ill-posed problem.

1. Introduction

Humans often take for granted their ability to recognize materials such as metal, plastic and paper under a wide range of viewing conditions. Humans succeed at this task despite the fact that different combinations of illumination and surface reflectance can produce identical images. For example, a chrome sphere can be made to look like any other sphere with just the right illumination.

Given that the task is formally underconstrained, how do subjects recognize surface reflectance properties across dif-

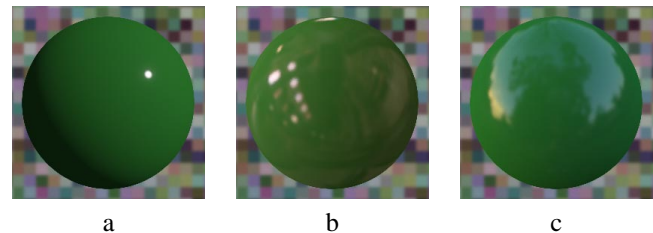


Figure 1. The sphere in (a) was rendered under point source illumination. The spheres in (b) and (c) were rendered under real-world illumination. All three spheres have the same reflectance properties and yet the impression of the material quality is clearer in (b) and (c) than in (a).

ferent illuminations? Under ordinary viewing conditions, an observer can draw on multiple sources of information. Potentially useful cues include motion, binocular stereopsis, knowledge of illumination conditions, and familiarity with the object. In this paper, we discuss how well subjects can judge reflectance properties in the absence of such cues, when forced to rely on stored assumptions about illumination. In order to do this we measure their ability to estimate reflectance properties from isolated images of single spheres viewed out of their original contexts. We find that subjects can perform the task well, as long as the illumination conditions are characteristic of real-world scenes.

Researchers in computer vision and graphics often assume that point source illumination simplifies the process of reflectance estimation. Figure 1 shows synthetic images of three identical spheres under different illuminations. Sphere (a) was rendered under point source illumination, while spheres (b) and (c) were rendered under photographically-captured real-world illumination [3]. The impression of the material quality is clearer in (b) and (c) than in (a). We show that humans estimate reflectance more reliably under complex realistic illumination than under simple synthetic illumination.

*A colour version of this paper may be downloaded from http://www.ai.mit.edu/people/rondror/papers/fleming_cvpr_01.pdf.

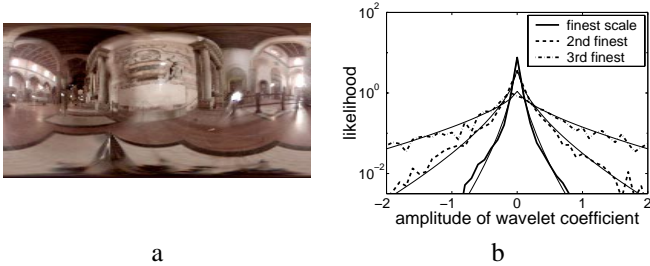


Figure 2. The image in (a) is a 2D panoramic projection of a spherical real-world illumination map. The graph in (b) shows the distribution of wavelet coefficients at several scales for this illumination map, which are characteristic of real-world images (for details, see [4] and [5]).

Why might real-world illumination facilitate reflectance estimation? In the real world, light is typically incident on a surface from nearly every direction, in the form of direct illumination from luminous sources or indirect illumination reflected from other surfaces. The illumination at a given point can be described by the spherical image acquired by a camera that looks in every direction from that point. Such images or “illumination maps” have been captured photographically and used to render objects for the purposes of realistic computer graphics. Recent work has shown that the spatial structure of such real-world illumination maps exhibits statistical regularity similar to that of natural images [5], see Figure 2. A computer system can take advantage of this regularity in recognizing surface reflectances [4]. We suggest that humans might also exploit stored assumptions about such regularities to solve this otherwise underconstrained problem. Under illuminations for which these assumptions do not hold, we expect systematic biases in estimates of surface reflectance. For example, a sphere in a low-contrast environment should look less glossy than it would under more typical conditions.

