massachusetts institute of technology — artificial intelligence laboratory

Surface Reflectance Estimation and Natural Illumination Statistics

Ron O. Dror, Edward H. Adelson and Alan S. Willsky

AI Memo 2001-023 September 2001

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

Humans recognize optical reflectance properties of surfaces such as metal, plastic, or paper from a single image without knowledge of illumination. We develop a machine vision system to perform similar recognition tasks automatically. Reflectance estimation under unknown, arbitrary illumination proves highly underconstrained due to the variety of potential illumination distributions and surface reflectance properties. We have found that the spatial structure of real-world illumination possesses some of the statistical regularities observed in the natural image statistics literature. A human or computer vision system may be able to exploit this prior information to determine the most likely surface reflectance given an observed image. We develop an algorithm for reflectance classification under unknown real-world illumination, which learns relationships between surface reflectance and certain features (statistics) computed from a single observed image. We also develop an automatic feature selection method.

1 Introduction

The upper portion of Figure 1 shows images of nine spheres, each photographed in two real-world settings. The two images of each sphere are completely different at the pixel level because illumination varies from one location to another. Yet, a human observer easily recognizes that the images in each column represent spheres of similar materials, while the images in different columns represent spheres of different materials. A human could classify a sphere photographed in a third setting into one of these nine categories according to its apparent material properties. This paper develops a computer algorithm with a similar ability to recognize surface reflectance.

The reflectance estimation problem proves ill posed in the absence of restrictions on the illumination. All the spheres in Figure 1 could be perfect chrome reflectors; because a mirrored surface simply reflects its surrounding environment, a properly illuminated chrome sphere could take on an arbitrary appearance. Different combinations of reflectance and illumination could explain the observed data even when images of the surface are available from all directions [22].

We wish to exploit information about the real world to determine the most likely surface reflectance given an observed image. By analogy, consider the situation where one wishes to estimate the variance of a Gaussian filter given only a single image which has been blurred by that filter. In the absence of information about the original image, one can only place an upper bound on the variance. If one assumes, however, that the original image was a properly focused, real-world photograph, then one can estimate the variance much more accurately by taking advantage of the fact that real-world scenes contain a great deal of statistical structure. For example, real-world images typically exhibit characteristic structures in the frequency and wavelet domains [23, 15, 28].

We believe that accurate reflectance estimation by humans or machines depends on the spatial structure of real-world illumination, which shares some of the statistical regularity studied in the literature on natural image statistics. We show that the relationship between the reflectance of a surface and the statistics of its image depends on the structure of natural illumination. We then extend our earlier work [7] on a system which classifies surface reflectance from images using a machine learning approach. We have also developed

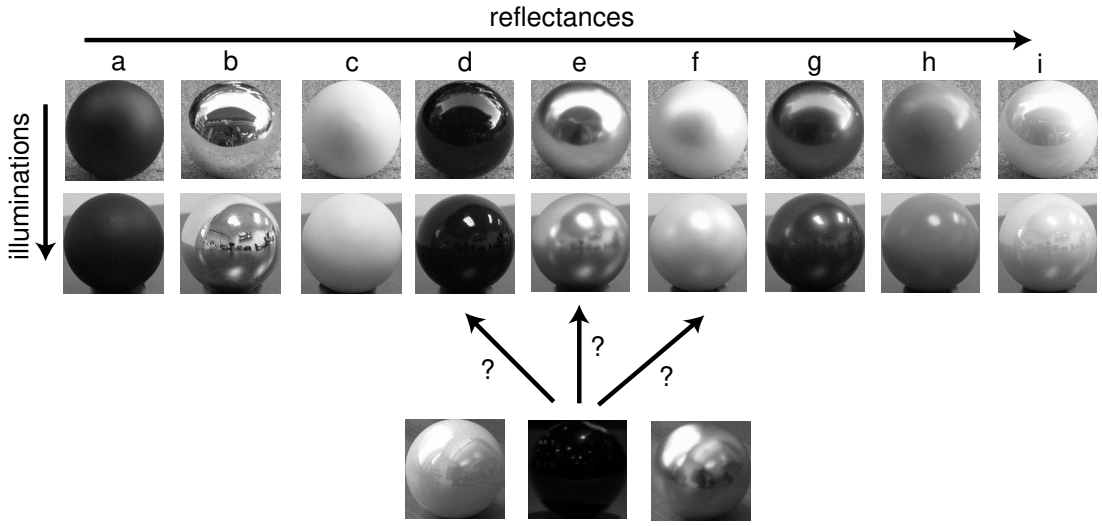


Figure 1: The problem addressed by our classifier, illustrated using a database of photographs. Each of nine spheres was photographed under seven different illuminations. We trained a nine-way classifier using the images corresponding to six illuminations, and then used it to classify individual images under the seventh illumination. The classification algorithm uses image data only from the surface itself, not from the surrounding background.

a principled method to choose from a bank of candidate features a subset that improve classification accuracy given a limited amount of training data.

Several practical applications in computer vision and graphics motivate our efforts. First, an ability to estimate reflectance under unknown illumination facilitates visual material recognition, because different physical surfaces such as metal, plastic, and paper possess different optical reflectance properties. Second, reconstruction of a scene from photographs for computer graphics requires inference of both the geometry and the reflectance of visible surfaces. Third, an ability to estimate reflectance from image data under unknown lighting conditions may help overcome the limitations of shape-from-shading algorithms which assume that reflectance is known in advance, and of classical algorithms for motion or stereo estimation, which assume Lambertian surface reflectance.

We classify reflectance from a single monochrome image of the surface of interest, without using contextual cues from the surrounding image. Our techniques could be improved by drawing on additional sources of information, including color spectral decompositions [26, 30], motion cues, and visual context. Humans, however, can effectively estimate certain surface reflectance properties even in the absence of these cues [12]. We chose to work with single monochrome images, because we wish to determine what information the basic image structure captures about reflectance. We simplify the problem at this stage by assuming that the surface under observation has homogeneous reflectance properties and known geometry.

2 Related Work

The importance of reflectance models in computer graphics has motivated several researchers to develop image-based reflectance estimation techniques. Tominaga *et al.* [30] present a method for estimating Phong model [18] parameters from an image of a uniform cylindrical surface illuminated by a point light source. Sato *et al.* [24] as well as Marschner [17] develop similar techniques which accommodate a more general geometry acquired through laser range scanning. These methods, unlike ours, do not apply to photographs taken in the natural world under complex lighting conditions.

The inverse global illumination techniques of Yu *et al.* [32] handle reflectance estimation without fully specified illumination, but require a collection of photographs representing all surfaces and primary light sources in a scene. This technique iteratively estimates both the illumination and reflectance of every surface patch in the scene. Our approach, on the other hand, requires only an image of the surface whose reflectance is in question. We avoid estimating illumination explicitly by characterizing it statistically.

Ramamoorthi and Hanrahan [22] perform mathematical analysis to determine when the reflectance estimation problem is well posed. Their work is complementary to ours; they show that reflectance estimation is ill posed in the absence of knowledge about illumination, while we handle the problem in precisely these cases. Unlike their approach for joint recovery of illumination and reflectance, our reflectance classification technique does not assume the presence of a known directional source, the availability of multiple images, or a particular form of reflectance.

