# Computational Luminance Constancy from Naturalistic Images

**Vijay Singh** Computational Neuroscience Initiative, Department of Physics, University of Pennsylvania, PA, USA

**Nicolas P. Cottaris** Department of Psychology, University of Pennsylvania, PA, USA

**Benjamin S. Heasly** Department of Psychology, University of Pennsylvania, PA, USA

**David H. Brainard** Department of Psychology, University of Pennsylvania, PA, USA

**Johannes Burge** Neuroscience Graduate Group, Department of Psychology, University of Pennsylvania, PA, USA

The human visual system supports stable percepts of object color even though the light that reflects from object surfaces varies significantly with the scene illumination. To understand the computations that support stable color perception, we study how estimating a target object's light reflectance value (LRV; a measure of the light reflected from the object under a standard illuminant) depends on variation in key properties of naturalistic scenes. Specifically, we study how variation in target object reflectance, illumination spectra, and the reflectance of background objects in a scene impact estimation of a target object's LRV. To do this, we applied supervised statistical learning methods to the simulated excitations of human cone photoreceptors, obtained from labeled naturalistic images. The naturalistic images were rendered with computer graphics, with the underlying scene descriptions generated stochastically using statistical models of natural spectral variation. Optimally decoding target object LRV from the responses of a small learned set of task-specific linear receptive fields that operate on a contrast representation of the cone excitations yields estimates that are within 13% of the correct LRV. Our work provides a framework for evaluating how different sources of scene variability limit performance on luminance constancy.

Keywords: color constancy, luminance constancy, supervised learning

## Introduction

The perceived color of an object has important behavioral implications because color helps to identify objects and their surface properties (Mollon, 1989; Jacobs, 1981). The computational challenge underlying object color perception is that the

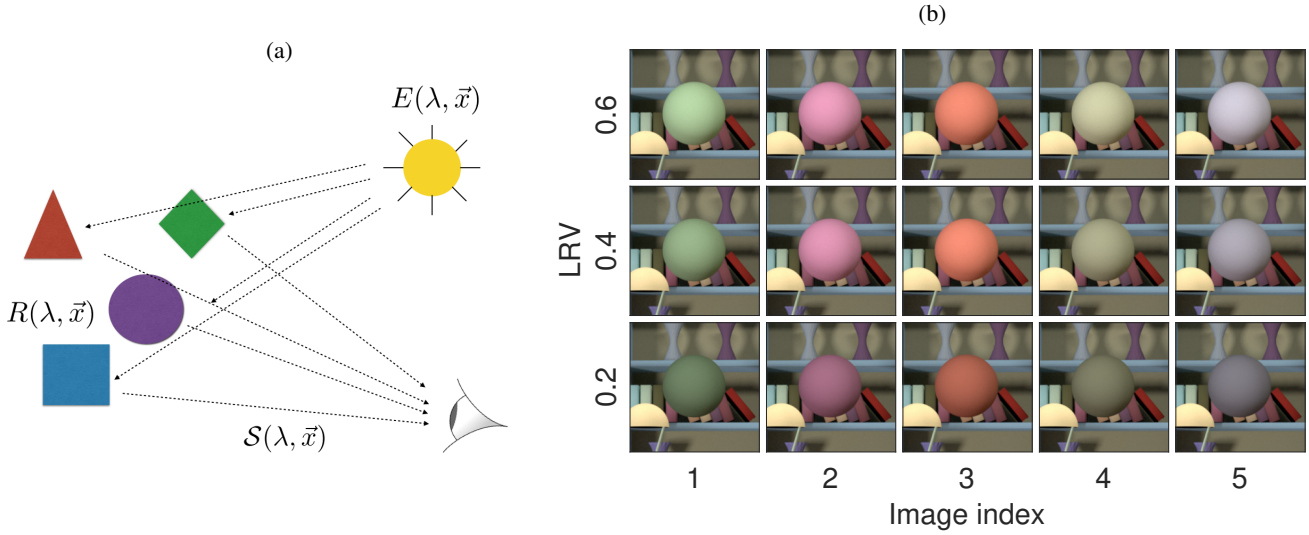


Figure 1: **Color and luminance constancy:** (a) The light reflected from an object to the eye depends both on the surface reflectance of the object and on the scene illumination. The reflected light also depends on geometric factors, such as the object's shape, pose, and position. The human visual system accounts for variations in reflected light due to object-extrinsic factors and produces a percept of object color that is relatively stable across such variations. (b) Images of a sphere under a fixed illuminant. Down each column, the sphere's light reflectance value (LRV) varies, but its relative reflectance spectrum (i.e. its spectral shape) is held fixed. Across each row, the relative reflectance spectrum varies, but the LRV is held fixed. We cast the problem of computational luminance constancy to be that of estimating the LRV of a target object from the image, across variation in other scene factors. Specifically, we study variation in target object relative surface reflectance, variation in illumination spectrum, and variation in the surface reflectance of the other objects in the scene.

light reflected from an object depends not just on the object's surface reflectance, but also on object-extrinsic factors such as the scene illumination, the pose of the object, and the viewpoint of the observer in the scene (Fig. 1a). To compute a perceptual representation of object color that is correlated with the object's physical surface reflectance, the visual system must account for these object-extrinsic factors. The ability of a visual system to compute a representation of object color that is stable against variation in object-extrinsic factors is called color constancy. Although human color constancy is not perfect, it is often very good (Foster, 2011; Brainard & Radonjic, 2014).

Early work on computational color constancy considered a simplified imaging model where scenes consisted of multiple flat matte objects and a single spatially diffuse illuminant (Land, 1977; Buchsbaum, 1980; Maloney & Wandell, 1986). Subsequent work incorporated probabilistic descriptions of the statistics of naturally occurring scenes (D'Zmura & Iverson, 1994; D'Zmura, Iverson, & Singer, 1995; Brainard & Freeman, 1997). A key insight from these computational studies is that stable color descriptors of a target object cannot be obtained solely from the light reflected from that target object. It is possible, however, to obtain stable color descriptors by jointly analyzing the light reflected from multiple objects in the scene. As a consequence, color constancy will be affected both by variation in the illumination and by variation in the surface reflectance of other objects in the scene. Color constancy algorithms should therefore be evaluated with respect to variation in both of these factors (?, ?; Brainard & Freeman, 1997).

Recently, supervised statistical learning methods have been applied to the computational problem of color constancy. These methods use the joint statistics of labeled images and scene variables to learn mappings that extract stable surface color descriptors from image input (Barron, 2015). This general approach has been useful for other perceptual problems in early and mid-level vision. For example, a recently developed technique called accuracy maximization analysis (AMA) learns linear receptive fields optimized for particular perceptual tasks (Geisler, Najemnik, & Ing, 2009; Burge & Jaini, 2017; Jaini & Burge, 2017). AMA has been used to develop ideal observers for estimating defocus blur, binocular disparity, and retinal motion estimation with natural stimuli. These ideal observers provide excellent predictions of human performance (Burge & Geisler, 2011, 2012, 2014, 2015).

Supervised learning requires large labeled datasets. Such datasets are not readily available for the study of color constancy. Although there are datasets of calibrated color images, these do not provide ground truth information about surface reflectance and illumination at each image location (Chakrabarti & Zickler, 2011; Nascimento, Amano, & Foster, 2016; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; Tkacik et al., 2011; Skauli & Farrell, 2013; Olmos & Kingdom, 2004). There are a few databases consisting of images of posed scenes where surface reflectances of individual objects have been measured (Funt & Drew, 1988; Ciurea & Funt, 2003), but these databases are not large enough to drive supervised learning. Finally, recent work has applied crowd-sourcing techniques to annotate natural images with surface reflectance descriptors (Bell, Bala, & Snavely, 2014).

In this paper, we use high-quality computer graphics to generate large datasets of naturalistic images where the surface reflectance corresponding to each image pixel is known. This approach allows us to investigate computational color constancy with naturalistic stimuli, while retaining the ability to control the properties of objects and illuminants. We tackle the specific problem of luminance constancy, a constitutive component of the more general color constancy problem (Fig. 1b).

We define the computational problem of luminance constancy as that of estimating the light reflectance value (LRV) of a target object's surface reflectance function. The LRV is a measure of the overall amount of light reflected by a surface (ASTM, 2017). Estimating the LRV from a surface reflectance function proceeds in two steps. First, one computes the luminance of the light that would be reflected from the surface under a reference illuminant. Second, one normalizes the result by the luminance of the reference illuminant itself. Here we use CIE daylight D65 as the reference illuminant and the CIE 1931 photopic luminosity function to compute luminance (CIE, 1986). Sometimes LRVs are expressed as a percentage. Here we express LRVs as proportions, so that they take on values between 0 and 1.

# Methods

## Overview

There are four key steps to our methods. First, we generate a labeled set of training images. Second, we use a model of the early visual system to compute the responses of the cone photoreceptor mosaic to the labeled images. Third, we apply AMA to the cone responses, as well as to a contrast normalized version of the cone responses, to learn task-optimal receptive fields (RFs). Fourth, we evaluate how well the responses of the RFs may be decoded to achieve luminance constancy.

