Camera calibration for natural image studies and vision research

Mark Brady[1,*] and Gordon E. Legge[2]

[1]*Department of Psychology, North Dakota State University, 115 Minard Hall, P.O. Box 5075, Fargo, North Dakota 58105, USA*
[2]*Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, Minnesota 55455, USA*
**Corresponding author: mark.brady@ndsu.edu*

Received March 6, 2008; revised October 10, 2008; accepted October 21, 2008; posted October 23, 2008 (Doc. ID 93504); published December 5, 2008

A dominant theme in vision research is that important characteristics of the visual pathway evolved to be effective in processing natural scenes. Given this perspective, one can learn about the nature of visual processing from a quantitative analysis of natural images. Such analysis can benefit from the camera as a measuring device. As such, the camera should not introduce arbitrary artifacts into the image formation process. This paper describes how to correct a number of unnecessary artifacts associated with obtaining natural scene statistics with a digital camera. For artifacts that are inherently part of image formation, and where elimination is not possible or appropriate, we describe methods for transformation and quantification. © 2008 Optical Society of America

OCIS codes: 100.2980, 330.1710.

1. INTRODUCTION

Natural images provide essential raw material for vision research. They are used for the testing of computer vision algorithms, as the source of image statistics, and to stimulate observers in psychophysical experiments. Natural images are captured by both biological and photographic systems. Biological vision has evolved to function in an environment of natural image structure and statistics [1], whereas photographic systems have been developed to meet various commercial criteria. As a result, biological and photographic systems will produce different image representations of the same three-dimensional scene. Given these differences, how can a camera be used so as to predict the response of a given biological vision system?

Geometric distortions in photography are a potential problem in computer vision. For example, suppose a computer vision algorithm is learning to recognize an object based on a number of image examples. The curvature of the object edges is likely to be an important property for learning. If the object is translated between images, field-dependent optical distortions, such as barrel distortion, will cause those curvatures to vary from image to image. Somehow, this has to be taken into account, or the distortion must be verified as being extremely small.

In the case of psychophysics, object recognition performance depends upon image contrast. Data on image contrast from a camera, in turn, depend upon the camera's spectral response function and its luminance response function. These functions vary from camera to camera. As a consequence, the contrast structure and statistical structure of the images will also vary. For this reason, two investigators collecting image data from the same natural objects, but using different cameras, may get different results.

This paper describes how to handle the artifactual characteristics of a camera. Once calibrated, a digital camera can be used to form useful databases of natural images [2].

2. DIGITAL CAMERA BASICS

There are many criteria to use in camera selection, and every project may require something different. Nevertheless, most investigators share some common needs, including lens quality, controllability, and compatible external software.

Resolution requirements vary with the study, but most studies can easily meet their resolution requirements at a reasonable cost. High dynamic range is required if one is interested in scenes combining shadowed regions and direct illumination. Quality complementary-metal-oxide semiconductor (CMOS) sensors in digital single-lens-reflex (SLR) cameras have dynamic ranges of 11.7 bits or more, which exceeds that of film [3,4].

These sensors also have a signal-to-noise ratio greater than ISO 50 film [5,6], and some have a pixel resolution greater than the best 35 mm film equivalent (16 megapixels) [7].

The advantage of the SLR design is that the viewfinder path shares the same lenses as the image capture path, allowing the photographer to see the image as it will be captured. This is especially important in a camera with changeable lenses. Most investigators will require

changeable lenses, since it is unlikely that a fixed lens will meet all of an investigator's requirements.

Zoom lenses are convenient, but a zoom setting can have a number of effects, as described below. Therefore, users may want to set zooms to a small number of calibrated settings or use fixed focal length lenses.

The second category of requirements, controllability, is very important for calibration. The following controls should have a manual mode: shutter speed, aperture, focus, and gain. The flash should be controllable as fixed on or fixed off. One should be able to turn off auto white balance. Finally, the output file format should be controllable and should include a raw format.

Traditionally, and in the image-processing domain, the term "raw" refers to a file containing only uncompressed pixel values. There is no file header. More recently, and in the photography domain, "raw" means something quite different. The photographic raw file does have a header and usually involves lossless compression. It is called "raw" because most of the image processing normally done in the camera is postponed or is reversible using off-camera software. The raw format usually has the additional advantage of having details of the camera settings saved in the header. Data that are not raw are undesirable because they will include the results of proprietary in-camera processing. In this paper, we shall use the term "raw" in the photographic sense.