2. Methods

In order to measure the accuracy of human surface reflectance estimation we asked subjects to perform a reflectance matching task. Subjects were presented with two images of spheres that had been computer rendered under different illuminations. Their task was to adjust two parameters of the surface reflectance of one sphere (the “Match”) until it appeared to be made of the same material as the other sphere (the “Test”), despite the difference in illumination. Example stimuli are shown in Figure 3. Nishida and Shinya [8] used a similar experimental design to study the effect of surface geometry on apparent reflectance properties.

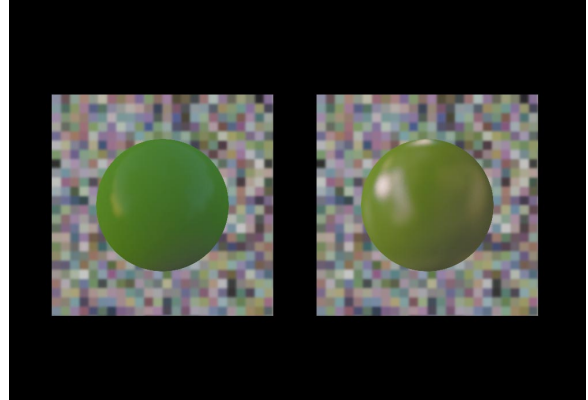


Figure 3. An example screen from the matching experiment. Subjects adjusted two parameters of the sphere on the right until it appeared to be made of the same material as the sphere on the left. Note that at the moment the spheres have different reflectance properties.

Observers. Four subjects with normal or corrected-to-normal vision participated in the experiments. One was an author (RF), two were experienced subjects who were naive to the purpose of the study (JM and MS) and one was a novice observer (RA) who was paid for participating.

Stimuli. Each Test or Match image contained a sphere that was computer rendered using the Radiance software package [12]. As the screen was viewed from approximately 1m, the spheres subtended a visual angle of roughly 5° . The spheres were viewed against a standard coloured random-check background, which can be downloaded from [1]. This background was used to avoid the Gelb effect, i.e. the tendency for an isolated surface to appear lighter in colour than it really is. Blur was added to separate the sphere from the background in depth. Because the same background was used for all stimuli, it provided no systematic information about the illumination.

The reflectances of the spheres were specified by the isotropic Ward model [11], a physically realizable variant of the Phong shading model. This model approximates the reflectance properties of a wide range of physical surfaces with a few free parameters by representing surface reflectance as the sum of two components: a diffuse (“matte” or “Lambertian”) component and a specular (“glossy”) component. To make the step-sizes perceptually equal, the reflectance parameters were reparameterized (i.e. the scales were stretched), according to the psychophysically uniform space proposed by Pellacini *et al.* [9]. As the perception of matte lightness has been studied extensively before [2, 6, 7], we held the diffuse reflectance fixed for all

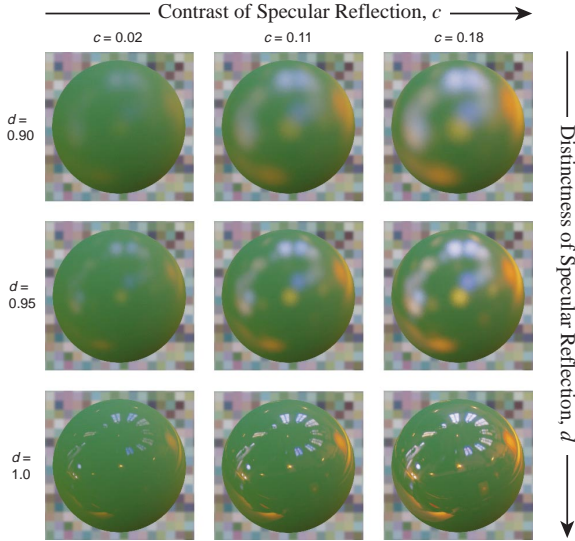


Figure 4. Grid showing range of reflectance properties used in the experiments for a particular real-world illumination map.

conditions. Subjects simultaneously adjusted two parameters controlling the specular component: the contrast of the specular reflection, c , and the sharpness of the specular component, d .¹ Increases in c lead to perceptually uniform increases in the glossiness of the surface; increases in d alter the surface from rough—like brushed aluminium—to smooth—like polished chrome. Ten values were used for c and eleven for d , making a total of 110 possible surface reflectances. These values spanned a range greater than but including the range of reflectances that isotropic dielectric materials such as gloss paint and sand-blasted plastic exhibit in the real world. Figure 4 shows example stimuli demonstrating the range of surface reflectances used in the experiment. The spheres used in all conditions were spatially uniform in surface reflectance (i.e. untextured).