3 Problem formulation

The reflectance of an opaque surface patch can be described by a bidirectional reflectance distribution function (BRDF), which specifies what proportion of the light incident from each possible illumination direction is reflected in each possible view direction. Since two angles are necessary to specify a direction in a three-dimensional world, the BRDF is a function of four continuous angular variables. Given the BRDF of a surface patch, one can perform a weighted integration over the illumination incident on the surface patch from every direction in the hemisphere surrounding its normal to determine the brightness of the patch in an image.[1]

Estimating a BRDF from a single observed image under unknown illumination proves highly ill posed. The observed image is a function of two continuous variables, whereas the

[1] Denote the BRDF by $f(\theta_i, \phi_i; \theta_r, \phi_r)$, where θ_i and θ_r are the angles of the incident direction and the view direction, respectively, from the surface normal, and ϕ_i and ϕ_r are their azimuthal angles. $f(\theta_i, \phi_i; \theta_r, \phi_r)$ indicates the ratio of the reflected radiance in direction (θ_r, ϕ_r) to incident irradiance from a differential solid angle centered on direction (θ_i, ϕ_i). If $I(\theta_i, \phi_i)$ gives the radiance of illumination incident from direction (θ_i, ϕ_i), the total reflected radiance of the surface patch in the view direction (θ_r, ϕ_r) is given by [13]

$$\int_{\phi_i=0}^{2\pi} \int_{\theta_i=0}^{\pi/2} f(\theta_i, \phi_i; \theta_r, \phi_r) I(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i.$$

If illumination is distant, the reflected light field (a collection of images of the surface from all angles) may be expressed as a convolution of the illumination and the BRDF [22]. In this paper we have access to only a single image, which cannot be expressed exactly as a convolution.

BRDF is a function of four continuous variables. The BRDF of a passive surface must satisfy conservation of energy and reciprocity [13], but the space of physically realizable BRDFs remains huge. Moreover, each pixel of the image depends not only on the BRDF, but on the unknown illumination, a function of two variables which may differ from one point on the surface to another. The unknowns occupy a much larger space than the observations.

Ideal Bayesian estimation of reflectance would require marginalizing over all possible illuminations to find the most likely BRDF $\hat{f}$ for a given observed image:

$$\hat{f} = \arg\max_f P(f|R) = \arg\max_f P(f) \int_I P(I)P(R|f,I)dI, \tag{1}$$

where I denotes illumination from every direction at each point on the surface, and R denotes the observed radiance of each point in the image. The prior probability over BRDFs $P(f)$ captures the fact that some reflectance functions are more likely than others in the real world; for example, white matte surfaces are more common than holograms. Likewise, $P(I)$ captures the statistical structure of real-world illumination.

Unfortunately, even if one could formulate the prior probabilities over illumination and reflectance explicitly, integration over all possible illuminations and maximization of this integral over all possible BRDFs would prove computationally daunting.[2] We use two alternative approaches in this paper. In Section 4, we utilize a parameterized reflectance model from computer graphics to examine the relationship between illumination statistics, image statistics, and reflectance. In Sections 5 and 6, we build a system to solve the reflectance classification task of Figure 1. This system works with a finite set of arbitrary reflectances, learning to differentiate between them based on training images photographed or rendered under natural illumination.

4 Dependence of image statistics on reflectance

4.1 Statistical regularity of real-world illumination

One can measure the illumination incident from every direction at a particular point in the real world using a camera located at the point of interest. By combining photographs from that point in every direction, one can compose a spherical illumination map. If all sources of direct and indirect illumination are relatively distant, the illumination map remains fixed as the camera moves through space. One can therefore use such an illumination map to render an object as it would appear at that location in space, using a generalization of traditional environment mapping [6].

[2] Replacing the integration with a maximum over illuminations may lead to incorrect reflectance estimates. In other words, joint estimation of illumination and reflectance, even when feasible, may fail to identify the most likely reflectance. Consider a photograph of a white matte sphere, corrupted by slight high-frequency imaging noise. One could explain this image approximately as a white matte sphere under any of a number of illuminations, but none of these would predict the noisy image exactly. On the other hand, one could explain the photograph precisely as a chrome sphere under just the right illumination. Thus the single most likely combination of reflectance and illumination might indeed involve a chrome sphere. Integrating over all possible illuminations would reveal that a more likely reflectance is white matte, because for that reflectance a large number of illuminations produce approximately the observed image. Unlike Yu *et al.* [32] and Ramamoorthi and Hanrahan [22], we wish to identify the most likely reflectance rather than explaining the observed image data as a single combination of illumination and reflectance.

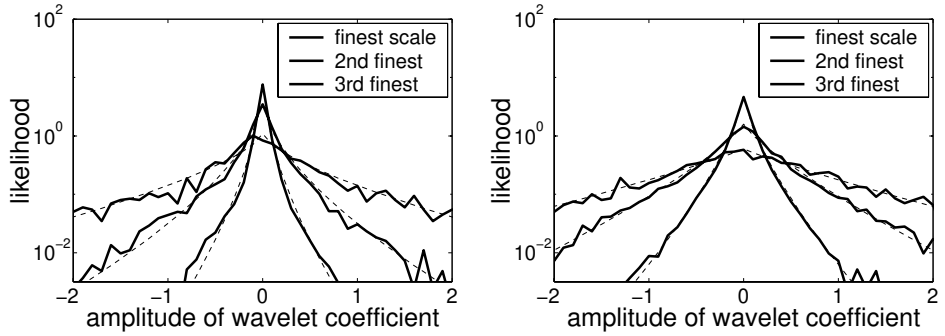
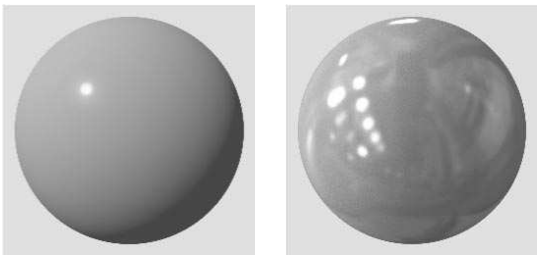


Figure 2: Thick lines indicate distributions of horizontally-oriented wavelet coefficients at three successive scales for an indoor illumination map (left) and an outdoor illumination map (right). Thin lines are maximum likelihood generalized Laplacian distributions fit to the empirical distributions. All the distributions have high kurtosis, with variance increasing at coarser scales. Wavelet coefficients were computed with a nine-tap quadrature mirror filter pyramid [27] from log-luminance equal-area cylindrical projections of the spherical maps. Both illumination maps (Galileo's tomb, eucalyptus grove) are due to Debevec [5].

A photographically-acquired illumination map is a type of real-world image. Although it differs in field of view and dynamic range from the photographs studied in the natural image statistics literature [11, 15], it shares many of the statistical regularities of these photographs, as shown in a separate paper [9]. Most natural illumination maps contain a wide range of intensities corresponding to illumination or lack of illumination from various directions. If pixel intensity is linear in luminance, the pixel histogram typically has positive skew due to the presence of sparse bright light sources. Marginal and joint distributions of wavelet coefficients at various scales and orientations exhibit similar heavy-tailed distributions from illumination map to illumination map (Figure 2).