In order to exploit the raw format, one requires both a camera to produce raw files and external software to read and manipulate those files. It was only with the 2003 release of Adobe Photoshop CS [8] that the ability to read and manipulate raw files was realized in commercial software. However, even applications such as Photoshop CS may leave the image somewhat processed. For example, whereas the user can refuse most suggested operations, the user is still forced to choose a white balance and a color space, with unpredictable results. In addition, commercial software typically performs gamma adjustment and demosaicing upon input, without notifying the user. One of the few available programs that provides access to the fully raw data in a raw file is dcraw [9].

We used three cameras in developing the methods described: the Canon EOS-D30, the Canon EOS-1Ds Mark II, and the Nikon D3.

3. RECOMMENDED EQUIPMENT AND MATERIALS

The procedures described here require the following equipment and materials or their equivalents: (1) digital raw capable camera, preferably with quality lenses and low-noise sensors; (2) software for reading fully raw image data; (3) Accu-Chart 11-Step Gamma Compensated Grey Scale Chart [10]; (4) monochromator; (5) broad-spectrum, high-intensity light source; (6) data set of reflectance spectra; (7) data set of illumination spectra; (8) LCD display; and (9) heavy tripod.

4. ARTIFACTS MANAGED BY CORRECTION OR ELIMINATION

Certain artifacts are considered to be undesirable limitations on resolution, and as such should be corrected or eliminated. However, limitations on resolution have their biological counterparts, and for some applications it may be desirable to reintroduce these limitations by mimicking a biological system.

A. Aberrations

Aberrations arise when design goals diverge from the idealized world of paraxial optics [11]. The two main categories of aberration are the chromatic and the monochromatic. Chromatic aberrations arise because the index of refraction is actually a function of wavelength. Monochromatic aberrations are those that arise even in images formed by monochromatic light. All aberrations can be reduced optically, but there is no known way to bring them all to zero in a single system. Therefore, it is best to choose a lens system design with a minimal amount of aberration in all categories and, where possible, to correct the remaining aberrations digitally.

1. Chromatic Aberrations

There are two types of chromatic aberration, longitudinal and lateral. Longitudinal chromatic aberration results from the fact that each wavelength has a different focal length, resulting in chromatic image blur. Lateral chromatic aberration is due to the fact that images in different wavelengths have different magnifications.

Longitudinal aberrations can be well managed optically. Typical solutions include the achromatic doublet lens with crown and flint glass or, more recently, the use of fluorite and ultra-low-dispersion (UD) glass. See Figs. 1(a)–1(c). Two of the lenses tested for this paper (Canon 16–35 mm f/2.8L USM and the 24–70 mm f/2.8L USM) utilized UD glass, one (Canon 100–400 f/4.5–5.6L IS) utilized UD and fluorite, and one lens (Canon 28–105 mm f/3.5–4.5 II UMS) did not utilize UD or fluorite glass.

However, even in those lenses using UD glass or fluorite, some lateral chromatic aberration remains. This can be corrected digitally as shown in Figs. 1(d) and 1(e) and described below.

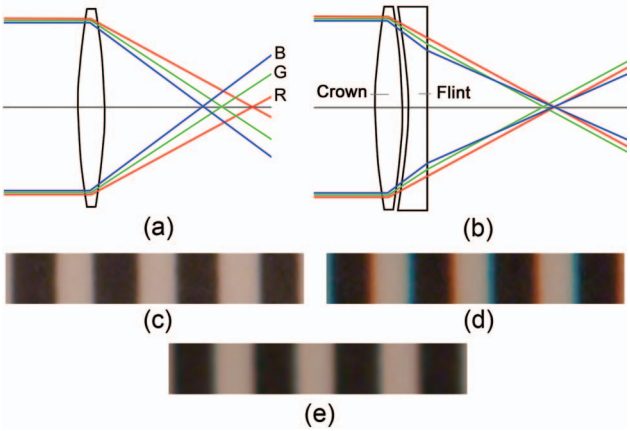


Fig. 1. Chromatic aberration and comparison of solutions. (a) Longitudinal chromatic aberration. Different wavelengths have different focal lengths on axis: R, red; G, green; B, blue. (b) Correction of longitudinal chromatic aberration by achromatic doublet. (c) On-axis test chart detail from the 1Ds camera. Chromatic aberration is corrected using UD glass. (d) Off-axis test detail from the same image as (c), showing lateral chromatic aberration. (e) Same as (d) but with digital correction.