Two classes of illumination were used to render the Test spheres: real-world and artificial. The real-world Test illuminations consisted of eight illumination maps from both indoor and outdoor scenes, which were captured photographically by a process described below. The artificially generated Test illuminations consisted of: a single point source; multiple point sources; a single extended rectangular source; Gaussian white noise; and Gaussian noise with a $1/f$ amplitude spectrum (pink noise).² The Match sphere

¹The original Ward model uses two parameters to describe specular reflection—the amount of specular reflection ρ_s , and the surface roughness α . Parameter c increases monotonically with ρ_s , while $d = 1 - \alpha$.

²We generated the white noise illumination map by summing spherical harmonics whose coefficients up to a fixed order were chosen from independent Gaussian distributions of equal variance. For the pink noise, the

that the subject adjusted was viewed under the same real-world illumination for all conditions. This allowed us to examine systematic effects of illumination upon matching performance. Examples of a sphere rendered under each of the illuminations used in the experiment are shown in Figure 5; the spheres all have the same surface reflectance. It should be noted that the illuminations contain finer detail than is visible in these spheres; this detail was visible in spheres with sharper (i.e. less blurred) specular reflections.

The real-world illuminations were based on high-dynamic range light probe images acquired photographically by Debevec *et al.* [3]. This was done by compositing wide-angle photographs at different exposures to create a spherical image that maps the illumination at a point. The overall brightness of each illumination was scaled such that a standard Lambertian patch oriented perpendicular to the observer would yield the same luminance under all illuminations. The artificial illuminations were created as spherical illumination maps and used to render spheres in the same way as the real-world illuminations.

There are in principle two basic ways that subjects can perform the task, and these depend on different features of the subject’s experience. One approach is to attend to the global, “high-level” percept of surface reflectance that emerges as a consequence of unconscious processing of the entire image. The second is to attend to “low-level” image features, such as local variations in contrast and blur, which correlate with the parameters of surface reflectance. Both strategies require subjects to make assumptions about the equivalence of features in the Test and Match images, and thus the task is ill posed irrespective of the strategy.

In order to determine how subjects perform the task, a control condition was used to complement the two main classes of illumination. The control stimuli were photographic negatives of four of the real-world illumination conditions. Specifically, only the luminances of the images were inverted; the hues were maintained. The process of inverting the image values preserves much of the spatial structure of the image, along with its odd-order image statistics. However, the percept of a uniform, coherent surface reflectance is interrupted, as can be seen in the example stimulus shown in Figure 6. Subjects adjusted a photographic negative of the standard probe. If subjects performed the task by attending to the “high-level” features of their experience, then performance with the negative controls should be compromised by the elimination of this source of information. If performance is the same for both positive and negative versions of the displays, then it is possible that subjects perceptually equate “low-level” properties of the Test

spherical harmonic coefficients were chosen independently from Gaussian distributions with standard deviation inversely proportional to the spherical harmonic order, which is analogous to frequency. This process produces noise whose power spectrum is similar to that of many real-world illuminations and natural images [5].