Figure 3 shows synthetic images of two identical spheres under different illuminations. Humans identify surface reflectance more easily in image B, rendered under a photographically-



A B

Figure 3: (A) A sphere rendered under illumination by a point light source. (B) The same sphere rendered under photographically-acquired illumination. The reflectance of the sphere is specified by the Ward model (see Section 4.2), with parameters $\rho_d = 0.083$, $\rho_s = 0.097$, $\alpha = 0.03$.

acquired illumination map, than in image A, rendered under point source illumination. Comparison of photographs of a sphere in a normally illuminated room and in a black room with a point light source reveals the same effect. The simplicity of true point source illumination violates the statistical regularities of typical natural illumination. Much previous work in reflectance estimation has considered the case of point source illumination as a convenient starting point. We wish instead to take advantage of the structure of natural illumination in estimating reflectance.

4.2 Relationship between image statistics and Ward model parameters

In this section, we use a parameterized reflectance model from computer graphics to gain insight into the relationship between illumination statistics, surface reflectance, and surface image statistics. Using synthetic images rendered under photographically acquired illumination, we examine the dependence of image statistics on model reflectance parameters. In Section 5, we apply some of our insights to develop a reflectance classification algorithm which handles arbitrary reflectances.

We use a popular reflectance model from computer graphics due to Ward [31]. The isotropic Ward reflectance model, which is a physically realizable variant of the common Phong shading model, specifies a BRDF of the form

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi} + \rho_s \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \frac{\exp(-\tan^2\delta/\alpha^2)}{4\pi\alpha^2}, \tag{2}$$

where δ is the angle between the surface normal and a vector bisecting the incident and reflected directions. The free parameters of this model are ρ_d, the fraction of incident energy reflected by the diffuse (Lambertian) component, ρ_s, the fraction of energy reflected by the specular component, and α, surface roughness measured as the standard deviation of surface slope. Higher α implies a more blurred specular component. Although developed empirically, the Ward model has been shown to provide an accurate fit for the BRDFs of certain classes of materials [31].

We eliminated some of the complicating effects of geometrical distortion by considering flat homogeneous surfaces under distant illumination. To a nearby viewer with a narrow field of view, the specular reflection of such a surface is approximately a projection of the illumination field blurred by a Gaussian filter. Because the surface is flat and illumination is distant, the diffuse component is constant across the surface. The top row of Figure 4 shows images of four different surfaces rendered using the same photographically-acquired illumination.

To determine how various statistics of the observed image vary with each reflectance parameter under natural illumination, we computed empirical derivatives of image statistics with respect to ρ_d, ρ_s, and α. We rendered flat surfaces using the previously described simplifying approximations under nine photographically-acquired illumination maps from Debevec's Light Probe Image Gallery (*http://www.debevec.org/Probes/*) [5]. These illumination maps represent diverse lighting conditions from four indoor settings and five outdoor settings. We normalized the overall intensity of each illumination as described in Section 5.4. We measured the sensitivity of a statistic to a parameter as the mean derivative of the statistic with respect to the parameter, normalized by the standard deviation of that statistic

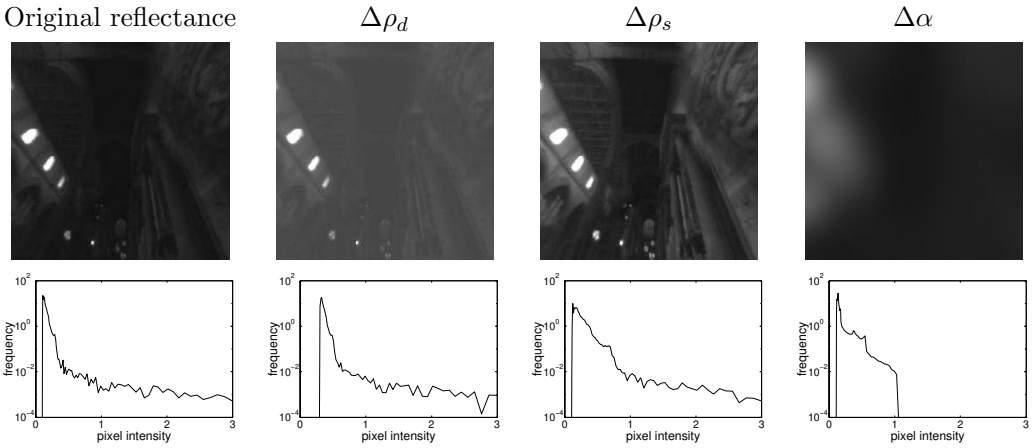


Figure 4: Effect of changes in Ward model reflectance parameters on an image and its pixel intensity histogram. The top row shows images of a flat surface under a photographically-acquired illumination map due to Debevec [5]. The bottom row shows corresponding histograms for each image, with logarithmic vertical axes. The leftmost column has reflectance parameter values $\rho_d = .1$, $\rho_s = .1$, and $\alpha = 0$. The remaining columns have the same parameter values as the first except that ρ_d increases to .3 in the second column, ρ_s increases to .3 in the third column, and α increases to .1 in the fourth column. The surface normal direction is 45° from the vertical in each case, and the images have a 90° field of view.

across illuminations.[3]

Inspired by previous work on natural image statistics and texture modeling [14, 4], we worked with statistics based on the distributions of pixel intensities and of wavelet coefficients at different scales and orientations. Heeger and Bergen [14] based their texture representation on distributions of the same quantities.

Figure 5 shows computed sensitivity values of several statistics characterizing the pixel intensity histogram, including the first four moments and several percentiles of the distribution. Large values, either positive or negative, indicate that variations in a particular statistic due to change in a parameter are large relative to variations of that statistic between illuminations. The results suggest that the 10th percentile of pixel intensity may be particularly useful in estimating ρ_d, that the mean and variance of intensities are relevant to estimation of ρ_s, and that the variance, skew, and kurtosis of the intensity distribution decrease as α increases.

To aid in interpreting these results, Figure 4 illustrates the effects of increasing each reflectance parameter on a particular image and its histograms. Because the diffuse reflectance component is constant for these flat images, increasing ρ_d simply shifts the entire histogram to uniformly higher values. Most illumination maps contain regions of low illumination,

[3]Let x_i denote the values of a particular statistic at a fixed parameter setting for illuminations $i = 1, 2, \ldots, N$. Let $\tilde{x}_i$ be the corresponding value of the statistic for each illumination when a particular parameter increases by a small quantity Δp. Let μ and σ denote the mean and standard deviation of x_i, while $\tilde{\mu}$ and $\tilde{\sigma}$ denote the corresponding quantities for $\tilde{x}_i$. We measure the local sensitivity of the statistic to the parameter being changed as $\frac{\tilde{\mu}-\mu}{\Delta p\sqrt{(\sigma^2+\tilde{\sigma}^2)/2}}$.