## Labeled training data

### *Virtual naturalistic scenes*

The light that reflects from objects to the eyes depends on many factors. These include the surface reflectance, texture, material and geometry of the objects, the spectral power distribution of the illuminants, and the position of the observer. We have developed a rendering package (https://github.com/BrainardLab/VirtualWorldColorConstancy) that allows us to construct

(a) Library (b) Mill (c) Table-Chairs

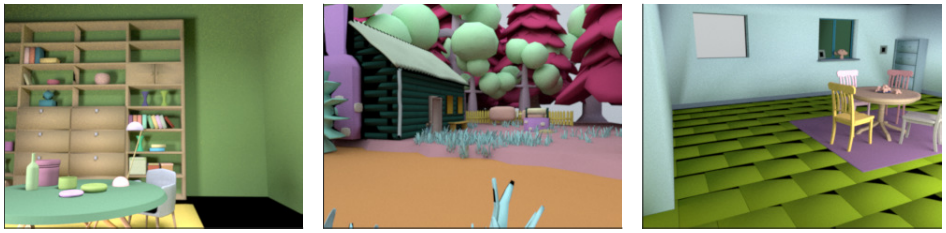

Figure 2: **Base scenes.** Each image shows a rendering of one base scene without additional inserted objects. The reflectance spectra of the distinct surfaces and the spectral power distributions of the light sources in each scene have been assigned randomly from statistical models of naturally occurring spectra. The images have been scaled and tone mapped individually for illustrative purposes.

.

(a) Big-ball (b) Xylophone (c) Barrel (d) Ring toy (e) Bottle

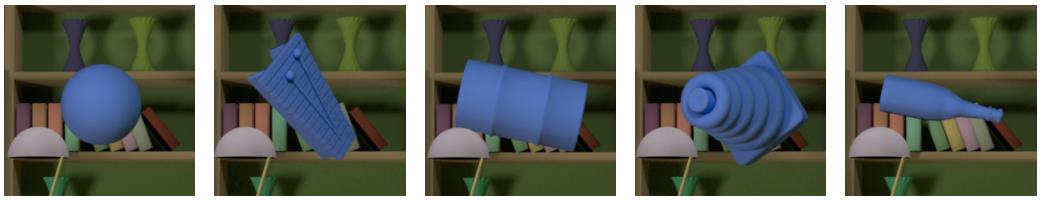

Figure 3: **Library base scene with inserted objects.** The rendering package can be used to insert objects into base scenes. Each image shows a different object inserted into the Library base scene. The rendering viewpoint was set so that the object was at the center of the images. The full rendered images have been cropped so that the inserted objects are easily visible.

models of naturalistic scenes, with key scene factors under programmatic control. The package builds on the RenderToolbox4 package (http://rendertoolbox.org) (Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014), and harnesses the open-source computer graphics renderer Mitsuba (https://www.mitsuba-renderer.org, Jakob, 2010) to produce physically-accurate images from the scene models. Because the images are rendered from known scene models, each image pixel can be labeled with the surface reflectance of the corresponding object. By using statistical models of daylight illumination and object surface reflectance to guide the rendering of scenes, the package allows us to produce large labeled sets of images that capture key aspects of the task-relevant statistical structure of natural scenes.

The RenderToolbox4 package includes a collection of base scenes (Fig. 2). Base scenes specify an arrangement of objects and light sources. Base scenes may be enriched by the insertion of additional objects, chosen from an object library (Fig. 3), and by the insertion of additional light sources. Once the position, size, and pose of the inserted objects and light sources have been set, a surface reflectance function can be assigned to each object in the scene, and a spectral power distribution function can be assigned to each light source (Fig. 4). The resulting scene model can then be rendered from any specified viewpoint.

In the present work, we used our package to generate datasets of naturalistic scenes and corresponding images. We used one fixed base scene and inserted a spherical target object of fixed size into this scene. We generated three distinct datasets, which

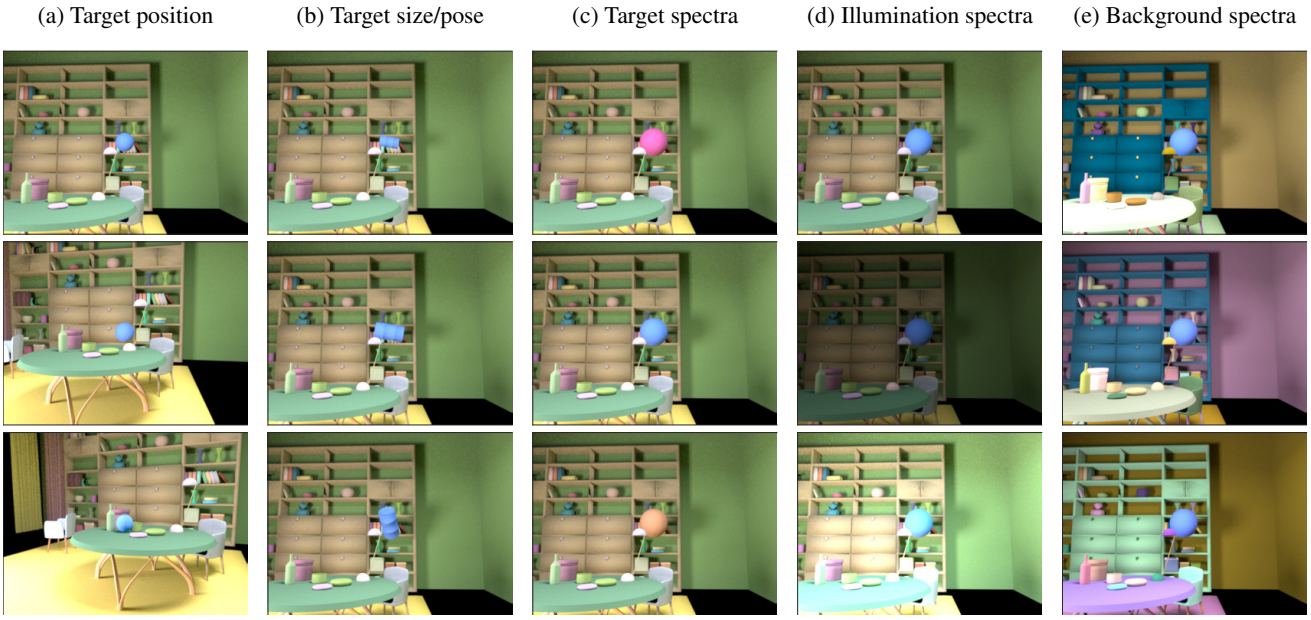


Figure 4: **Scene transformations.** The properties of a visual scene can be broadly classified into two groups: geometrical (a-b) and spectral (c-e). Our rendering package provides control over these properties as illustrated by the columns of the figure. (a) Variation in inserted object positions. (b) Variation in object size and pose. (c) Variation in the surface reflectance of the target object. (d) Variation in the spectral power distributions of the light sources. (e) Variation in surface reflectance of the background objects. All images in this figure are rendered with common scaling and tone mapping.

we refer to as Conditions 1-3 (Fig. 5). These are described below. Across these datasets the luminance constancy problem (i.e. estimating the target object LRV) becomes progressively more difficult.

The variation within our datasets captures the essence of the computational problem of lightness constancy up to effects of scene geometry, an additional richness that we do not address in this paper (see Funt & Drew, 1988; Barron & Malik, 2012).

In Condition 1, the relative reflectance spectrum of the target object was the only source of scene variation other than the target object LRV. The reflectance spectra of the background objects and the spectral power distribution of the light sources were held fixed. In Condition 2, the spectral power distribution of the light sources varied in addition to the relative reflectance spectrum of the target object. Finally, in Condition 3 three distinct scene factors–target object relative reflectance spectrum, spectral power distribution of the light sources, and reflectance spectra of the background objects –varied in addition to the target object LRV. The datasets for Conditions 1 and 2 consisted of 1000 scenes and images, 100 for each of 10 target object LRV values. The dataset for Condition 3 consisted of 3000 scenes and images, 300 each for the 10 target LRV values. The LRV values were equally spaced between 0.2 and 0.6. For each LRV value, we generated a different relative target object surface reflectance for each scene.