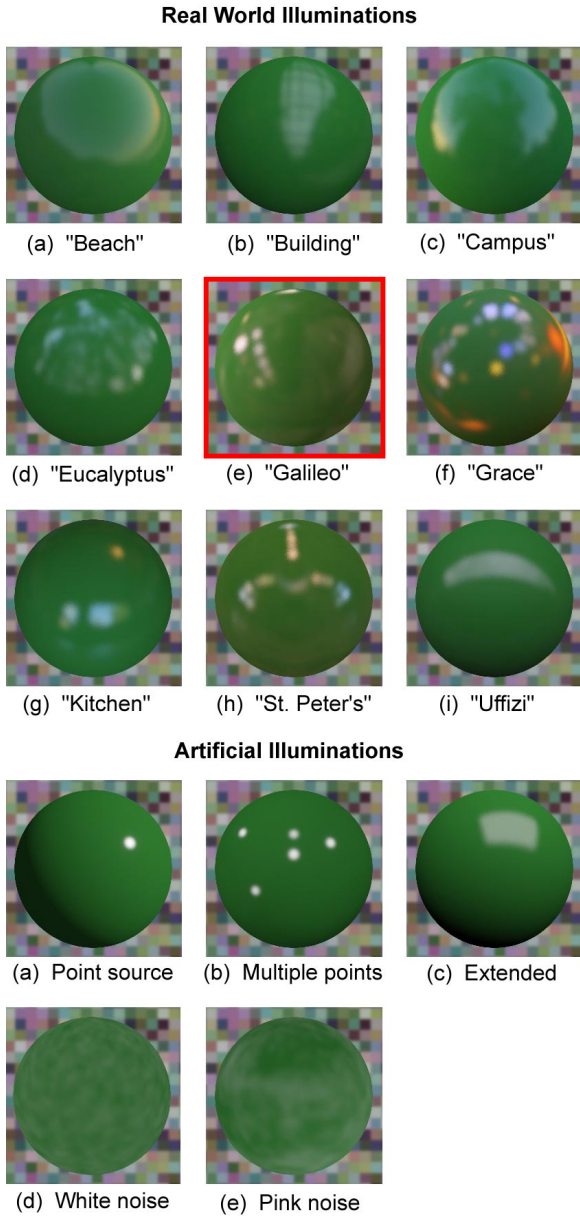


Figure 5. Spheres rendered under each of the illuminations used in the experiments. All spheres have the same surface reflectance. The fifth real-world illumination (highlighted with a perimeter) was the standard Match illumination.

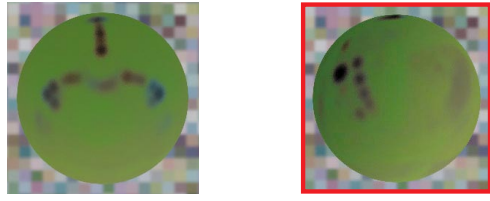


Figure 6. Photographic negatives of spheres rendered under the real-world illuminations that were used in the main experiment. The image with a perimeter, a polarity inverted version of the standard Match illumination, was used as the standard probe in the control condition.

and Match stimuli that correlate with surface reflectance.

Dealing with the limitations of the monitor. The dynamic range of a good monitor is several orders of magnitude smaller than the full range of luminances that result from viewing a specular surface under typical conditions. We used a number of presentation devices in an attempt to overcome this.

First, all images were presented in a black room with the lights off, to decrease the luminance of the darkest blacks in the image. We estimated that as a consequence of this we were able to achieve a high spatial frequency dynamic range of about 30.

Second, rather than allowing the image values to clip, the images were passed through a compressive non-linearity of the type described by Tumblin *et al.* [10]. This is a sigmoidal non-linearity that is linear for intermediate luminances but compresses low and high values. The same tone-mapping function was used for every experimental condition. The monitor was calibrated to ensure linearity before every session of the experiment.

Third, we applied synthetic glare to the rendered images in order to mimic the optical effects of viewing high luminances with the human eye. This was done according to specifications derived by Ward *et al.* [13] from empirical measurements of the optical properties of the eye. This process simulates the glare that would be experienced had the brightest points in the images really been shown at their correct luminances. The process has little effect except for bright point sources.

Procedure. Each illumination condition was run in a separate block and the order of the blocks was randomized across subjects. Thus each subject ran a total of 17 blocks (8 real-world + 5 artificial + 4 control); these were run at different times. Within a block, subjects made 110 observations: one for each of the possible reflectances of the Test sphere. Thus for a given value of c , the subject would per-

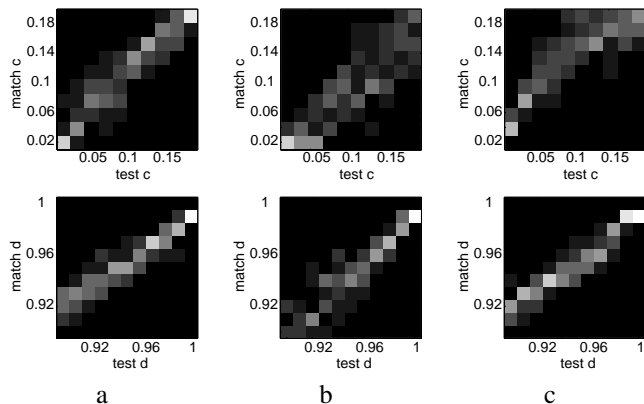


Figure 7. Match values plotted as a function of Test values for individual subjects. Graphs in the top row are matches for specular contrast, c ; graphs in the bottom row are for specular distinctness, d . The graphs in (a) are subject RF’s matches for spheres under the “St. Peter’s” illumination; (b) shows RA’s matches for spheres under the “Eucalyptus” illumination; (c) shows subject MS’s matches for spheres under the “Grace” illumination.