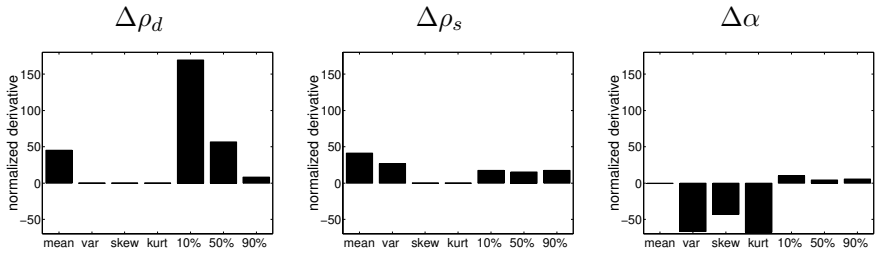


Figure 5: Normalized derivatives of pixel histogram statistics with respect to ρ_d, ρ_s, and α. The bars in each plot represent sensitivity of the mean, variance, skew, kurtosis, 10th percentile, median, and 90th percentile of the pixel intensity distribution. Derivatives were computed at parameter values $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .01$.

where the specular component contributes little to observed radiance. The darkest areas of the observed image, as measured by the 10th percentile of pixel intensities, therefore prove indicative of its diffuse reflectance.

Increasing the strength of specular reflectance ρ_s scales the specular component of the reflectance, increasing the mean and variance of the pixel distributions. Increasing the surface roughness α blurs the specular reflection, eliminating outliers. Because most outliers are due to localized light sources or reflections, blurring the image not only reduces variance and kurtosis of the histogram, but also reduces its skew.

We performed similar analysis for statistics summarizing the distributions of wavelet coefficients. We constructed a two-dimensional pyramid from each image using nine-tap symmetric quadrature mirror filters [27]; our experience is that the precise choice of the pyramidal decomposition is not critical in gathering image statistics. We then computed moments and percentiles of the coefficients at each scale and orientation. Figure 6 shows the normalized derivatives of several such statistics with respect to the α blur parameter. As α increases, the variance and kurtosis of each subband decrease. Variances of the wavelet coefficient distributions at different scales provide an approximate measure of the spectral power in different frequency bands. Kurtoses of these distributions provide a rough measure of image edginess, because edges tend to produce extreme values of wavelet coefficients, leading to the heavy-tailed (highly kurtotic) distributions typical of natural illumination. At the small value of α (0.01) used in computation of the derivatives for this figure, the variances of the finest scale subbands are particularly sensitive to a change in blur. At higher values of α, the variances of the coarser subbands become more sensitive than the variances of the finer subbands.

For curved geometries or for wider view angles, the relationships between parameter values and image statistics become more complicated. The radiance of the diffuse component is not constant across the surface, and statistics of the observed image are non-stationary even if illumination statistics are stationary (Section 5.1). However, computations of normalized derivatives from rendered images yield similar results. Figure 7 shows how differently illuminated surfaces of a particular reflectance tend to cluster in a space defined by appropriately chosen image statistics. The horizontal axis of the scatter plot represents the 10th percentile of pixel intensities, while the vertical axis represents the variance of the horizontal bandpass

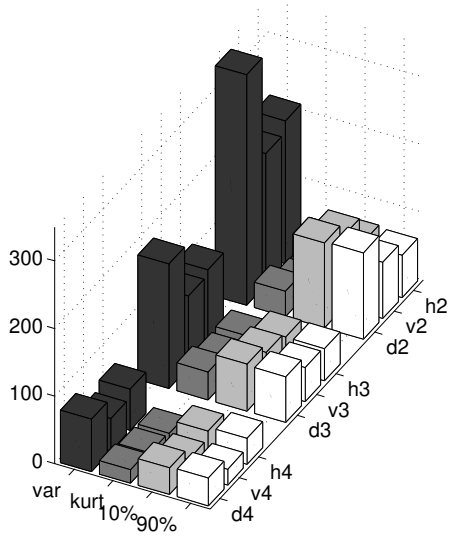


Figure 6: Absolute values of normalized derivatives of wavelet coefficient statistics with respect to α parameter. Statistics shown are the variance, kurtosis, 10th percentile and 90th percentile of the horizontal, vertical, and diagonal wavelet coefficients at the second, third, and fourth finest scales. The variance, kurtosis, and 90th percentile of each subband decrease with increasing α, while the 10th percentile increases. As in Figure 5, derivatives are computed at parameter values $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .01$.

filter output at the second finest scale. The former correlates with diffuse reflectance, while the latter depends on the strength and sharpness of the specularity.

In order to estimate reflectance accurately, we wish to choose not a single informative statistic, but several image statistics which jointly characterize reflectance. Normalized derivatives do not indicate whether two statistics capture similar or different information about reflectance. Section 5.3 describes one approach to addressing this issue.

If the normalized derivative computations are repeated with illuminations whose statistics differ from those of natural images, the results may differ significantly. For example Figure 8 is analogous to Figure 5, except that the natural illumination maps were replaced by eight-by-eight grids, with each grid square chosen by an unbiased coin toss as black or white. The relative values of the normalized derivatives are quite different from those of Figure 5. When α increases, for example, variance tends to increase, but skew does not change significantly, and kurtosis actually increases. At low α, the pixel histogram is binary; as α increases, it approaches a Gaussian shape. The normalized derivatives of Figure 8 have significantly larger magnitudes than those of Figure 5, due to the greater regularity of the "checkerboard" illuminations.

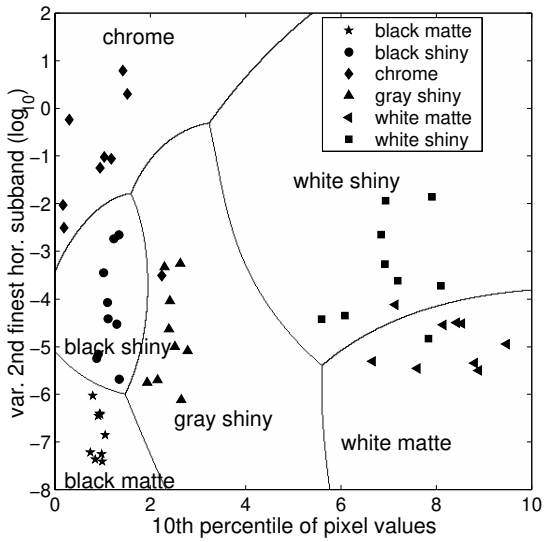


Figure 7: Solid symbols indicate locations in a two-dimensional feature space of images of spheres of six reflectances (Figure 11), each rendered under nine different real-world illuminations due to Debevec. Lines separate the regions which the SVM classifier (Section 5.2) assigns to different reflectances. Statistics were computed after the images were warped as described in Section 5.1.

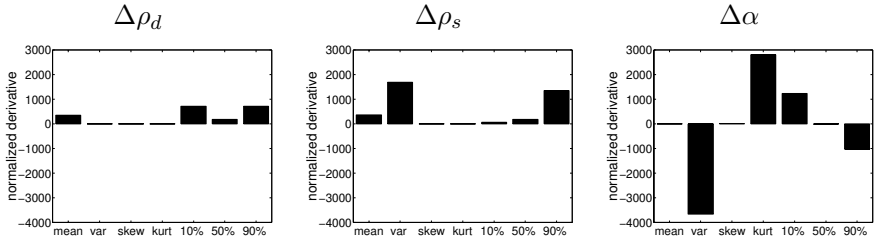


Figure 8: Normalized derivatives corresponding to those of Figure 5, but for images rendered under random "checkerboard" illuminations. Vertical axes are different from Figure 5.