We used the Library base scene and a spherical target object. The Library base scene contains 2 area lights. We inserted one additional spherical light source into the scene. The position and size of the inserted object, the inserted light source, and the viewpoint on the scene were held fixed across all scenes. The geometry is shown in Figure 9a below. Surface and illuminant spectra were sampled according to statistical models of naturally occurring spectra. Multispectral images ($320 \times 240$ pixels)

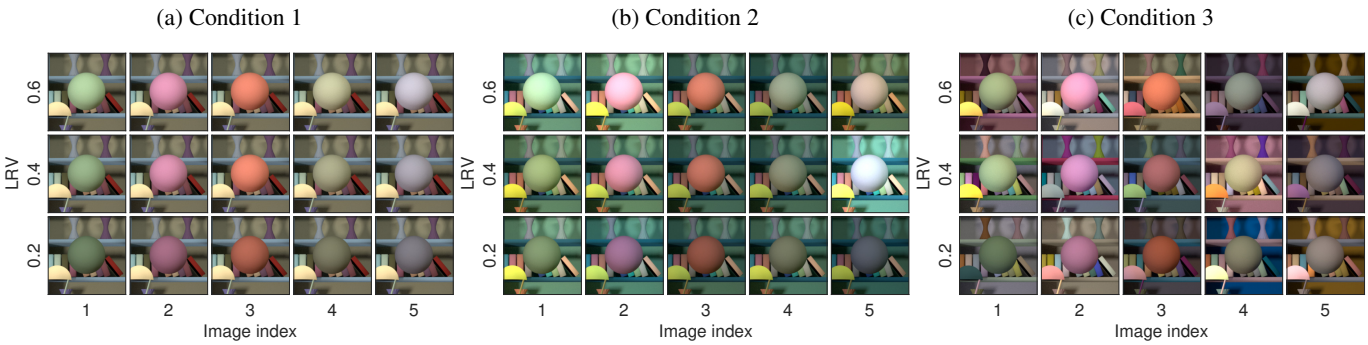


Figure 5: **Conditions 1-3:** Illustrative images for the three types of spectral variations studied in this paper. (a) Condition 1: variable target object relative reflectance spectrum, fixed light source spectra, fixed background object spectra (same as Fig. 1b). (b) Condition 2: variable target object relative reflectance spectrum, variable light source spectra, fixed background object spectra. (c) Condition 3: variable target object relative reflectance spectrum, variable light source spectra, variable background object spectra. The spheres in each row of each panel have the same LRV, while the spheres in each column of each panel have the same relative surface reflectance spectrum. Across panels a-c, spheres in corresponding locations have the same surface reflectance. In all three panels, the overall light source spectra scale factors (see Methods) were drawn from a uniform distribution on the range [0.5, 1.5]. This is smaller than the variation we studied, but allows us to show the variation within the dynamic range available for display in publication. All images in this figure are rendered with common scaling and tone mapping.

were rendered at 31 evenly-spaced wavelengths between 400nm and 700nm.

### *Illumination spectra*

To generate illumination spectra, we developed a statistical model of the Granada daylight measurements (Hernández-Andrés, Romero, Nieves, & Lee, 2001; http://colorimaginglab.ugr.es/pages/Data) and then sampled randomly from the model. To match the wavelength spacing we use in rendering, we resampled the wavelength spacing of the Granada spectra to 31 evenly-spaced wavelengths between 400 nm and 700 nm.

Our statistical model approximates normalized spectra from the Granada dataset using the first 6 principle components of the normalized measurements from that dataset. The variation in the weights on these components is then modeled with a 6-dimensional Gaussian. The mean and covariance matrix of the Gaussian match the sample mean and covariance of the weights (Brainard & Freeman, 1997). To generate a random relative illuminant spectrum, we sample weights randomly from this Gaussian and then perform a weighted sum of the components to generate the corresponding spectrum. Figure 6b illustrates the relative spectra of draws obtained using this procedure. Chromaticities and rendered color renditions of these draws are shown in Figures 6c and 6d. Details of the statistical model for illumination are provided in the appendix.

The multivariate Gaussian model described above is based on normalized spectra. Thus, it does not embody the large variation in overall intensity of natural daylights. To restore the overall intensity variation in the Granada dataset, which spans approximately three orders of magnitude, we scaled each randomly generated spectrum with a random number sampled from a log-uniform distribution spanning three orders of magnitude. The particular range of scale factors was chosen so that the mean number of cone isomerizations from the target object ranged from approximately 100 to 100,000, given a 100 ms cone integration time (see below). This range of cone isomerizations corresponds approximately to that of natural daylight viewing.

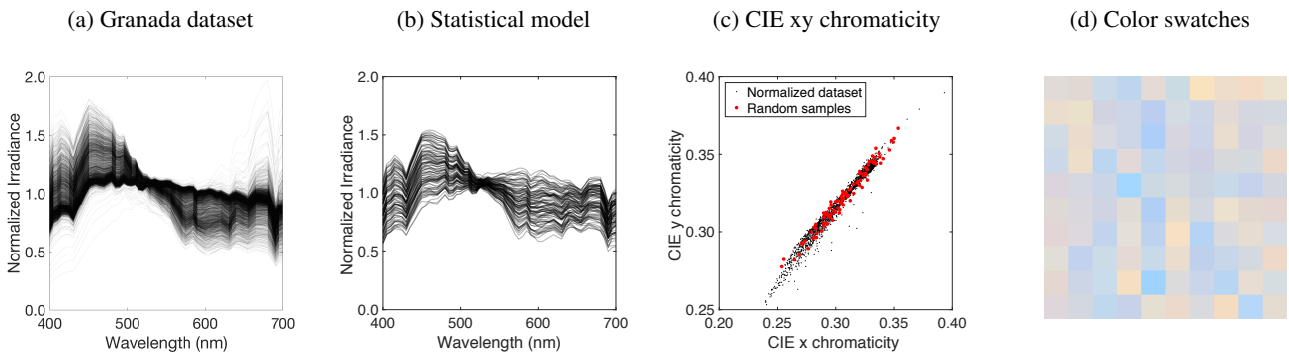


Figure 6: **Statistical model of illumination spectra:** (a) Normalized Granada dataset. Each spectrum is normalized by its mean power. (b) Sample spectra generated using the statistical model derived from normalized Granada dataset. (c) CIE xy chromaticities of the Granda dataset (black) and the samples shown in panel b (red). (d) sRGB renditions of the samples shown in panel b.

### *Surface reflectance spectra*

To generate random reflectance spectra, we employed principles similar to those used to generate random illuminant spectral power distributions. We used measurements from the Munsell (Kelly, Gibson, & Nickerson, 1943) and Vrhel (Vrhel, Gershon, & Iwan, 1994) surface reflectance datasets to create a dataset containing 632 reflectance spectra (462 from the Munsell data and 170 from the Vrhel data) (Figure 7a). We resampled the spectra to 31 evenly-spaced wavelengths between 400 nm and 700 nm. Figures 7b, 7c, and 7d illustrate reflectance sampled obtained using the model. Details of the statistical model for surface reflectance spectra are provided in the appendix.

To generate the target object reflectance at a particular LRV, the sampled reflectance spectrum was scaled such that its LRV had the desired value (see appendix). Figure 8 shows color renderings of target reflectance spectra under CIE illuminant D65, for the evenly spaced LRV values we studied.

## Model of early visual system

Light entering the eye is focussed and blurred by eye's optics to form the retinal image. This image is then sampled by a mosaic of cone photoreceptors. The excitations of these photoreceptors provide the information available to the neural visual system for further processing. Because we are interested in how well luminance constancy may be achieved by the human visual system, we simulated the cone excitations to our scenes using a model of the early visual system.

We focussed our analysis on image regions local to the target by cropping the rendered images to $1 \times 1$ degrees of visual angle around the target object ($51 \times 51$ pixels; Figure 9b). The local analysis is motivated by the fact that neural receptive fields early in the visual pathways (e.g., retina, primary visual cortex) pool information locally. In primary visual cortex, foveal receptive fields have a maximum spatial extent of approximately 1 degree of visual angle (Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988).

We modeled a visual system having typical optical blurring (including axial chromatic aberration)(Marimont & Wandell, 1994), and typical spatial sampling with an interleaved rectangular mosaic of long (L), middle (M) and short (S) wavelength-sensitive cones (Brainard, 2015; Fig. 9c). The cone mosaic contained L:M:S cones in the ratio 0.6:0.3:0.1 and with spectral

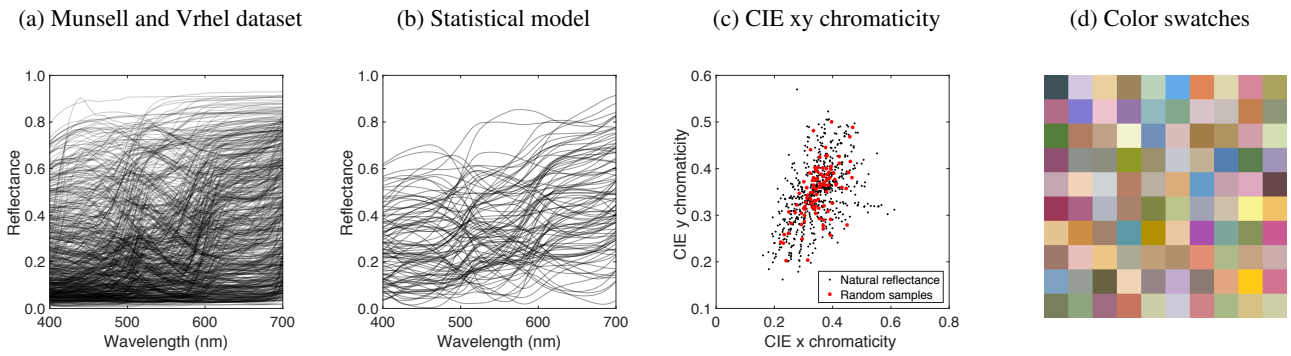


Figure 7: **Statistical model of surface reflectance:** (a) Surface reflectance spectra from the Munsell and Vrhel datasets. (b) Sample spectra generated using the surface reflectance statistical model. (c) CIE xy chromaticity diagram of the Munsell and Vrhel reflectances (black) and the sample spectra shown in panel b (red). (d) sRGB renditions of the samples shown in panel b, rendered under the CIE D65 daylight spectrum.