form 11 matches, although each of these would have a different value of d . Conversely, for each given value of d , the subject would perform 10 matches, although each would have a different values of c . The reflectances within a block were shown in random order.

Subjects could adjust both parameters simultaneously using the keyboard and pressing the mouse button to move onto the next trial when satisfied with the match. Subjects were informed by a beep if they tried to exceed the range of Match reflectances.

3. Results

3.1. Can subjects match surface reflectance under unknown illumination?

Figure 7 shows example data from the matching experiments. Subjects’ responses for specular contrast, c and specular distinctness, d are plotted as two independent graphs; the top row represents matches for c , the bottom row for d . Each pair of plots shows one subject’s matches for a single real-world illumination. The x -axis represents the true value of the Test reflectance property (c or d) and the y -axis is the subject’s match. Thus Figure 7a, for example, shows matching performance for the c parameter (top) and d parameter (bottom) for subject RF viewing a sphere that was illuminated as if it had been placed inside St. Peter’s cathedral in the Vatican (the illumination depicted in Fig-

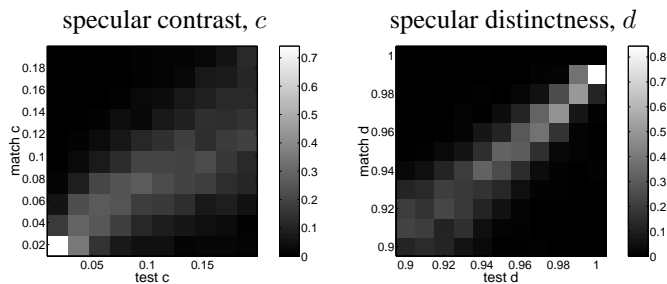


Figure 8. Data pooled across all subjects and across all real-world illuminations for parameters c and d . Grey value represents density of responses.

ure 4h). The grey value represents the density of responses for a given Test value. Thus, if subjects always responded with the same Match value to a given Test value, the corresponding sample is white; the rarer the response, the darker the grey.

The complete data set is summarized in Figure 8. These are the pooled results from all subjects across all real-world illuminations. These data show that, in fact, subjects can estimate surface reflectance in the absence of contextual information to specify illumination. This is important given that the task is formally ill posed. If subjects found the task impossible, the Match values would not depend on the Test values, i.e. the slope of the graphs would be zero. In fact, subjects’ matches are closer to veridical than to chance performance.

3.2. How accurate are subjects’ matches?

In order to quantify precisely how well subjects performed, we calculated the root mean square (r.m.s.) error between their Match values and the true values of the Test stimulus. This measure of accuracy can be expressed as a percentage of the total range of c and d that we used: the worse the performance, the larger the percentage.

For specular contrast, c , the r.m.s. error for data pooled across subjects and across all real-world illuminations was 28% of the total range of values that we used. For parameter specular distinctness, d , the r.m.s. error across subjects and across illuminations was 16% of the total range of values that we used.

3.3. Are the parameters perceptually independent?

In Figures 7 and 8 the two parameters c and d were plotted on separate graphs. This is only appropriate if the two parameters are perceptually independent; i.e. if perceived c is not a function of d , and vice versa. When Pellacini *et al.* [9] proposed their reparameterization of the Ward model,

they reported that the two parameters were perceptually independent. Our data support this finding.

On average there was no statistical dependence of perceived c on the Test value of d , nor of perceived d on the Test value of c . By binning the values of both Match parameters into small, intermediate and large values, we were able to identify the origins of any interactions we found. Although some illuminations did show small significant interactions, these were inconsistent across bins and across illuminations. This suggests that the small interactions are biases introduced by the specific illuminations, rather than a fundamental lack of independence in the visual representation of these two parameters.