It is likely that the lateral chromatic aberrations of a zoom lens will vary with focal length. For example, lateral chromatic aberration decreases as focal length increases with the Canon 16–35 mm, whereas it increases as focal length increases with the 24–70 mm. The 100–400 mm lens shows little effect of zoom. No effect of zoom setting is discernable on longitudinal chromatic aberration in any of the test lenses.

2. Distortion

Most of the monochromatic aberrations are best corrected optically. These include the four blurring aberrations: spherical aberration, coma, astigmatism, and Petzval field curvature. A fifth monochromatic aberration is distortion, where straight lines in the scene are imaged as curves.

Distortion varies according to the distance between a lens and the aperture stop (the element in a multielement lens that determines the maximum bundle diameter), producing a difference in magnification as a function of distance from the image center. When the stop is between the object and the lens, barrel distortion is produced. When the stop is between the lens and the focal plane, pincushion distortion is produced. When the stop and lens are coincident, there is zero distortion. What is perhaps less well appreciated is that distortion is also dependent on object distance.

The 28–105 mm lens on the D30 camera displays a slight barrel distortion at wide-angle zoom settings and a slight pincushion distortion at telephoto settings, as does the 24–70 mm; the 16–35 mm lens exhibits barrel distortion at wide angle settings but no distortion at 35 mm; and the 100–400 mm lens exhibits increasing pincushion distortion as the focal length increases. Correction procedures, described below, produce grid lines that are straight to within 1 pixel.

3. Correcting Distortion and Chromatic Aberration

Distortion is the only monochromatic aberration that does not involve blur and can therefore be corrected digitally. It can be expressed as a remapping of the ideal undistorted image into the distorted image

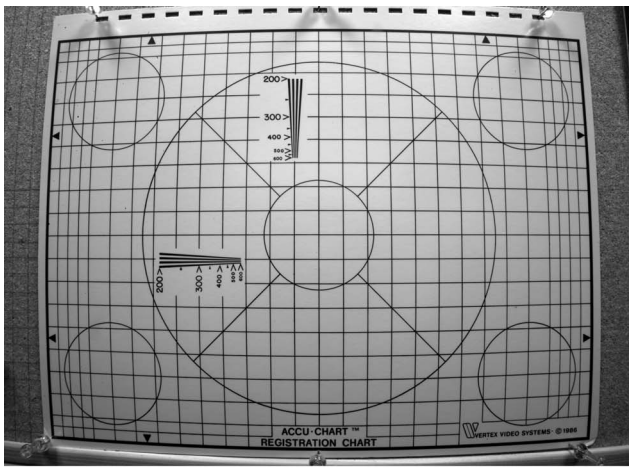
$$r_d = f(r_u), \tag{1}$$

where the distorted radial position r_d is a function of the undistorted radial position r_u. Where the derivative of f is negative, one has barrel distortion, and where the derivative of f is positive, one has pincushion distortion. If f is not monotonic, then one will have both barrel and pincushion distortion in one image, a case called mustache distortion. An example of barrel distortion is shown in Fig. 2(a).

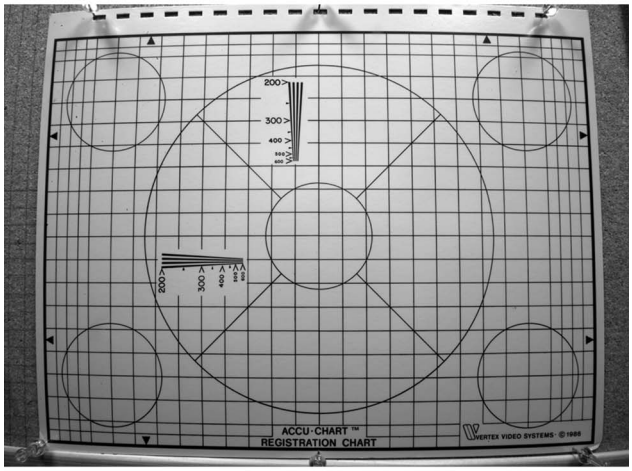
In order to make the correction, we want the inverse of the distortion function f. The inverse f^{-1}, or correction function, is unknown, but we can assume there is a Taylor series

$$r_u(r_d) = a_0 + a_1 r_d + a_2 r_d^2 + a_3 r_d^3 + a_4 r_d^4 + a_5 r_d^5 + \cdots \tag{2}$$