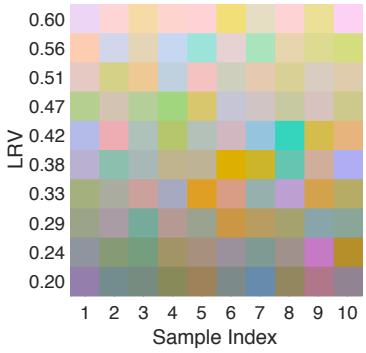


Figure 8: **Target object surface reflectance spectra:** sRGB renditions of the target object surface reflectance rendered under CIE D65 daylight spectrum. The figure shows 10 random samples at 10 equally spaced LRV levels in the range [0.2, 0.6]. Each row contains 10 random samples of reflectance spectra generated at the LRV shown on the left.

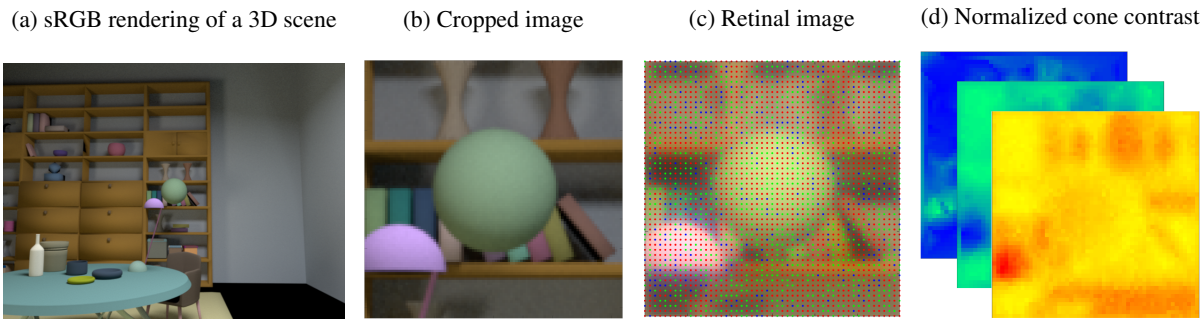


Figure 9: **Pipeline for generating labeled datasets:** The labeled images used in this paper are generated as follows: (a) A 3D virtual scene, containing a target object is created. The rendering viewpoint is selected so that the target object is in the center of the image. Spectra are assigned to the target object, illuminants, and other objects in the scene using the statistical models described in the text. A multispectral image of the scene is then rendered. (b) The central portion of the rendered image is cropped around the target object. (c) The retinal image corresponding to the rendered image is computed, which is then used to compute the number of photopigment isomerizations in the cone photoreceptors. The figure shows a rendering of the retinal image after optical blurring, with the location and identity of the cones indicated by an overlay (L cones: red, M cones: green, S cones: blue). (d) The cone excitations are linearly interpolated to estimate the responses of all the three types of cone at each location (demosaicing). Finally, the demosaiced images are contrast normalized.

sensitivities derived from the CIE physiological standard (CIE, 1986). Cone excitations were taken as the number of photopigment isomerizations, assuming an integration time of 100 msec. The Poisson nature of the photopigment isomerization was also included (Hecht, Shlaer, & Pirenne, 1942). This modeling was implemented using the software infrastructure provided by ISETBio (https://isetbio.org).

To put the computed cone excitations into a form for further analysis, we demosaiced the excitations for each cone class using linear interpolation to obtain a cone excitation image for each cone class. The lens and macular pigment absorb more short than long wavelength light, so that the excitations of the S cones tend to be smaller than those of the L and M cones. To make the magnitude of the three cone excitation images similar across cone classes, we normalized each cone excitation image by the summed (over wavelength) quantal efficiency of the corresponding cone class. The demosaicing and normalization do not alter the available information relative to that carried by the cone excitations themselves.

To capture key properties of post-receptoral processing, such as contrast normalization (Heeger, 1992; Albrecht & Geisler, 1991; Carandini & Heeger, 2012), we processed the cone excitation images as follows: We first converted each cone excitation image to a corresponding cone contrast image. This was accomplished by computing the mean response over the three cone excitation images, and then subtracting off and dividing by this mean. To model contrast normalization, we then divided the contrast images by the sum of squared contrasts taken over image pixels and cone classes.

## Computational luminance constancy

We used our datasets to determine how well target object LRV can be estimated from cone excitations and from normalized cone contrasts. Studying both representations allows us to understand how early contrast coding and normalization affect luminance constancy. We applied accuracy maximization analysis (AMA) to learn the optimal receptive fields for estimating LRV,

and evaluated the performance obtained when the responses of these receptive fields are optimally decoded.

### *Learning optimal receptive fields*

AMA is a task-specific Bayesian method for dimensionality reduction. When provided with a labeled training set, a receptive field response model, a decoder that uses these responses to estimate the stimulus label, and a cost function, AMA returns a set of N linear receptive fields, where N is a parameter of the procedure. These receptive fields are chosen so that decoding their responses leads to minimum expected-cost estimation, relative to any other choice of N linear receptive fields. This is possible because given a training set together with a set of N linear receptive fields and an explicit cost function, it is possible to determine the minimum expected-cost estimator that maps receptive field responses to the stimulus labels. Moreover, a close approximation to the corresponding expected estimation cost may be explicitly computed. Thus, AMA searches over the space of linear receptive fields to find the set that minimizes the expected estimation cost. For each condition, we learned the receptive fields with a training set consisting of 90% of the images.

In our implementation of AMA, we explored both the Kullback-Leibler divergence cost function (corresponding to the maximum a posteriori estimator) and the mean squared error cost function (corresponding to the posterior mean estimator) and assumed that receptive field responses were corrupted by scaled Gaussian noise (i.e. Poisson-like noise with a fano factor of 1.3). Training with both cost functions yielded similar estimation performance; the results reported here are for the Kullback-Leibler divergence cost function. We chose this cost function because the estimates it yields at the boundaries of the LRV range are less biased than those obtained with the mean squared error cost function.

Details of how AMA learns receptive fields and how the receptive field responses are optimally decoded are provided in previously published work (Geisler et al., 2009; Burge & Jaini, 2017; Jaini & Burge, 2017).

### *Decoding optimal receptive fields*

Once the AMA receptive fields have been learned, a general decoder is needed that can be used with arbitrary test images. A general decoder is necessary because the decoder used to learn the receptive fields can be used only with the training set. Specifically, the decoder used to learn the receptive fields requires the response mean and response variance of each receptive field to every labelled stimulus in the training set (Geisler et al., 2009; Burge & Jaini, 2017). To proceed, first we use the AMA receptive fields and the training dataset to find the distributions of the receptive field responses conditional on the target object LRVs. Then, we model these conditional response distributions with multivariate Gaussians. To estimate the LRV of the target object in a test image, we use Bayes' rule together with the Gaussian conditional distributions and a uniform prior to obtain the posterior distribution over LRV values, given the receptive field responses. The optimal estimate is the LRV value that minimizes the expected value of the cost function, where the expectation is taken over the posterior. For the results we report, we again used the Kullback-Leibler divergence cost function.

The Gaussian approximations for the posterior were estimated using the same training set that was used to learn the receptive fields. Estimation performance was evaluated on a test set consisting of the 10% of images in the dataset that were not in the training set. For our primary results, estimation was based on six receptive fields.