5 A method for reflectance classification

To solve the classification problem of Figure 1, we attempt to select from a finite set of reflectances the one which most closely represents the reflectance of an observed surface. The candidate reflectances may be defined by different parameter settings of a reflectance model such as the Ward model, but they could also be arbitrary BRDFs specified as lookup tables or by a set of photographs of a real surface. Given images of several surfaces under various illuminations, we wish to classify photographs under unknown, novel illumination.

We base our classifier on statistics such as those examined in Section 4.2. Specifically, we characterize pixel intensity distributions and distributions of wavelet coefficients at each scale and orientation using their mean, variance, skew, kurtosis, and 10th, 50th, and 90th percentiles. We also consider several additional families of statistics, including spectral power in oriented frequency bands. The following sections consider the effects of surface geometry on the observed image as well as our choice of a classification technique and of classification features.

5.1 Geometry

If we assume that both the observer and illumination are distant relative to the curvature of a convex surface under observation, the brightness a surface point depends only on the local surface orientation. Hence an image of the surface determines a relationship between surface orientation and observed radiance for a particular illumination condition. We have chosen to illustrate our techniques using spheres because we can capture their geometry precisely from image contours. In a separate paper, we extend our technique to other objects of known geometry, and show that it proves robust to inaccurate geometry estimates [8].

We wish to compute statistics on an image representing radiance as a function of orientation. An image of a sphere possesses non-stationary image statistics because a perspective or orthographic projection compresses features near the edges. One could reduce these effects by considering radiance as a function of orientation and performing the analysis in a spherical domain,[4] using spherical wavelet transforms [25]. We have chosen initially to simply warp the observed image into one with more nearly stationary statistics. In particular, our algorithm extracts an annulus of the spherical image and unwraps it into a rectangle using a polar-to-rectangular coordinate transformation (Figure 9). Other warping methods yield similar results [8].

5.2 Classification techniques

Whether we choose features by hand or automatically, we would like our classifier to be robust to the inclusion of some features whose values vary significantly from image to image but provide little information about the reflectance class. We chose support vector machines

[4]Even when the image is defined as a function on the sphere instead of the plane, its statistics are not stationary. Consider a typical reflectance function consisting of a diffuse and a specular component. A localized light source contributes most to the diffuse component for a surface patch whose normal is in the light source direction. The same light source contributes most to the specular component for a surface patch whose normal bisects the light direction and the view direction. Even if illumination is isotropic, image statistics will vary spatially.

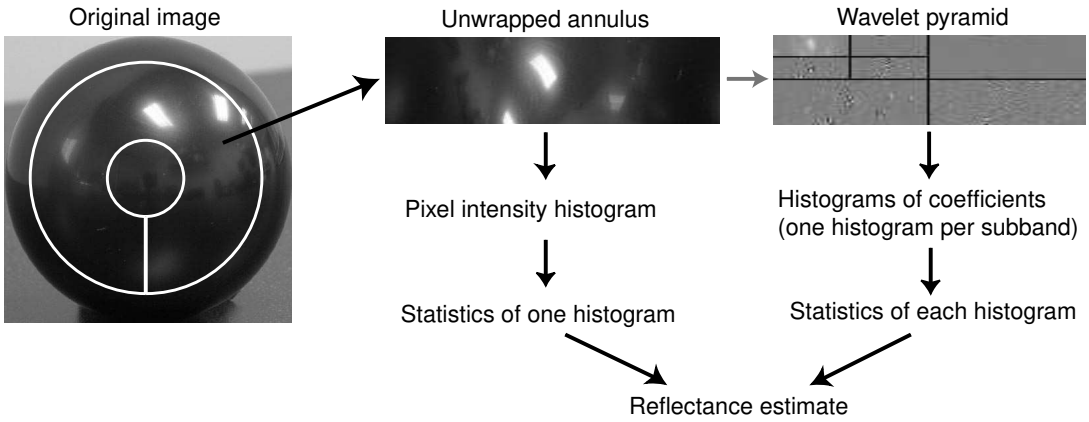


Figure 9: Flowchart for computation of image features, which applies to both testing and training of the classifier. The features are histogram statistics, computed on the original image and on its wavelet transform.

(SVMs) for classification because they tend to generalize well given a limited number of training samples and a large number of features [10]. Our implementation utilizes the SVMTorch software [3] with Gaussian kernels to train and apply SVM classifiers. SVMTorch uses a one-against-all voting scheme to perform multiclass classification. As discussed in Section 6, we found that SVM classifiers substantially outperform nearest neighbor and k-nearest neighbor classifiers when the number of features is large compared to the number of training images. Figure 7 shows the class boundaries determined by an SVM classifier based on two statistics.

5.3 Feature selection

In practice, we have only a limited amount of training data, because either the surface images themselves or the illumination maps used to render them must be acquired photographically in real-world environments. Given a large set of candidate features, we wish to automatically identify a small subset which will yield high classifier performance given limited training data. In other words, we wish to select those features whose variation from one image to another captures the most information about the reflectance class corresponding to an image. Mathematically, we wish to find a set of N features which minimizes the Bayes error for this classification problem.

To perform this minimization explicitly, one must compute the Bayes error for every combination of N features. One might limit the combinatorial explosion by using a greedy algorithm which chooses features sequentially, but choosing the Nth feature would still require estimates of joint probability densities for N features. The number of training samples required to estimate an N-dimensional probability distribution grows exponentially with N.

Because we have a limited number of training samples, we need to avoid high-dimensional joint density estimation. We eliminate the need for multidimensional probability density estimation by using an iterative method which alternates between the following two steps:

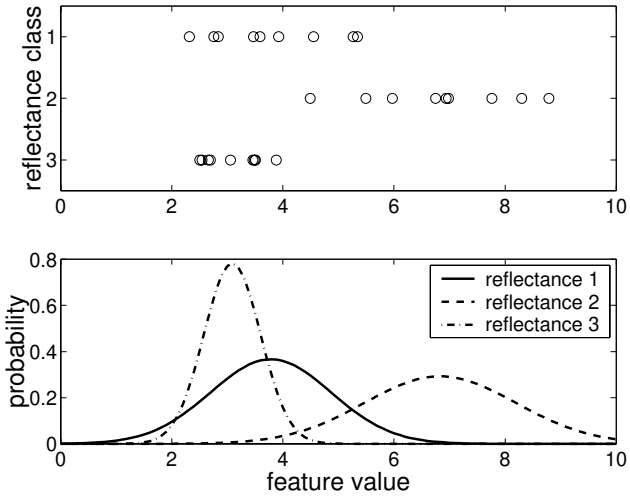


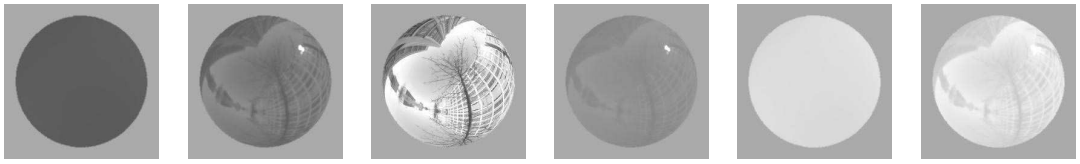
Figure 10: Bayes error calculation for a three-way classification problem for a single feature. The upper subplot shows the scatter of values of a feature (the variance of coefficients in one wavelet subband) for three different reflectance classes. The lower subplot shows Gaussian distributions which have been fit to the feature values for each class. The Bayes error is the error probability of a classifier which always chooses the most likely class given a particular feature value. One computes it by integrating over the appropriate portions of each distribution.