3.4. Comparison between real-world and artificial illuminations.

Subjects are more accurate at matching surface reflectance properties when objects are illuminated under real-world conditions than when illuminated with a single point source, for example. Figure 9 shows r.m.s. error across subjects for each illumination condition. The illuminations have been ranked by error. Light bars represent the real-world illuminations and dark-bars represent the various artificial illumination conditions. In both (a) and (b) the tendency for subjects to perform better with real-world than with artificial illuminations is clear.

Of the spheres shown in Figure 5, the one illuminated under a single extended rectangular source looks more similar to the “real-world” spheres, than to the “point source” sphere. This similarity in appearance is reflected in the results. As can be seen in Figure 9, accuracy for spheres illuminated under the extended source was comparable to the “real-world” spheres, at least for parameter d . This suggests that extended luminance edges might play an important role in our ability to perceive surface reflectance.

Although r.m.s. error is a reasonable metric of performance accuracy, it fails to capture quite how inaccurate subjects’ matches are for some of the artificial illuminations. Figure 10 shows mean matching plots across subjects for the specular contrast parameter, c . The “Uffizi” illumination shown in (a) resulted in the least accurate matches of any of the real-world illuminations, and yet these errors reflect a systematic bias, presumably introduced by idiosyncratic statistical properties of that illumination. By contrast, the matches for the white noise and pink noise conditions are disorganized and noisy. Presumably this unreliability reflects the difficulty that subjects experienced in interpreting these patterns as specular reflections. Indeed, the example images shown in Figure 5 demonstrate that random patterns of illumination do not lead to reliable percepts of surface reflectance. This is particularly important in the case of the pink noise, which has a similar amplitude spectrum to

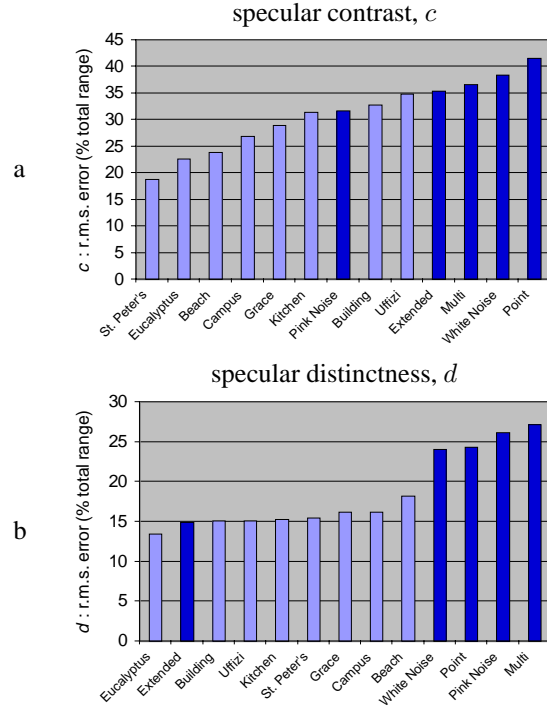


Figure 9. (a) R.m.s error for data pool across subjects for specular contrast c . (b) R.m.s. error across subjects for specular distinctness d . The different illuminations have been ranked in ascending order of error. Light bars represent real-world illuminations, dark bars represent artificial illuminations. Artificial illuminations tend to have higher error (worse performance) than real-world illuminations.

typical real-world illuminations [5]. That subjects perform worse for this condition suggests that the representation of stored assumptions about the regularities of real-world illuminations does not solely contain information about spatial frequencies.

3.5. Are subjects matching “high-level” or “low-level” features?

As discussed above, there are two ways that subjects can perform the matching task: by attending to their “high-level” impression of surface reflectance, or by perceptually equating “low-level” features that correlate with surface reflectance. Both tasks are underconstrained, but it is interesting to investigate which approach subjects take. To this end, a control condition was run featuring photographic negatives of some of the real-world illuminations.