### *Baseline methods*

To provide baselines for evaluation of the estimation method described above, we used linear regression. First, we solved for the weights on the average L, M and S cone excitations corresponding to the target object that best predicted the target LRV values. We took the average cone responses from a $3 \times 3$ pixel region at the center of the target object. This baseline method

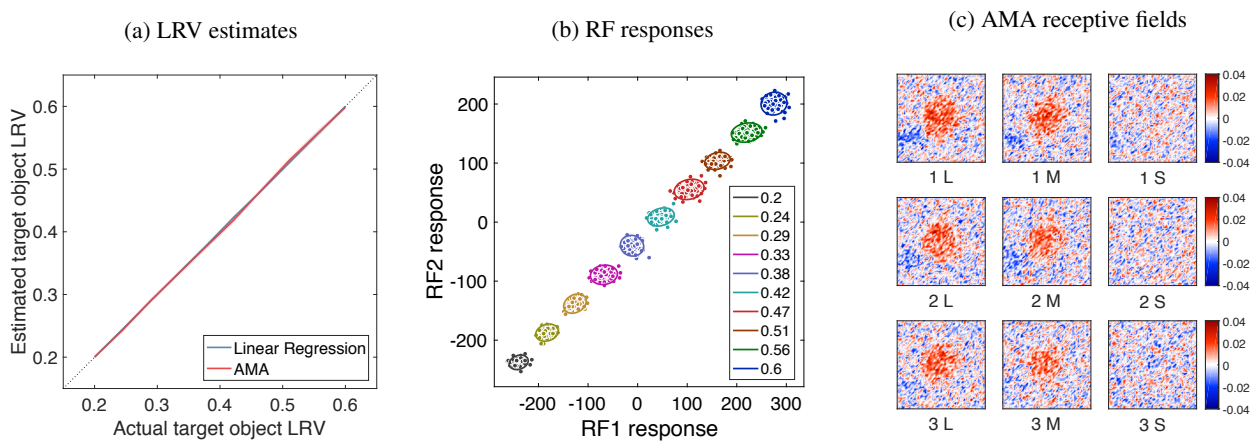


Figure 10: **Condition 1 results:** (a) Mean estimated LRV $\pm$ 1 standard deviation for images in the test set obtained using the AMA-based estimator and linear regression on the cone excitations. Solid lines show the mean estimate, the filled region in lighter color shows $\pm$ 1 standard deviation. The diagonal identity line (dotted blue) indicates veridical estimation. The standard deviations are too small to be visible. Relative RMSE (see Fig.14) is 0.7% for linear regression and 1.1% for AMA-based estimation. (b) Training image cone excitations in Condition 1 projected onto the first two AMA receptive fields. Each cluster of responses represents the image patches associated with a particular LRV. The response clusters are approximated by a multivariate Gaussians whose means and covariances are represented by the ellipses shown in the figure. (c) The first three AMA receptive fields.

disregards information carried by the image regions corresponding to the background objects. We also performed regression on the contrast normalized version of the cone excitation images. This second baseline method indirectly incorporates information carried in the image regions corresponding to the background objects, through their effect on the contrast normalized responses at the center of the target object. We also evaluated the performance of a naive model that uses the mean LRV value in the training set as the estimate, irrespective of the image data; that is, the naive model always estimated the LRV to be equal to 0.4.

# Results

## Condition 1

In Condition 1, only the LRV and the relative reflectance spectrum of the target object vary across scenes. We used accuracy maximization analysis (AMA) to learn a set of linear receptive fields that are optimal for estimating target LRV from the cone excitations in this condition. Decoding performance is shown in Figure 10a. Figure 10a also shows the performance of the baseline linear regression method. Both methods are trained on 90% of the images in the dataset and tested on the remaining 10%. The LRV estimates obtained for both methods are essentially perfect.

Figure 10b shows the responses of the first two AMA receptive fields to all the image patches in the training set. Each individual point represents the receptive field responses to an individual image patch and is color coded by target object LRV. The responses segregate according to the target LRV, which supports the high level of estimation performance shown in Figure 10a. Note that even the responses of the first RF separate well (vertical lines would accurately distinguish the LRVs), so that that one receptive field is sufficient for accurately performing the task.

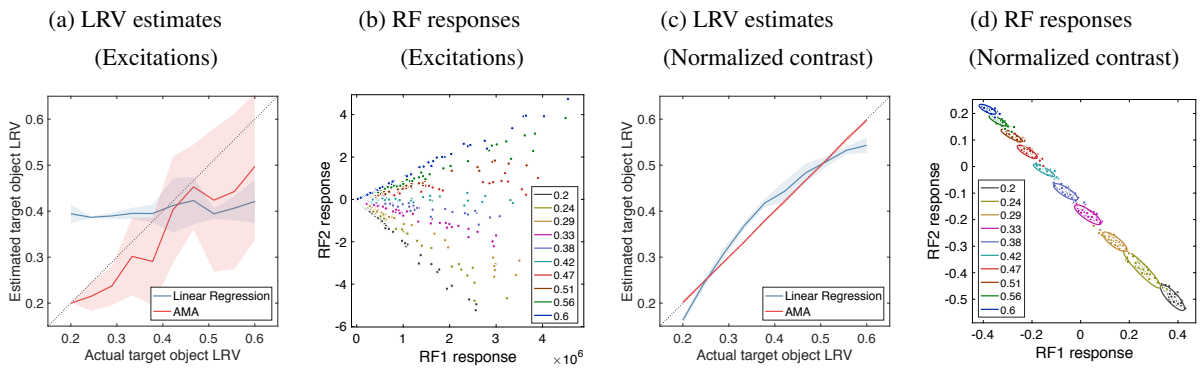


Figure 11: **Condition 2 results:** (a) Mean estimated LRV $\pm$ 1 standard deviation for images in the test set obtained using the AMA-based estimator and linear regression from the cone excitations. The relative RMSE is high (41.4% for linear regression and 26.4% for AMA-based estimation). (b) Cone excitations projected onto AMA receptive fields. (c) Mean estimated LRV $\pm$ 1 standard deviation for images in the test set obtained using the AMA-based estimator and linear regression from the contrast normalized cone excitations. The relative RMSE of estimation is 9.3% for linear regression and 1.2% for the AMA-based estimator. (d) The contrast normalized cone excitations projected along the first two AMA receptive fields. The receptive fields were learned using AMA and the contrast normalized cone excitations. For Condition 2, the response corresponding to individual image patches at each LRV level separates much better when the input is the contrast normalized cone excitations than when the input is the raw cone excitations.

Figure 10c shows the first three AMA receptive fields. Each receptive field performs a weighted sum of the L, M, and S cone excitations. The L and M cone excitations at the target object location receive large weights. The L and M cone excitations at the background object locations and the S cone excitations at all locations receive small weights and are thus largely ignored. For this condition, the background regions provide very little task-relevant information. In addition, luminance is primarily determined by L and M cones, so the S cones should make little direct contribution. The first three RFs are similar to each other, reflecting the fact that here the main benefit of multiple receptive fields is to reduce the impact of noise on estimation performance.

The results in Figure 10 were obtained for one particular draw from our statistical model of illumination. We verified that performance remained excellent for other draws.

## Condition 2

Condition 2 includes variation in the spectral power distribution of the light sources, in addition to the variation present in Condition 1. This illumination variation makes the task more difficult because it causes variation in the cone excitations that is not due to target object LRV. We learned AMA receptive fields on the cone excitations to the images in this new condition, and again evaluated decoding performance (Fig. 11(a,b)). Performance is poor. Indeed, the estimates deviate considerably from the true LRV, as seen by the fact that the mean estimate (red line) does not lie on the positive diagonal and by the fact that there is large estimation variability for each value of the true LRV. For this case, linear regression also does poorly. Indeed, the linear regression estimates are essentially the same as would be obtained by simply guessing the mean LRV of the training set (0.4).

Recall that there are two qualitatively distinct factors contributing to the variation in illumination spectra: changes in overall intensity, and changes in the relative spectral power distribution. To separate the effect of these two factors, we trained and

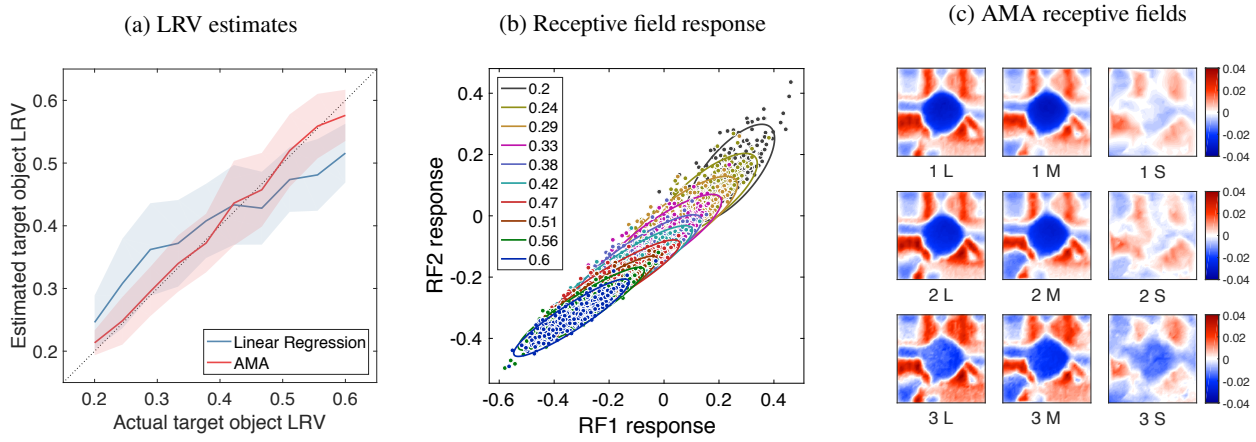


Figure 12: **Condition 3 results:** (a) Mean estimated LRV $\pm$ 1 standard deviation for images in the test set obtained using the AMA-based estimator and linear regression from the contrast normalized cone excitations. The relative RMSE is 23.9% for linear regression and 12.6% for AMA. (b) Contrast normalized cone excitations for the training set projected onto the first two AMA receptive fields. (c) First three AMA receptive fields learned using the contrast normalized cone excitations.

evaluated AMA on the contrast normalized cone excitations. By contrast normalizing the cone excitations, the contribution of overall intensity is essentially removed. Figure 11(c,d) shows that performance improves greatly with contrast normalization. Estimates obtained from the AMA receptive field responses are essentially perfect. Estimates obtained from linear regression are substantially improved. This observation is consistent with previous results on computational color constancy that show that contrast-based representations are effective for supporting color constancy when only illumination spectra vary (e.g. Land, 1986). However, contrast-based representations are less effective at supporting constancy when the spectra of background objects in the scene vary (e.g. Brainard & Wandell, 1986).