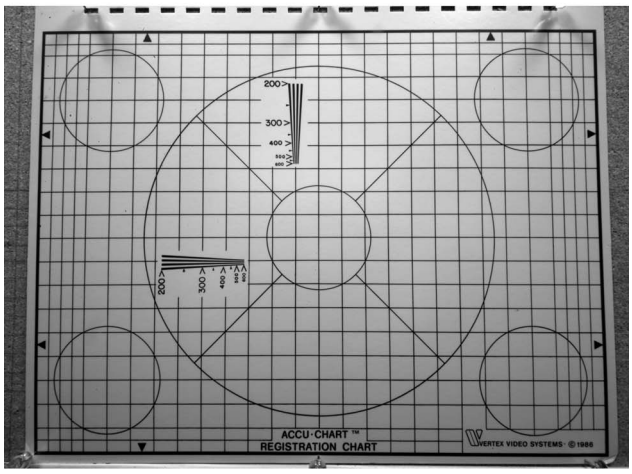
that represents it, where a_i is the nth derivative of f^{-1}. We use a fourth-degree polynomial approximation like the one originally implemented by Dersch [11]:



(a)



(b)



(c)

Fig. 2. Comparison of distortions in a test chart. (a) Unprocessed image, (b) image processed with commercial software, (c) image processed using Eq. (3); (a4, a3, a2) = (0, 0, −0.075).

$$r_u(r_d) = a_1 r_d + a_2 r_d^2 + a_3 r_d^3 + a_4 r_d^4. \tag{3}$$

In Eq. (3), coefficient $a_0 = 0$. If it were not, the correction would leave a hole in the center of the image. If one normalizes all images to have a unit half-diagonal, one can

set a_1 to control image magnification, which we will manipulate separately when dealing with chromatic aberration. The case

$$a_1 = 1 - a_2 - a_3 - a_4 \quad (4)$$

causes no overall magnification. Since the causes of geometric distortion are not color dependent, distortion in all color channels can be corrected using the same set of coefficients.

Lateral chromatic aberration can then be corrected by applying Eq. (3) to two of the three color channel images, choosing a_1 so as to ensure that all color channels have equal magnification. Adjustments to a_1 are made with the aid of a test chart consisting of black and white stripes from the center to the periphery. The adjusted image should show minimal chroma at all eccentricities. Since lateral chromatic aberration is a wavelength-dependent magnification aberration, $a_2=a_3=a_4=0$ for both chromatic corrections.

If lateral chromatic aberration is significant in an image, it cannot be corrected entirely. This is because each color-sensitive element responds to light across many different wavelengths and each wavelength is blurred to a different extent.

The process of finding the appropriate coefficients for correction may be greatly simplified if a commercial software vendor supplies the appropriate coefficients for your lens [12,13]. However, such software assumes an object distance of infinity. In those cases where the object was close, we found that applying commercially available software left images with a significant distortion. Figure 2(a) shows barrel distortion from the 1Ds camera with a 16–35 mm lens and focal plane 29 cm from the test chart. Figure 2(b) shows the output of commercial software, which assumes an infinite object distance.

In order to fully correct the distortion to within 1 pixel, we implement the above described corrections in Matlab. Equation (3) is used to map source (distorted) to target (corrected) pixels. Since the source pixel locations are fractional, we also utilize bicubic interpolation to improve the mapping quality. The results are shown in Fig. 2(c).

B. Vignetting

Vignetting appears in an image as a darkening of the image corners, as shown in Fig. 3. If one were to photograph a uniform field, there would be a darkening of the image as a function of increasing distance from the image center. Vignetting usually appears as a subtle effect because it occurs over relatively large image distances. However, the actual magnitudes are often large, so vignetting can have significant effects on the use of cameras in scientific research. In one test, we found that the amount of light passed by the lens to the corner of the image is about nine times less than the light passed to the center [see Fig. 3(c)].