Figure 11 features mean matching plots across subjects for an example real-world illumination (Eucalyptus) and its negative counterparts. For specular contrast, c , subjects’

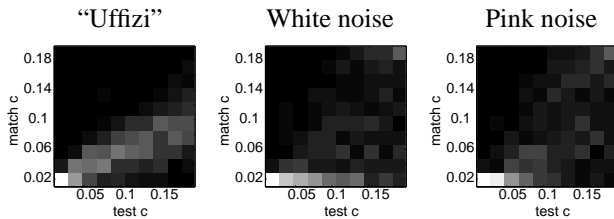


Figure 10. Matches for parameter c for spheres viewed under three different illuminations. In (a) matches for the “Uffizi” real-world illumination have been pooled across subjects. Of all the real-world illuminations, this illumination had the highest r.m.s. error across subjects. The error reflects a large consistent bias in performance. (b) and (c) show similar plots for the white noise and pink noise conditions respectively. Note the noisy performance of subjects for the artificial illuminations, relative to the real-world illumination.

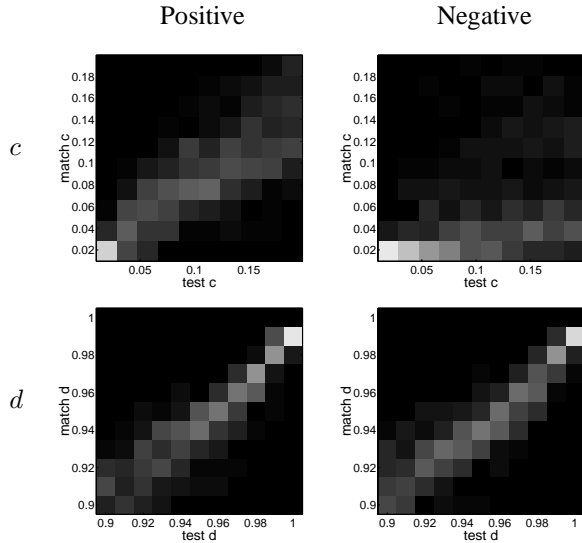


Figure 11. Matching plots for positive (left) and negative (right) versions of the “Eucalyptus” real-world illumination. The top graphs are for specular contrast, c and the bottom graphs for specular distinctness, d . Data have been pooled across subjects.

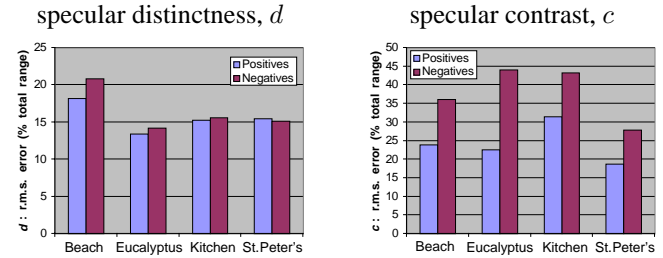


Figure 12. Summarized r.m.s. error in estimates of c and d for positive and negative images. Light bars represent positive images and dark bars represent their negative counterparts.

matches are less accurate in the negative Control condition than in the positive version. When subjects are matching specular distinctness, d , performance is approximately equal for positive and negative versions of the stimuli.

These data are summarized in Figure 12. Figure 12(a) shows r.m.s. error across subjects for the d parameter. Errors are statistically equal for positive and negative versions of the stimuli. This does not guarantee that subjects are not using their percept of reflectance to perform the task. However, it does not exclude the possibility that the perception of spatial spread may be based on the low-level representations, rather than additional information carried by high-level representations of surface reflectance. Figure 12(b) shows r.m.s. error for the c parameter. Subjects are significantly better with the positive stimuli than their negative counterparts. This suggests that the high-level representation of surface reflectance contains additional information that helps subjects perform the task, because interruption of this percept reduces performance accuracy. This also suggests that subjects did indeed follow their instructions to attend to the material quality of the spheres.