1. Estimate the marginal probability density of each individual feature for images of a particular class (Figure 10). Choose the feature which, when used alone for classification, leads to minimal Bayes classification error assuming the estimated probability densities.

2. Regress each of the features which have not yet been selected against the selected feature to obtain a predictive relationship. Modify each feature which has not yet been selected by subtracting off its predicted value based on the selected feature.

We have found empirically that the features discussed in Section 4.2 are unimodally distributed for any particular reflectance class. In fact, the empirically based image model of Portilla *et al.* [20], which describes natural images using a Gaussian scale mixture model in the wavelet domain, suggests that features based on marginal distributions of wavelet coefficients will be log normally distributed. We therefore fit Gaussian distributions to our candidate features in the log domain. Besides the ease and robustness with which one can estimate the mean and variance of a Gaussian given a limited number of data samples, this simple model offers the advantage that the Bayes error reduces to a sum of normal error functions.

We have also found empirically that features tend to vary monotonically with one another. We therefore use simple affine regression to remove the dependence of one feature on another.



black matte black shiny chrome gray shiny white matte white shiny

Figure 11: Synthetic spheres of 6 different reflectances, each rendered under one of Teller's illumination maps. Ward model parameters are as follows: black matte, $\rho_d = .1$, $\rho_s = 0$; black shiny, $\rho_d = .1$, $\rho_s = .1$, $\alpha = .01$; chrome, $\rho_d = 0$, $\rho_s = .75$, $\alpha = 0$; gray shiny, $\rho_d = .25$, $\rho_s = .05$, $\alpha = .01$; white matte, $\rho_d = .9$, $\rho_s = 0$; white shiny, $\rho_d = .7$, $\rho_s = .25$, $\alpha = .01$.

5.4 Lightness ambiguity

Because our analysis techniques rely solely on the image of the surface of interest, they suffer from ambiguity between the overall strength of illumination and the overall lightness of the surface. A white matte sphere under dim illumination and a gray matte sphere under bright illumination will produce identical images. Resolution of this ambiguity requires contextual information from the remainder of the image or scene. Because color constancy and lightness estimation have been studied separately [1, 2], we eliminate this problem from the current study by normalizing our images for overall strength of illumination, as measured by the brightness of a standard white surface positioned perpendicular to the viewer at the position of the surface under observation.

6 Results

6.1 Image sets

We trained classifiers using five different data sets, one comprised of photographs and the others consisting of synthetic images rendered under photographically-acquired illumination maps. We photographed spheres of nine different materials under seven diverse illumination conditions, including both indoor and outdoor settings. Images were acquired in 24-bit RGB format using a Nikon D1 digital camera and then converted to 8-bit gray scale images for further processing. Figure 1 shows examples of these images. The entire image set is available at *http://www.ai.mit.edu/people/rondror/sphere_photos/*. Because the digital camera applies a compressive nonlinearity, these images are not linear in luminance.

Synthetic images have the advantage that the surface BRDF is known exactly. To create synthetic images, we used Ward's *Radiance* package [16], which efficiently implements the Ward reflectance model. Our rendering methodology is similar to that of Debevec [6], but we prefilter the illumination map in the spherical harmonic domain [21] before rendering the diffuse component. Our reflectance algorithm treats the rendering machinery as a black box, using only the final rendered images in the training process.

Our four synthetic data sets consist of images of spheres from two different sets of reflectances, each rendered under two different illumination sets. The first set of six reflectances (Figures 11) was specified by Ward model parameters chosen to correspond to common materials of distinctly different appearances. The second set of reflectances in-

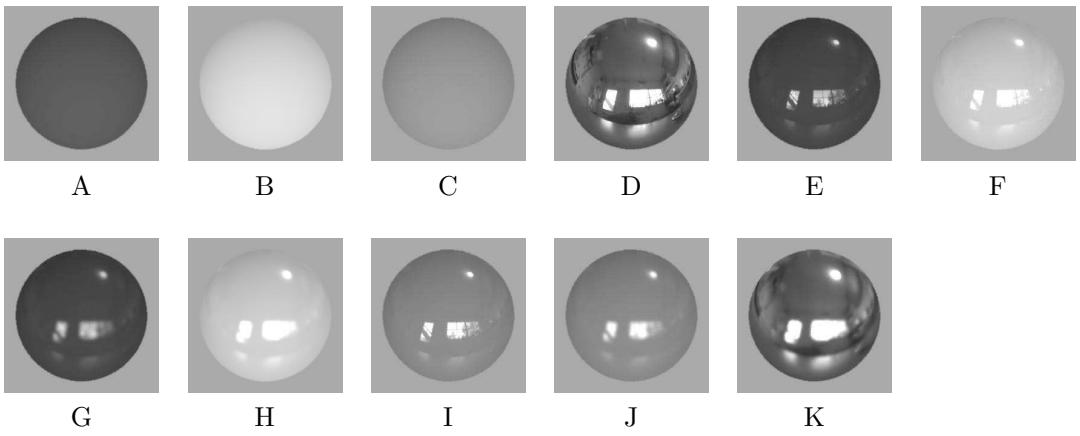


Figure 12: Synthetic spheres representing 11 reflectances, each rendered under an illumination map captured by one of Debevec's light probes. Ward model parameters are as follows: (A) $\rho_d = .1$, $\rho_s = 0$, (B) $\rho_d = .9$, $\rho_s = 0$, (C) $\rho_d = .35$, $\rho_s = 0$, (D) $\rho_d = 0$, $\rho_s = .75$, $\alpha = 0$, (E) $\rho_d = .05$, $\rho_s = .05$, $\alpha = 0$, (F) $\rho_d = .7$, $\rho_s = .2$, $\alpha = 0$, (G) $\rho_d = .05$, $\rho_s = 05$, $\alpha = .02$, (H) $\rho_d = .7$, $\rho_s = .25$, $\alpha = .02$, (I) $\rho_d = .25$, $\rho_s = .1$, $\alpha = 0$, (J) $\rho_d = .25$, $\rho_s = .1$, $\alpha = .02$, (K) $\rho_d = 0$, $\rho_s = .75$, $\alpha = .02$. These parameters were chosen such that several spheres share each total reflectance ($\rho_s + \rho_d$).

cludes 11 combinations of Ward model parameter settings (Figure 12), chosen such that classification based solely on mean brightness of the observed image produces poor performance.