There are three potential sources for vignetting, which will be dealt with below: mechanical vignetting, optical vignetting, and natural vignetting [14]. In the case of mechanical vignetting, the darkening can be rather abrupt. In the case of optical or natural vignetting, the darkening is continuous.

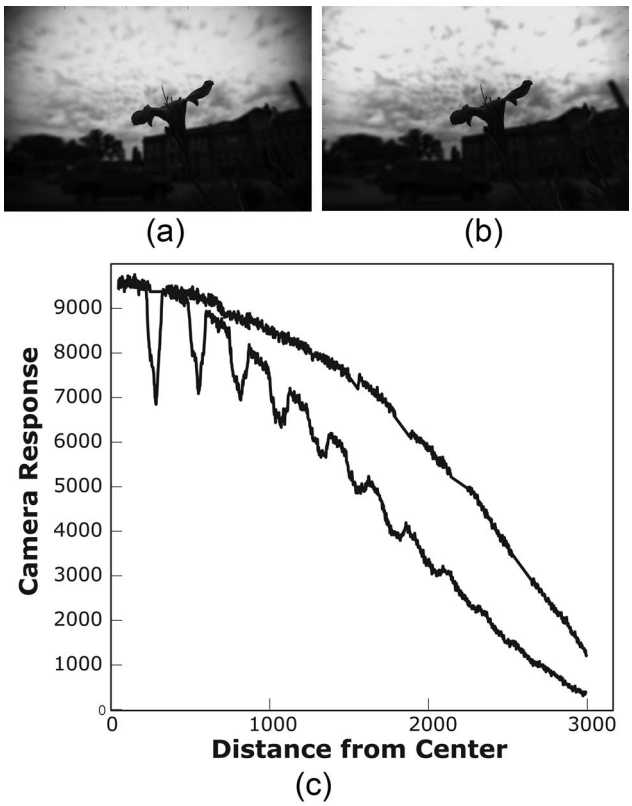


Fig. 3. Vignetting correction. (a) Uncorrected image showing vignetting. The scene is of a flower against a cloudy sky. The unusual appearance is due to the fact that the image is raw, with no gamma or color balance applied. The focus is on the flower in the foreground. The lens is the 16–35 mm set to 16 mm. The aperture=f/2.8. (b) Corrected image, using the algorithm described in the paper. (c) Comparison of correction curves, with and without camera rotation. Top curve, vignetting measured using test chart and camera rotation. Bottom curve, vignetting as measured without camera rotation. Nonuniformities across the test chart are confounded with actual vignetting, which would lead to significant overcorrection. Dips are due to registration marks in the test chart.

In cameras, mechanical vignetting is caused by occluding extensions of the lens barrel, such as lens shades and stacked filters. This may result in completely black corners and cannot be digitally corrected. Fortunately, mechanical vignetting can be corrected simply by removing the extensions.

Optical vignetting occurs because, for oblique rays, the effective aperture size is reduced by the diameters of the front and rear elements. A smaller aperture stop will suffer less from this delimiting. Thus, reducing the aperture can correct the problem. However, if a larger aperture is required, optical vignetting can still be corrected digitally.

Natural vignetting is caused by the angle made between the rays from the exit pupil and the sensor surface. As the angle decreases, the collected light decreases as well. This drop off is typically modeled using a $\cos^4(\theta)$ law [14]. However, there are many lens designs where the $\cos^4(\theta)$ model does not hold, so an empirical approach must be taken when correcting natural vignetting. Natural vignetting cannot be corrected via aperture adjustment but can be corrected digitally.

The degree and structure of vignetting varies with aperture, zoom setting, and focus. Vignetting tends to in-

crease with aperture size, since portions of large ray bundles are more easily occluded by aperture stops. However, the effects of zoom setting depend on the lens. The 16–35 mm lens exhibited more vignetting at small zoom settings, where as the 100–400 mm lens exhibited the opposite trend. Overall, the wide-angle lens showed more vignetting than the telephoto.

As in the case of distortion, we take some diagnostic photographs, measure the artifact, and correct it with an inverse function.