3.6. What cues are subjects using? Comparisons between illuminations.

Even though subjects match reflectances accurately under unknown real-world illumination, they exhibit biases dependent on illumination. When we pooled data across all subjects, we found that estimates of c for each illumination tended to differ from the veridical values by a multiplicative factor, while estimates of d differed from the veridical values by an additive factor. In other words, estimates \hat{c} of c cluster along a line $\hat{c} = mc$, where m is a constant specific to the test illumination. Estimates \hat{d} of d cluster along a line $\hat{d} = d + a$, where a is a constant specific to the test illumination. Figure 13 shows the best fit coefficients m and a for each illumination, making systematic biases between

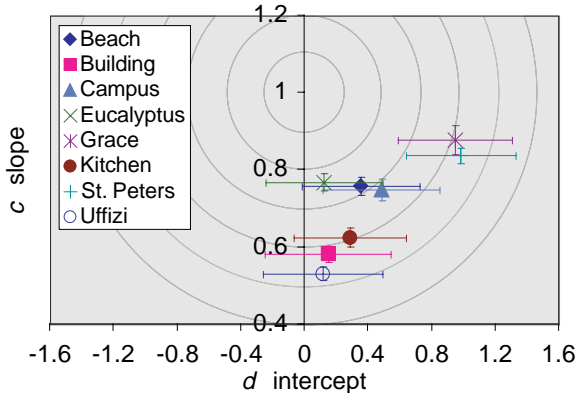


Figure 13. Comparison between matches under real-world illuminations. The x -axis represents an additive bias term for estimates of d , while the y -axis represents a multiplicative error factor for estimates of c . Point $(0, 1)$ represents unbiased performance.

illuminations explicit. The point where $m = 1$ and $a = 0$ corresponds to unbiased estimates of c and d , respectively.

These biases correlate with certain properties of the illumination maps. For example, illumination maps containing bright, point-like sources typically lead to increased estimates of both the contrast and distinctness of the specular reflectance. The “Grace” and “St. Peter’s” illuminations, which are dominated by bright indoor lights, exhibit the highest values for both m and a in Figure 13. In future work, we plan to test the relationship between illumination map characteristics and reflectance estimation biases by measuring subject performance with modified versions of illumination maps. For instance, we can test the importance of the brightest light source by removing it while leaving the remainder of the illumination map unaltered.

One salient aspect of Figure 13 is that m is consistently less than 1, while a is consistently greater than 0 across illuminations. There are two possible explanations. One is that the “Galileo” illumination, which was used as the Match probe, is a statistical outlier in terms of the features that subjects use to estimate c and d . Another possibility is that the matching method introduces some systematic response biases, as subjects prefer to avoid the ends of the scales, for example. A pilot experiment featuring a different illumination for the Match probe suggests that both of these factors play a role.

4. Conclusions

Estimating surface reflectance properties under unknown illumination is underconstrained because many combinations of lighting and reflectance can give rise to the same

images. However, we have shown that in the absence of contextual cues to illumination, subjects can match surface reflectance properties.

Real-world illumination contains statistical regularities and subjects match spheres illuminated under real-world illumination more accurately than they can match spheres illuminated under point light sources. It seems likely that subjects make use of stored assumptions about the regularities of real-world illumination in order to perform the task.

This research was supported by NIH grant EY12690-1 to E.H.A., by a grant from NTT to the MIT Artificial Intelligence Lab, by a contract with Unilever Research, and by NDSEG and Whitaker Fellowships to R.O.D.

References

- [1] http://www.ai.mit.edu/people/rondror/papers/rf_bg.tiff.
- [2] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge, MA, 1999.
- [3] P. E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. *Computer Graphics (SIGGRAPH)*, 2000.
- [4] R. O. Dror, E. H. Adelson, and A. S. Willsky. Surface reflectance estimation and natural illumination statistics. In *Proc. of IEEE Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, July 2001.
- [5] R. O. Dror, T. Leung, A. S. Willsky, and E. H. Adelson. Statistics of real-world illumination. In *Proceedings of CVPR*, 2001.
- [6] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou. An anchoring theory of lightness perception. *Psychological Review*, 106:795–834, 1999.
- [7] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am. A*, 61:1–11, 1971.
- [8] 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.
- [9] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. *Computer Graphics (SIGGRAPH)*, 2000.
- [10] J. Tumblin, J. K. Hodgins, and B. K. Guenter. Two methods for display of high-contrast images. *ACM Transactions on Graphics*, 18:56–94, 1999.
- [11] G. J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH)*, 26(2):265–72, 1992.
- [12] G. J. Ward. The RADIANCE lighting simulation and rendering system. *Computer Graphics (SIGGRAPH)*, 1994.
- [13] G. Ward Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3, 1997.