## Condition 3

Condition 3 introduces variation in the reflectance spectra of the background objects, in addition to the sources of variation in Conditions 1 and 2. This condition models the sources of real-world spectral variation that are most relevant for the computational problem of luminance constancy. We again trained and evaluated AMA for labeled image patches, using the contrast normalized cone excitations. The LRV estimates obtained via AMA are more variable and less accurate than for the previous conditions (Figure 12a). Nevertheless, the estimates provide useful information about the target LRV. Indeed, the estimates track the true LRV on average, as seen by the fact that the mean estimate (red line) lies along the positive diagonal. The increased estimation variability is indicated by the increased width of the shaded red region, relative to its width for the results of Condition 1 and 2. Performance of the baseline linear regression method (also evaluated using contrast normalized cone excitations) is worse than that of the AMA-based estimates.

Figure 12b shows the responses of the first two AMA receptive fields to the training stimuli. Although the responses vary systematically with target object LRV, there is considerable overlap in the receptive field responses for stimuli having different LRV values. This overlap is due to the combined effect of the variation in the illumination and background surface spectra. Recall, however, that the performance shown Figure 12a is based on six rather than two receptive fields; there is likely to be

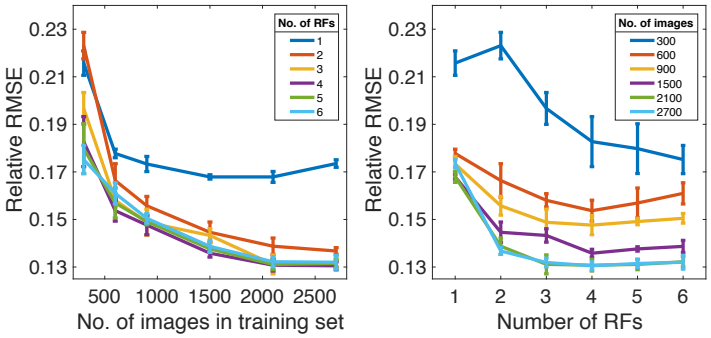


Figure 13: **Change of relative RMSE with number of receptive fields (RFs) and size of training set:** (a) Change of relative RMSE for Condition 3 with the size of training set. The relative RMSE converges around 2000 images. To make this plot, the 300 image test set for Condition 3 was used in each case, while the training set size varied between 300 to 2700 images. The number of receptive fields used was varied between 1 and 6. (b) Same data as in panel a, but plotted as a function of number of receptive fields. Performance converges around 3 RFs. The error bars in each panel show +/- 1 SEM, with the variation taken across multiple runs with different random initializations for the numerical search used in our implementation of AMA.

less overlap in the full six-dimensional response space. That is, inclusion of more receptive fields will in general reduce the ambiguity that is seen when we visualize the responses of just the first two receptive fields. This effect is shown in the right panel of Figure 13. That figure also illustrates the effect of training set size on estimation performance.

Figure 12c shows the first three receptive fields learned for Condition 3. The first two are similar to the receptive fields for Condition 1, in that the L and M cone excitations receive large weights, while the S cone excitations are less heavily weighted. In contrast to the Condition 1 RFs, there is also systematic contribution of cone excitations from background object locations, with both positive and negative contributions from background objects. The specific geometry of the receptive fields, however, should not be taken to be a strict prediction for receptive fields in the human visual system, as the particular RF geometry arises partly as a consequence of the fixed scene geometry used in the training set. Indeed, aspects of spatial structure of the training images can be seen in the RFs. Also, the third receptive field for Condition 3 places weight on the S cone excitations. Although S cone signals do not contribute directly to luminance, S cone responses are correlated with L and M cone responses to natural spectra (Burton & Moorhead, 1987; Benson, Manning, & Brainard, 2014). Information from the S cones can therefore contribute usefully to performance in the task. Finally, we note that there is some run-to-run variation in the structure of the optimal receptive fields that depends on the random initialization of the numerical search used in our implementation of AMA. (The cost landscape is not convex so local minima are possible.) However, such run-to-run RF variation leads only to very small changes in estimation error (see error bars in Figure 13). Furthermore, even if the order varies somewhat, the same set of RFs tends to be learned.

Our rendering software allows us to compare the effect of background surface reflectances on target object LRV with and without simulation of secondary reflections of light from one object onto another. Such secondary reflections were simulated for our primary results. When we turn off this feature of the rendering, we find (data not shown) that that the secondary reflections have minimal effect on LRV estimation: the estimates without secondary reflections were similar to those with reflections. This suggests that the primary source of the estimation error in Condition 3 is caused by image-to-image variation in the reflectance of the background objects, which in turn affects the normalized contrast representation of the target.

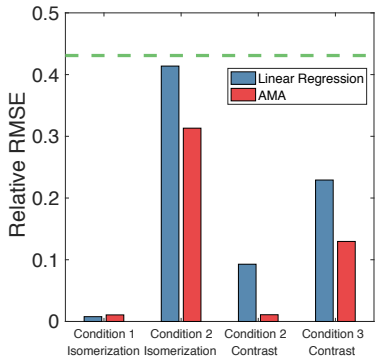


Figure 14: **Summary of luminance constancy performance:** Relative RMSE estimation error for each method and condition. Relative RMSE is the square root of the mean of the squared difference between the estimated and true LRV divided by the true LRV. The mean is taken over all stimuli in the test set. The dotted line represents the relative RMSE for a naive method that estimates the LRV of each image as the mean LRV over the images in the set (LRV = 0.4).

# Discussion

## Summary

In this paper, we studied luminance constancy using naturalistic images. We used a software pipeline, which was developed for this work, to render datasets of multispectral images from scene descriptions. Because we rendered the images, we could label each image in the dataset by the light reflectance value (LRV) of the target object. We used the labeled datasets to learn estimators for the target object LRV. Across scenes, we varied the LRV of the target object, the relative reflectance spectrum of the target object, the spectral power distributions of the light sources, and the reflectance spectra of the background objects in the scene. These spectral variations were based on statistical models of natural surface reflectance spectra and natural illumination spectra. We studied the how the performance of the learned estimators changed with systematic manipulation of the spectral variation in the datasets.

Figure 14 summarizes our main findings, by showing the overall relative root mean squared estimation error (relative RMSE) for each condition. In Condition 1, where only the relative reflectance spectrum of the target object varies, luminance constancy is an easy computational problem, and performance is essentially perfect for both the baseline linear regression and AMA methods. In Condition 2, where the spectral power distribution of the illumination also varies, the problem of luminance constancy is more difficult. Indeed, neither of our methods perform well when they operate directly on the cone excitations. Performance with both methods improves greatly when the cone excitations are contrast normalized. Indeed, AMA performance for Condition 2 is essentially perfect, as it is in Condition 1. Linear regression does less well, in part because it only has access to cone excitations at image locations corresponding to the target object. Condition 2, however, does not include scene-to-scene variation in the background surface reflectances, as occurs in natural viewing. The results for Condition 3 show performance when we add such variation. Performance here represents the overall level of luminance constancy that our methods achieve for the most realistic condition that we tested. In this case, AMA-based estimates of target object LRV are, on average, within 13% (relative RMSE) of the true value. Introducing variation in the surface reflectance spectra of the background objects makes the luminance constancy problem substantially more difficult. Variation in the reflectance spectra of the background objects surfaces reduces how reliably

light reflected from those objects can be used as a cue to estimate the spectral power distribution of the illuminant.

## Images of Virtual vs. Real Scenes

As noted in the introduction, supervised learning requires large labeled datasets. Labeled datasets of natural images have been useful for developing normative models of the estimation of defocus blur, binocular disparity, retinal image motion, and 3D surface orientation (Burge & Geisler, 2011, 2012, 2014, 2015; Sebastian, Burge, & Geisler, 2015; Kim & Burge, 2018; Burge, Fowlkes, & Banks, 2010; Girshick, Landy, & Simoncelli, 2011; Burge, McCann, & Geisler, 2016; Goncalves & Welchman, 2017).