A uniformly reflective and uniformly illuminated surface would make a good subject for our diagnostic photograph. Unfortunately, such conditions are difficult to produce in the real world. Reflective surfaces typically have nonuniform bidirectional reflectance distribution functions and are nonuniformly illuminated. Luminous surfaces, such as LCD displays, are viewpoint dependent and therefore create nonuniform images, even in the absence of vignetting. Because of the confound between nonuniformity of the test surface and vignetting, it is preferable to create a virtual version of the ideal scene by compositing several photographs. The subject of the photograph should be a small and approximately uniform colored rectangular patch. Using a heavy tripod, the patch may be photographed so that it appears in the center of the image.

Keeping the tripod in place, the camera body can then be rotated such that the patch moves along the image diagonal, a distance r_I/n from the center, where r_I is the diagonal radius of the image. The diagonal is used because it contains the full range of radii in an image. The degree of vignetting in the corners is not easily estimated from vignetting along the shorter horizontal or vertical radius. The set of resulting patches are the same patch on the test surface but occur at different locations in each image.

For each rectangular patch, collect pairs $(r(i),R(i))$ along the patch diagonal, where $r(i)$ is the distance from the center of the image and $R(i)$ is the raw camera response. Let $v(r)$ be any well-fitting interpolation function of the points $(r(i),R(i))$. It is then possible to multiply each pixel $I(x,y)$, at $(column,row)=(x,y)$, by $1/v(r(x,y))$ in order to correct the vignetting.

This procedure is dependent on the lens, aperture, zoom setting, and focusing distance. Therefore, these parameters must be set during calibration to match those expected during camera application.

C. Dark Current

A pixel's dark current is its response in complete darkness and may have a number of design-dependent factors. Such design details are beyond the scope of this paper. Dark current can be measured by averaging over the entire sensor. However, dark current can vary from pixel to pixel, creating fixed patterns of hot pixels and cold pixels. Hot pixels are those with higher than average responses, and cold pixels are those with lower than average responses. These atypical pixels may occur individually, in neighborhood clusters, or in columns. If hot or cold pixel responses are too extreme, they may be considered to be defective. The best way to treat defective pixels is to discard their responses and replace them with interpolated values. Subsection 4.D describes how to correct dark current in nondefective pixels.

D. Factors Affecting Dark Current

Dark current usually has a thermal noise component. Therefore, it should be measured at the same temperature as the intended application environment. Dark current also typically depends upon gain.

One might expect the accumulated dark current to be proportional to the exposure time. With the EOS-D30, this is not the case at all. For short exposures (<0.5 s), the dark current is found to be a function of sensor color and otherwise constant. For longer exposures, dark current is a function of color but also increases monotonically with shutter speed.

The median raw camera response for all exposures on the blue channel is 0, 16 for all exposures in the green channel, and 0 for the majority of exposures on the red channel; 1/1000, 20, and 25 s. are the exceptions for the red channel, where the median dark current is 18.

Dark current can be measured simply by capturing images in darkness with the lens cap on at various shutter speeds and then computing a mean or median dark pixel value for each color channel to be subtracted from the application image.

In some cases, dark current is corrected automatically on the sensor chip [15]. In such cases, dark current subtraction may not be appropriate.

5. OPTOELECTRONIC CONVERSION FUNCTION AND COLOR MATCHING FUNCTIONS

Given a certain amount of energy falling on a particular pixel, the optoelectronic conversion function (OECF) returns that pixel's response. Given a certain wavelength, the color matching function gives the relative pixel responses of the three color channels. This may all sound very simple. However, in an actual camera system, the energy response and the color matching function of a pixel are superimposed into a single pixel response. Therefore, in order to characterize the energy response and color matching function of a camera, one must tease these two functions apart. In this section, we shall describe how to separate and characterize the two functions.

A. Measuring the Optoelectronic Conversion Function

The goal in measuring the OECF of a camera is to determine the function $R(Y)$, where Y is the scene luminance in cd/m^2 and $R(Y)$ is the camera response. This is accomplished by measuring a series of gray test patches and recording the corresponding raw camera responses in each color channel.

In determining our camera's general response to light energy, we would like to temporarily avoid the issue of color. This suggests using a white light, which has equal power at all wavelengths. However, in actual practice, what we recognize as a white surface comes about by heterochromatic illumination of a surface, which in addition may not reflect all wavelengths equally. Furthermore, there are many different spectra that are considered

white, such as the various CIE daylight spectra. All of this suggests that we must choose some reference white spectrum. The reason this works is that the procedure described here manipulates the energy of the various wavelengths in our arbitrary spectrum by an equal factor, thus avoiding any influence of color.