Each of these two sets of spheres was rendered under Debevec's nine spherical illumination maps. Figure 12 shows spheres rendered under ones of these illuminations. We also rendered all the spheres under a larger set of 100 illuminations, based on high dynamic range imagery acquired by Teller *et al.* [29] in outdoor urban environments and available at *http://city.lcs.mit.edu/data*. Because these illumination maps capture only the upper hemisphere of the environment, we mirrored the upper hemisphere into the lower hemisphere before rendering the spheres (Figure 11). Some ensemble statistics of these illumination maps will be affected by the fact that each map contains two skies, and that each was collected in an urban environment. We therefore expect a classifier trained on a data set rendered under the Teller illuminations to behave somewhat differently from a set trained under the Debevec illuminations, even when the underlying reflectance classes are the same. We believe, however, that both sets of illumination maps preserve many of the typical statistical characteristics of real-world illumination and are therefore worthy of study.

Each sphere was rendered from a view angle midway between a top view and a horizontal view. The resulting renderings were converted from *Radiance*'s native high dynamic range format to floating point images for further processing.

6.2 Evaluation metrics

To test classification accuracy for a particular choice of features using a limited number of available sample images, we performed a variant of leave-one-out cross-validation. Working

	Image set				
	Debevec illuminations (9)		Teller illuminations (100)		Photos (7)
Feature set	6 classes	11 classes	6 classes	11 classes	(9 classes)
Chance (0)	16.7	9.1	16.7	9.1	11.1
Image mean (1)	50.0	19.2	73.5	32.5	42.9
Hand-selected (2)	90.7	68.7	97.0	74.4	50.8
Hand-selected (6)	98.1	83.8	99.7	98.5	93.7
Auto-selected (6)	96.3	66.7	99.3	94.4	74.6
PCA (6)	79.6	61.6	96.0	86.8	71.4
All features (117)	79.6	85.9	96.0	97.9	79.4

Table 1: Cross-validation performance (in % accuracy) for SVM classifiers using different feature sets on all five image sets described in Section 6.1. "Auto-selected" refers to the feature selection method of Section 5.3. Numbers in parentheses indicate the number of features, the number of illuminations, or the number of classes.

with any one of the image sets described in Section 6.1, we left out one illumination at a time. We trained a classifier on the images corresponding to the remaining illuminations and then applied it to test images corresponding to the illumination which had been left out. By leaving out each illumination and repeating this process, we were able to use each image in a set as a test image. Because our feature selection technique produces similar rankings of features when applied to the entire data set or to the data set with one illumination omitted, we used entire data sets for automatic feature selection.

6.3 Performance

Table 1 compares the performance of classifiers trained using various combinations of features on each of the five data sets. As a baseline, the table includes the performance of a completely random classifier, and of a classifier based on only the mean brightness of the observed sphere. The next two lines of the table list performances for feature sets which we hand-selected based on the analysis of Section 4.2 and experimentation with multiple data sets. One of these, illustrated in Figure 7, consists of only two features, but performs significantly better than the classifier based on the image mean. The other consists of six features, namely the mean and tenth percentile of the original unwrapped image, the variance of coefficients in the finest and second-finest radially (vertically) oriented subbands, the ratio of these two variances, and the kurtosis of the second-finest radially oriented subband. A classifier based on this feature set achieves nearly perfect performance (99.7% accuracy) on a set with 100 examples of each reflectance class. As expected, classifier accuracy decreases as the number of classes increases. Accuracy also decreases as the number of illuminations decreases, because the number of available training samples decreases.

We considered a total of 117 features, based on the distributions of image pixel values and of wavelet coefficients at various scales and orientations, as well as power spectral content in several frequency bands and ratios of variances of wavelet coefficients at successive scales. Table 1 shows the performance of classifiers trained using all 117 statistics. The SVM classifiers perform surprisingly well given the small number of training samples — as few

Classification technique	6 features		117 features	
	9 illuminations	100 illuminations	9 illuminations	100 illuminations
SVM	83.8	98.5	85.9	97.9
nearest neighbor	78.8	96.1	67.7	63.6
5-nearest neighbor	75.8	95.9	68.5	84.6

Table 2: Comparison of performances of SVM, nearest neighbor, and k-nearest neighbor (k = 5) classifiers, using the set of six hand-selected features and the complete set of 117 features. Classification accuracies are shown (in %) for synthetically rendered image sets with 11 reflectances, under both the nine Debevec illuminations and the 100 Teller illuminations.

as six per class for the photographic data set.[5] However, the large number of features compared to the number of available samples per class leads to overfitting in the classifier, and performance typically suffers compared to that obtained with a few well-chosen features.

We applied our feature selection method to the synthetic image set consisting of spheres of six reflectances rendered under the nine Debevec illuminations. For comparison, we performed principle components analysis (PCA) on the same data set and selected six linear combinations of features corresponding to the first six principle components.[6] Our feature selection method leads to significantly better performance than PCA, because PCA identifies features which vary significantly from image to image rather than features which allow discrimination of one reflectance class from another. For the image set on which we performed feature selection and for the other set based on the same reflectances (the first and third columns of performance values in Table 1, respectively), our automatic feature selection method led to significantly better results than those obtained using the entire feature set. For image sets based on other reflectance classes, this particular set of features did not perform as well. Features which discriminate well among one set of reflectance classes may not discriminate as well among other reflectance classes. Also, our automatically selected features never performed as well as the set of 6 hand-selected features, showing that the automatic selection technique leaves room for improvement.

Table 2 compares the performances of support vector machines, nearest neighbor classifiers, and k-nearest neighbor classifiers. SVM classifiers outperform the others due to the large number of features and classes compared to the number of available training samples. Nearest neighbor classifiers also suffer from the fact that different features exhibit different variances across the image sets. As expected, the performance difference is larger when the classifiers are trained with the full feature set than when they are trained with only a few selected features.

6.4 Classification failures

The SVM classifier based on the set of six hand-selected features misclassified only four of the 63 images in the set of photographs, outperforming the first author of this paper. Fig-

[5]Because we leave out one illumination at a time in the cross-validation process, the number of available training samples is one less than the number of illuminations.

[6]To improve performance, we normalized each of the selected principal components to have the same variance.

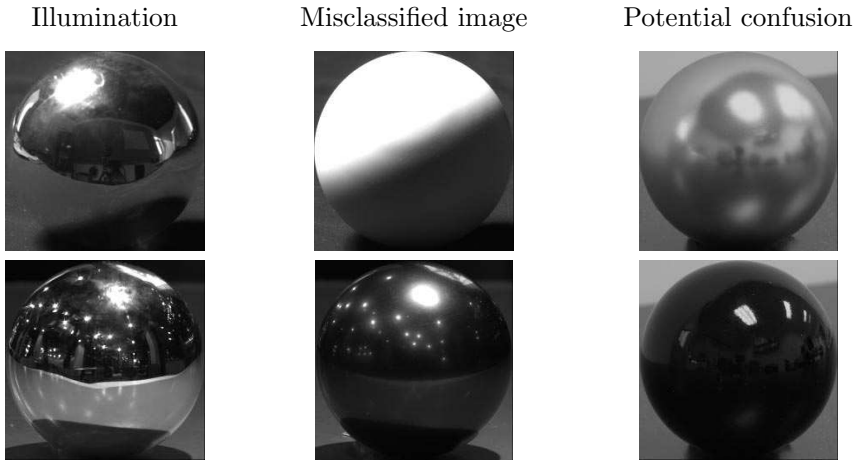


Figure 13: Classification errors for the set of photographs. In each row, the middle column shows a misclassified image. The left-hand column shows an image of a chrome sphere under the same illumination as the misclassified sphere. The right-hand column shows a sphere in the training data set belonging to the incorrectly chosen class; the image shown is the one with minimum Euclidean distance from the misclassified image in the feature space.

ure 13 illustrates two of the four misclassifications. In each row, the middle column shows a misclassified image, the left-hand column shows a chrome sphere under the same illumination, and the right-hand column shows a sphere of a different reflectance that represents a potential source of confusion.