Unfortunately, such datasets are not readily available for studying correlates of object surface reflectance. To obtain ground truth information about the reflectance of objects in a natural image, it is necessary to make an independent measurement of the of the reflectance function of each object (or of the illumination impinging on each object). Although this can be done at a small number of image locations by inserting discretely spaced reflectance standards into the scene, interpolating to locations between such standards requires strong assumptions that are often unjustified in real scenes.

Here we used labeled images rendered from virtual scene descriptions. There are many advantages to using rendered images (Butler, Wulff, Stanley, & Black, 2012). One advantage is that they allow us to work with large number of labeled images where object reflectance is precisely known at each pixel. A second advantage is that we can control the variation in distinct scene factors that might affect the difficulty of the estimation problem. This flexibility allows the study of individual scene factors as well as their combined effects. Here, we exploited this flexibility to quantify how variation in the relative reflectance spectrum of the target object, the spectrum of the illumination, and the reflectance spectrum of the background objects limit LRV estimation. We also exploited our use of rendered images to explore how the presence or absence of secondary reflections from background objects affected estimation of target object LRV. This type of question cannot be addressed using real images. The basic approach we use here can be extended to include parametric control over the amount of variation of different factors. For example, we could systematically vary the variances of the distribution over the weights that control the relative spectrum of the illumination.

There are also disadvantages associated with using rendered images. Virtual rendered scenes are not guaranteed to capture all of the task relevant variation that occurs in real scenes. Even casual inspection of our images reveals that they are rendered, and not real. However, computer graphics is getting better and we expect that the gap between virtual and natural image databases will steadily close. Indeed, carefully constructed graphics images are now quite difficult to differentiate from real images.

To increase the representativeness of our rendered images, we used datasets of natural surface reflectance spectra and natural daylight illumination spectra. Although we believe these datasets provide reasonable approximations of the statistical variation in reflectance and illumination spectra, they can be extended and improved. For example, the surface reflectance datasets capture regularities in naturally occurring reflectance functions, but the surfaces in the datasets were not sampled to reflect the relative frequency of different reflectance spectra in natural viewing. In addition, the dataset of illuminations does not include the artificial illuminations that we now live much of our lives under. Future work with virtual image datasets will benefit from targeted efforts to characterize additional features of the natural statistics of the scene components that are varied.

## Future Directions

In the work presented here, we studied computational luminance constancy in virtual scenes with naturalistic spectral variation in light sources and in surface reflectance functions. It is natural to begin with spectral variation, because this variation is at the heart of what makes luminance constancy a rich computational problem. In natural scenes, however, there are other sources of variation that add additional richness and make the problem more difficult to solve. These include variation in non-spectral

properties of objects and lighting in the scene, including object texture, material, and shape as well as lighting geometry. The methods we developed here may be generalized to study the effects of variation in these factors. One could incorporate these other sources of variation into the generation of the scenes and again learn estimators from the corresponding labeled images. A challenge for this approach will be to thoughtfully control the increase in problem complexity, both to keep compute time feasible and to ensure that it is possible to extract meaningful insight from the results.

Luminance constancy is a special case of the more general problem of color constancy. The approach we developed here could also be generalized to the estimation of additional surface reflectance descriptors, such as object hue or chroma. A further advance would be to develop methods for learning receptive fields that are optimal for multivariate estimation problems. With such methods one could directly estimate three-dimensional color descriptors.

### Linking to Human Performance

An important motivation for studying the computational problem of luminance constancy is to gain insight about human vision. The approach we developed can be used to make predictions of how humans would perform in a psychophysical task that probes the visual representation of object LRV. For example, one could study discrimination thresholds for target LRV using the stimuli that were generated with the methods used here. More specifically, we could ask how LRV discrimination thresholds are impacted by spectral variation in the illuminant, the background objects, and in the target object itself. Comparing human thresholds with the precision of AMA-based estimates would then allow inferences about how well human observers make use of the information available in the images for performing LRV discrimination. We think experiments and analyses along these lines represent an important future direction.

## Acknowledgments

Supported by startup funds from the University of Pennsylvania (JB), NIH grant R01 EY028571 from the National Eye Institute and the Office of Behavioral and Social Science Research (JB), NIH grant RO1 EY100106 (DHB) from the National Eye Institute, Simons Foundation Collaboration on the Global Brain Grant 324759 (DHB), and the University of Pennsylvania Computational Neuroscience Initiative (VS). Preliminary versions of this work were presented at the Computational Cognitive Neuroscience Conference (2017) and the Annual Meeting of the Society for Neuroscience (2017).

## References

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, *7*(6), 531-546.

ASTM. (2017). Standard test method for luminous reflectance factor of acoustical materials by use of integrating-sphere reflectometers. *Renovations of Center for Historic Preservation*, *98*(A), E1477.

Barron, J. T. (2015). Convolutional color constancy. *Proceedings of the IEEE International Conference on Computer Vision*, 379-387.

Barron, J. T., & Malik, J. (2012). Color constancy, intrinsic images, and shape estimation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 57-70.

Bell, S., Bala, K., & Snavely, N. (2014). Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, *33*(4), 159.

Benson, N. C., Manning, J. R., & Brainard, D. H. (2014). Unsupervised learning of cone spectral classes from natural images. *PLoS Computational Biology*, *10*(6), e1003652.

Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science*, *1*, 519-546.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A*, *14*(7), 1393-1411.

Brainard, D. H., & Radonjic, A. (2014). Color constancy. In L. M. Chalupa & J. S. Werner (Eds.), *The New Visual Neurosciences*. MIT Press, 545-556.

Brainard, D. H., & Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, *3*(10), 1651-1661.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, *310*(1), 1-26.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *Journal of Neuroscience*, *30*(21), 7269-7280.

Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, *108*(40), 16849-16854.

Burge, J., & Geisler, W. S. (2012). Optimal defocus estimates from individual images for autofocusing a digital camera. *Digital Photography VIII*, *8299*, 82990E.

Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of Vision*, *14*(2), 1-18.

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature Communications*, *6*, 7900.

Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks: Computational improvements, filter robustness, and coding advantages for scaled additive noise. *PLoS Computational Biology*, *13*(2), e1005281.

Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3d tilt from local image cues in natural scenes. *Journal of Vision*, *16*(13), 1-25.

Burton, G., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*(1), 157-170.

Butler, D. J., Wulff, J., Stanley, G. B., & Black, M. J. (2012). A naturalistic open source movie for optical flow evaluation. *European Conference on Computer Vision (ECCV)*, 611-625.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51-62.

Chakrabarti, A., & Zickler, T. (2011). Statistics of real-world hyperspectral images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 193-200.

CIE. (1986). *Colorimetry, second edition* (Tech. Rep. No. 15.2). Bureau Central de la CIE.

Ciurea, F., & Funt, B. (2003). A large image database for color constancy research. *Color and Imaging Conference*, *2003*(1), 160-164.

D'Zmura, M., & Iverson, G. (1994). Color constancy. iii. General linear recovery of spectral descriptions for lights and surfaces. *Journal Of The Optical Society Of America A*, *11*(9), 2389-2400.

D'Zmura, M., Iverson, G., & Singer, B. (1995). Probabilistic color constancy. *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*, 187-202.

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674-700.

Funt, B. V., & Drew, M. S. (1988). Color constancy computation in near-mondrian scenes using a finite dimensional linear model. *Computer Society Conference on Computer Vision and Pattern Recognition, 1988*, 544-549.

Gattass, R., Gross, C., & Sandell, J. (1981). Visual topography of V2 in the macaque. *Journal of Comparative Neurology*, *201*(4), 519-539.

Gattass, R., Sousa, A., & Gross, C. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of*

*Neuroscience*, *8*(6), 1831-1845.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision*, *9*(13), 1-6.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926-932.

Goncalves, N. R., & Welchman, A. E. (2017). "What not" detectors help the brain see in depth. *Current Biology*, *27*(10), 1403-1412.

Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014). Rendertoolbox3: Matlab tools that facilitate physically based stimulus rendering for vision research. *Journal of Vision*, *14*(2), 1-22.

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, quanta, and vision. *The Journal of General Physiology*, *25*(6), 819-840.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181-197.

Hernández-Andrés, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern europe. *Journal of the Optical Society of America A*, *18*(6), 1325-1335.

Jacobs, G. H. (1981). *Comparative Color Vision*. New York: Academic Press.

Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of neural response. *Journal of Vision*, *17*(12), 1-26.

Jakob, W. (2010). Mitsuba renderer. *http://www.mitsuba-renderer.org*.

Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the munsell book of color from spectrophotometric measurements. *Journal of the Optical Society of America*, *33*(7), 355-376.

Kim, S., & Burge, J. (2018). The lawful imprecision of human surface tilt estimation in natural scenes. *eLife*, *7*, e31448.