Our procedure roughly follows that described in ISO 14524:1999(E) method B [16]. In comparison, ISO 14524:1999(E) method A measures the response of the sensor minus the lens. Since the camera will always be used with a lens, we find it more practical to calibrate the camera–lens as a single unit. Furthermore, we have found that our primary test camera (the 1Ds) gives identical responses to a light energy stimulus with a variety of lenses regardless of the zoom setting. This makes it possible to measure the OECF with one lens and zoom setting and then apply the camera with an entirely different lens and zoom. However, the response of the Nikon D3 proved to be lens dependent using the AF Nikkor 50 mm $f/1.8$ D and the AF DX Fisheye-NIKKOR 10.5 mm $f/2.8$ ED.

In order to collect luminance and camera response data, we require a standardized gray-scale target. For this, we use the Accu-Chart 11-Step Gamma Compensated Grey Scale Chart [10]. The ISO Camera OECF Test Chart is also useful for this purpose [16]. However, we prefer the 11-Step Chart because it is counterbalanced for horizontal and vertical positional variation.

Uniformity of spectral reflectivity across gray patches is important. This can be verified by recording not only the luminance for each patch but also the (x,y) chromaticity values for each gray band in the CIE 1931 space to ensure that the gray bands are approximately achromatic and form a small cluster. They should cluster near the achromatic point $(x,y)=(1/3,1/3)$. See Fig. 4.

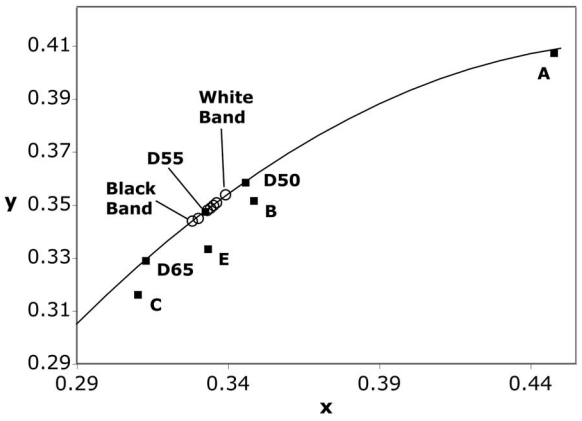


Fig. 4. Color check of direct sun illuminated bands on the 11-Step Chart. Each circle represents the color of a gray test chart band in CIE chromaticity space. These points should be close to the various CIE daylight colors and the achromatic point E, and they are; B, direct sunlight; D55, cloudy daylight; D65, CIE daylight; C, average sunlight. The gray band colors fit very closely to the Stiles daylight locus function. The Stiles function is based on a curve fit to a large number of daylight observations [23]. The color temperature of D55, in the midst of the gray cluster, is 5500 K. This temperature is an exact match to the EOS-D30's color temperature assumption for daylight color balance. For comparison, nondaylight whites are A, tungsten, and D50, bright tungsten.

Cameras with high-quality CMOS sensors have a high dynamic range. This requires the use of multiple exposure settings in order to exercise the entire range of the sensor's response. By combining images of the same chart at two exposure settings, the chart becomes a virtual chart with a broader luminance range than possible with any real single chart. This requires adjustment to an effective luminance, which is

$$Y_E = Y_o * T_o/T_r, \tag{5}$$

where Y_o is the luminance measured with a photometer, T_r is an arbitary reference exposure, and T_o is the other exposure. Essentially, measuring the light from a patch for twice as long is equivalent to measuring a patch that is twice as bright.

For our reference illuminant we use natural daylight at noon, latitude 46.9. When reflected off the 11-Step Chart, this very closely approximates CIE daylight D55. Certain artificial sources may have more constant spectra with respect to time of year. However, we prefer the directional nature of natural daylight, which provides the most uniform illumination of a test chart.

We find that the OECF of the 1Ds is exactly linear with color-channel-dependent slopes, as shown in Fig. 5.