The first row of Figure 13 illustrates the most egregious of the misclassifications. A white matte ball (column c in Figure 1) illuminated by a bright desk lamp is misclassified as a rough metallic ball (column e). This illumination, which accounted for two of the four classification errors, was the only one created by the photographer expressedly for the purpose of collecting these photographs. Images under this point-source-like illumination also proved most difficult for humans to classify, lending credence to the claim that the statistical regularity of "natural" illumination plays an essential role in reflectance estimation. The classifier might perform better if the training set included any illuminations similar to this one.

The second row of Figure 13 shows a more subtle error, in which a bright shiny metallic ball (column e in Figure 1) under a collection of bright incandescent lights is misclassified as a black shiny ball (column d). In this case, the images in the second and third column represent a potential source of confusion even to a human observer.

7 Discussion

Real-world illumination exhibits predictable statistical structure which may facilitate reflectance estimation under unknown illumination, an otherwise ill-posed problem. This paper demonstrates the feasibility of reflectance classification from single monochrome images in unknown real-world scenes. Although our classification algorithm rivals human

performance for test data sets, it leaves a number of open research questions.

First, we have not solved the problem in an explicitly Bayesian manner. While we have analyzed the relationship between illumination statistics, image statistics, and reflectance, we do not translate the results directly to an optimal technique for reflectance classification or reflectance parameter regression. We would like to put our classification method on a more rigorous theoretical foundation.

Second, we believe that significant performance gains may be attained through the use of a more general feature set. The features we use at present do not explicitly model image edges. Features which capture dependencies between wavelet coefficients at neighboring scales and orientations have proven useful in texture modeling [19]. Such features could be incorporated into our classification scheme and feature selection algorithm.

Finally, our method for feature selection may prove more widely applicable to texture classification problems. Our approach bears some similarity to the work of Zhu, Wu, and Mumford [33], who developed a general feature selection method for texture modeling. However, a reflectance or texture classification problem differs from a texture analysis and synthesis problem in two important regards. First, we wish to find not the features which best *characterize* each image, but those which best *distinguish* reflectances from one another. Second, because machine learning techniques such as SVMs require inputs to be a collection of scalar features rather than a collection of probability distributions, we select *features* rather than simply selecting the *filters* whose output they characterize. That is, we select particular summary statistics of distributions of filter outputs, rather than entire histograms. One could generalize our approach by fitting more general parametric forms to the marginal feature distributions, or by replacing the affine regression with nonlinear regression.

Acknowledgments

Marc Talusan and Rebecca Loh carried out the photography of the spheres analyzed in Section 6. Seth Teller, Neel Master, and Michael Bosse shared the data set from the MIT City Scanning Project and helped us use it to construct illumination maps. This work was supported by an NDSEG fellowship to R.O.D., by NIH Grant EY11005-04 to E.H.A., by a grant from NTT to the MIT Artificial Intelligence Lab, by a contract with Unilever Research, and by ONR Grant N00014-00-1-0089 and AFOSR Grant F49620-00-0362 to A.S.W.

References

[1] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge, MA, 1999.

[2] J. R. Boyack and A. K. Juenger. Brightness adjustment of images using digital scene analysis. U.S. Patent US5724456, March 1998.

[3] R. Collobert and S. Bengio. Support vector machines for large-scale regression problems. Technical Report IDIAP-RR 00-17, IDIAP, Martigny, Switzerland, 2000.

[4] J. S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. *Computer Graphics (SIGGRAPH)*, pages 361–368, 1997.

[5] P. 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.

[6] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Computer Graphics (SIGGRAPH)*, 1998.

[7] R. O. Dror, E. H. Adelson, and A. S. Willsky. Estimating surface reflectance properties from images under unknown illumination. In *Proceedings of SPIE 4299: Human Vision and Electronic Imaging VI*, pages 231–242, San Jose, 2001.

[8] R. O. Dror, E. H. Adelson, and A. S. Willsky. Surface reflectance estimation from a single image under unknown real-world illumination. In *Proceedings of CVPR Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation*, 2001.

[9] R. O. Dror, T. Leung, A. S. Willsky, and E. H. Adelson. Statistics of real-world illumination. In *Proceedings of CVPR*, 2001.

[10] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1):1–50, 2000.

[11] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–94, 1987.

[12] R.T. Fleming, R. O. Dror, and E. H. Adelson. How do humans determine reflectance properties under unknown illumination? In *Proceedings of CVPR Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation*, 2001.

[13] Andrew S. Glassner. *Principles of Digital Image Synthesis*, volume 2. Morgan Kaufmann, San Francisco, 1995.

[14] D. J. Heeger and J. R. Bergen. Pyramid-based texure analysis/synthesis. *Computer Graphics (SIGGRAPH)*, 1995.

[15] J. Huang and D. Mumford. Statistics of natural images and models. In *CVPR*, pages 541–47, 1999.

[16] G. W. Larson and R. Shakespeare. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaugmann, San Francisco, 1998.

[17] S. R. Marschner. *Inverse Rendering for Computer Graphics*. PhD thesis, Cornell University, Ithaca, NY, 1998.

[18] Bui-Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–7, 1975.

[19] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, pages 49–71, 2000.

[20] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Adaptive wiener denoising using a gaussian scale mixture model in the wavelet domain. In *ICIP*, 2001.

[21] R. Ramamoorthi and P. Hanrahan. An efficient representation for environment irradiance maps. *Computer Graphics (SIGGRAPH)*, 2001.

[22] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. *Computer Graphics (SIGGRAPH)*, 2001.

[23] D. L. Ruderman. The statistics of natural images. *Network: Comput. Neural Syst.*, 5:517–48, 1994.

[24] Y. Sato, M. D. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. In *Computer Graphics (SIGGRAPH)*, 1997.

[25] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. *Computer Graphics (SIGGRAPH)*, 1995.

[26] S. Shafer. Using color to separate reflection components. *Color Research and Application*, 10:210–218, 1985.

[27] E. P. Simoncelli and E. H. Adelson. Subband transforms. In John W Woods, editor, *Subband Image Coding*, chapter 4, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.

[28] E.P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc SPIE, 44th Annual Meeting*, volume 3813, Denver, July 1999.

[29] S. Teller, M. Antone, M. Bosse, S. Coorg, M. Jethwa, and N. Master. Calibrated, registered images of an extended urban area. *IJCV*, submitted.

[30] S. Tominaga and N. Tanaka. Estimating reflection parameters from a single color image. *IEEE Computer Graphics and Applications*, 20:58–66, September/October 2000.

[31] G. J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH)*, 26(2):265–72, 1992.

[32] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 215–24, Los Angeles, 1999.

[33] S. C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME) – towards a unified theory for texture modeling. *IJCV*, 27:1–20, 1998.