Land, E. H. (1977). The retinex theory of color vision. *Scientific American*, *237*(6), 108-128.

Land, E. H. (1986). An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, *83*(10), 3078-3080.

Maloney, L., & Wandell, B. A. (1986). Color constancy: a method for recovering surface spectral reflectances. *Journal of the Optical Society of America A*, *3*, 29-33.

Marimont, D. H., & Wandell, B. A. (1994). Matching color images: the effects of axial chromatic aberration. *Journal of the Optical Society of America A*, *11*(12), 3113-3122.

Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision. *Journal of Experimental Biology*, *146*, 21-38.

Nascimento, S. M., Amano, K., & Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision Research*, *120*, 39-44.

Olmos, A., & Kingdom, F. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, *33*(12), 1463-1473.

Parraga, C. A., Brelstaff, G., Troscianko, T., & Moorehead, I. R. (1998). Color and luminance information in natural scenes. *Journal of The Optical Society of America A*, *15*, 563-569.

Sebastian, S., Burge, J., & Geisler, W. S. (2015). Defocus blur discrimination in natural images with natural optics. *Journal of Vision*, *15*(5).

Skauli, T., & Farrell, J. (2013). A collection of hyperspectral images for imaging systems research. *Digital Photography IX*, *8660*, 86600C1-C7.

Tkacik, G., Garrigan, P., Ratliff, C., Milcinski, G., Klein, J. M., Sterling, P., et al. (2011). Natural images from the birthplace of the human eye. *PLoS ONE*, *6*(6), e20409.

Vrhel, M. J., Gershon, R., & Iwan, L. S. (1994). Measurement and analysis of object reflectance spectra. *Color Research & Application*, *19*(1), 4-9.

# Appendix: Luminance constancy under fixed geometry

Vijay Singh, Nicolas P. Cottaris, Benjamin S. Heasly, David H. Brainard, Johannes Burge

## 1 Illumination Spectra

Denote the $i$'th spectrum in the Granada dataset as $I_i^{\mathrm{G}}(\lambda)$, where $\{i \in [1, M]\}$ and $M$ is the total number of spectra in the dataset. Since the measured spectra vary over several orders of magnitude in overall intensity, we normalize each spectrum by dividing it by its mean power to obtain what we refer to as rescaled spectra:

$$I_i^{\mathrm{S}}(\lambda) = \frac{I_i^{\mathrm{G}}(\lambda)}{\int d\lambda I_i^{\mathrm{G}}(\lambda)}. \tag{1}$$

For simplicity of notation, we denote wavelength $\lambda$ as a continuous variable; in the actual calculations wavelength is discretely sampled and integrals are approximated by sums. The Granada dataset was measured at 5 nm sampling intervals between 300 and 1100 nm. We subsampled the spectra to the 400-700 nm interval, 10 nm spacing representation used for rendering, and performed our calculations at this sampling.

The rescaled spectra $I_i^{\mathrm{S}}(\lambda)$ were mean centered for PCA by subtracting out the mean ($\bar{I}^{\mathrm{S}}(\lambda)$), over all rescaled spectra in the dataset,

$$\bar{I}^{s}(\lambda) = \frac{1}{M} \sum_i I_i^{\mathrm{S}}(\lambda). \tag{2}$$

The mean centered dataset, $I_i^{\mathrm{MC}}(\lambda) = I_i^{\mathrm{S}}(\lambda) - \bar{I}^{\mathrm{S}}(\lambda)$ was decomposed as:

$$I_i^{\mathrm{MC}}(\lambda) = \sum_j w_{ij} e_j^{\mathrm{PCA}}(\lambda), \tag{3}$$

where the $e_j^{\mathrm{PCA}}(\lambda)$ are the PCA basis vectors obtained using singular value decomposition (SVD) applied to $I_i^{\mathrm{MC}}(\lambda)$, and $w_{ij}$ are the weights obtained by projecting each of $I_i^{\mathrm{MC}}(\lambda)$ onto the basis vectors $e_j^{\mathrm{PCA}}(\lambda)$ $\left(w_{ij} = I_i^{\mathrm{MC}}(\lambda) \cdot e_j^{\mathrm{PCA}}(\lambda)\right)$. We used the basis vectors corresponding to the largest six SVD eigenvalues, so that $\{j \in [1, 6]\}$. For the rescaled Granada dataset, these six eigenvalues account for more than 90% of the variance.

To get a random illuminant spectrum $\tilde{I}_i(\lambda)$, we sample random weights $\tilde{w}_{ij}$ from a multivariate Gaussian distribution with mean $\bar{w}_j = \frac{1}{M} \sum_i w_{ij}$, and covariance matrix $\Sigma$ given as:

$$\Sigma_{jj'} = \frac{1}{M} \sum_i \left(w_{ij} - \bar{w}_j\right)\left(w_{ij'} - \bar{w}_{j'}\right). \tag{4}$$

The corresponding spectrum is generated from these weights as $\left(\sum_j \tilde{w}_{ij} e_j^{\mathrm{PCA}}(\lambda) + \bar{I}^{\mathrm{S}}(\lambda)\right)$. This spectrum can sometimes have values that are less than zero. In such cases, the weights are discarded and a new

draw obtained, until the condition $\left(\sum_j \tilde{w}_{ij} e_j^{\mathrm{PCA}}(\lambda) + \bar{I}^{\mathrm{S}}(\lambda)\right) > 0$, is satisfied for all $\lambda$. The random illuminant spectrum is

$$\tilde{I}_i(\lambda) = \left(\sum_j \tilde{w}_{ij} e_j^{\mathrm{PCA}}(\lambda) + \bar{I}^{\mathrm{S}}(\lambda)\right). \tag{5}$$

## 2 Surface Reflectance Spectra

Denote the $i$'th sample in the reflectance spectrum dataset as $R_i(\lambda)$, where $\{i \in [1, M]\}$ and $M$ is the total number of spectra in the dataset. We first calculated the mean reflectance spectrum, $\bar{R}(\lambda)$, by taking the sample mean over all spectra in the dataset

$$\bar{R}(\lambda) = \frac{1}{M} \sum_{i=1}^{M} R_i(\lambda). \tag{6}$$

The dataset is then mean centered by subtracting the mean spectrum, $R_i^{\mathrm{MC}}(\lambda) = R_i(\lambda) - \bar{R}(\lambda)$ and decomposed using PCA as:

$$R_i^{\mathrm{MC}}(\lambda) = \sum_j w_{ij}\, e_j^{\mathrm{PCA}}(\lambda), \tag{7}$$

where $e_j^{\mathrm{PCA}}(\lambda)$ are PCA basis vectors obtained using SVD applied to $R_i^{\mathrm{MC}}(\lambda)$ and $w_{ij}$ are the weights obtained by projecting each of $R_i^{\mathrm{MC}}(\lambda)$ onto the basis vectors $e_j^{\mathrm{PCA}}(\lambda)$ $\left(w_{ij} = R_i^{\mathrm{MC}}(\lambda) \cdot e_j^{\mathrm{PCA}}(\lambda)\right)$. We used the basis vectors corresponding to the largest six SVD eigenvalues, which account for more than 90% of the variance in the combined Munsell and Vrhel datasets.

To get a random reflectance spectrum, we generate samples of weights $(\tilde{w}_{ij})$ drawn from the multivariate Gaussian distribution with mean $\bar{w}_j = \frac{1}{M}\sum_i w_{ij}$, and covariance $\Sigma_{jj'} = \frac{1}{M}\sum_i (w_{ij} - \bar{w}_j)(w_{ij'} - \bar{w}_{j'})$. If the randomly sampled weights satisfy the condition $\left(0 < \sum_j \tilde{w}_{ij} e_j^{\mathrm{PCA}}(\lambda) + \bar{R}(\lambda) < 1\right)$ at every $\lambda$, the randomly generated reflectance spectrum is given as:

$$\tilde{R}_i(\lambda) = \sum_j \tilde{w}_{ij} e_j^{\mathrm{PCA}}(\lambda) + \bar{R}(\lambda). \tag{8}$$

Otherwise the draw is discarded and a new set of weights is drawn.

For generating the target object reflectance at a particular target LRV ($Y_{\mathrm{T}}$), the values in a generated spectrum were scaled by

$$\frac{Y_{\mathrm{T}}}{\int d\lambda \tilde{R}(\lambda) D_{65}(\lambda) \bar{y}(\lambda)} \tag{9}$$

with $\bar{y}(\lambda)$ being the CIE photopic luminosity (or luminous efficiency) function, which describes the average spectral sensitivity of human visual perception of luminance. The target reflectance is then given by:

$$\tilde{R}_i^{\mathrm{T}}(\lambda) = \tilde{R}_i(\lambda) \cdot \left(\frac{Y_{\mathrm{T}}}{\int d\lambda \tilde{R}(\lambda) D_{65}(\lambda) \bar{y}(\lambda)}\right). \tag{10}$$