When using exposures, apertures, and gain other than those used here, one must adjust the (R,G,B) values in order to retain the same OECF. For instance, given reference exposure T_r, reference aperture diameter D_r, and reference gain G_r, the adjusted (R,G,B) would be

$$(R_{adj},G_{adj},B_{adj}) = T_r D_r^2 G_r (R,G,B)/(TD^2G), \tag{6}$$

where T, D, and G are the new exposure, diameter, and gain, respectively.

B. Measuring the Color Matching Function

Our goal here is to measure the camera's response to color at constant energy. One complication is that the space of color spectra has an infinite number of dimensions. We can simplify things by studying an incomplete, yet repre-

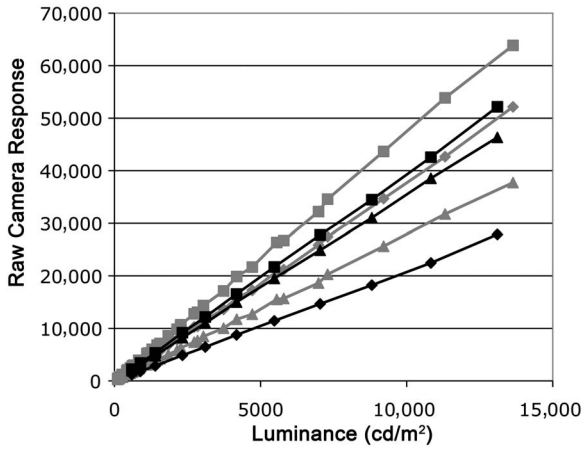


Fig. 5. Camera response as a function of luminance. Reference exposure for both cameras is 0.002 s. Reference apertures are $f/5.6$ and $f/8.0$ for the Canon 1Ds and Nikon D3, respectively. The lens for the D3 is the AF DX Fisheye-NIKKOR 10.5 mm $f/2.8$ ED. Black lines indicate the 1Ds and gray lines the D3. Diamonds, squares, and triangles indicate red, green, and blue channels, respectively.

sentative, subset of all possible spectra. The representative subset normally used for this purpose is the one-dimensional space of monochromatic lights at constant energy. We used wavelengths spaced at 10 nm from 400 nm to 720 nm.

Once we have a description of the wavelength-dependent response properties of our camera, we can combine this description with the fact that our sensor responds linearly to energy in order to predict the camera's response to any spectrum. In fact, we will do this in order to map camera responses to predicted human cone responses.

Our approach roughly follows that described in ISO 17321-1:2006(E) [17] method A and the spectroradiometric approach of Martínes-Verdú *et al.* [18]. In comparison, ISO 17321 method B involves measuring the camera's response to complex spectra. However, it is not obvious how the resulting data can be used to predict camera responses to novel spectra. We also do not use the optical approach mentioned by Martínes-Verdú *et al.* because it requires proprietary knowledge of the optical design.

In order to produce monochromatic light, we use a monochromator. One complication is that the output of the monochromator depends on wavelength. This variable energy has two particular consequences that must be dealt with. The first consequence is that we are forced to use various exposure times to avoid sensor activations near dark current levels and saturation levels. Saturation levels can be especially problematic, since some whitening of the camera response may occur. Various exposure times can easily be adjusted for, since collected energy is proportional to exposure time. We use the adjustments as described in Subsection 5.A.

Having adjusted for various exposure times, the variable lamp energy must also be corrected for. This is accomplished by applying the lamp's relative energy quotient as measured with a radiometer. The resulting normalized lamp energy output, as a function of wavelength, is constant.

Our monochromator is composed of a CRi Varispec tunable filter illuminated with a xenon lamp. Unlike most artificial light sources, the xenon lamp provides significant energy at all visible wavelengths. Wavelengths outside the visible range are filtered out using heat glass. This prevents damage to the monochromatic filter and limits fluctuations due to temperature. Fans are also used to keep the heat glass and filter at a constant temperature.

Radiometric measures of the monochromatic signal and camera response measures can be taken simultaneously or sequentially. The simultaneous approach has the advantage that there is no drift in signal strength over time. This can be achieved by using a broadband nonpolarizing beam splitter. However, we found that even broadband splitters exhibit some wavelength dependence in how the beam is split. Therefore, we used a sequential approach and bracketed our camera measurements with repeated radiometer measures to ensure temporal stability. We found no measurable drift over time.

Both the camera measures and the radiometer measures are taken by aiming directly into the filter. This avoids any concern about wavelength-specific properties of a reflecting target.

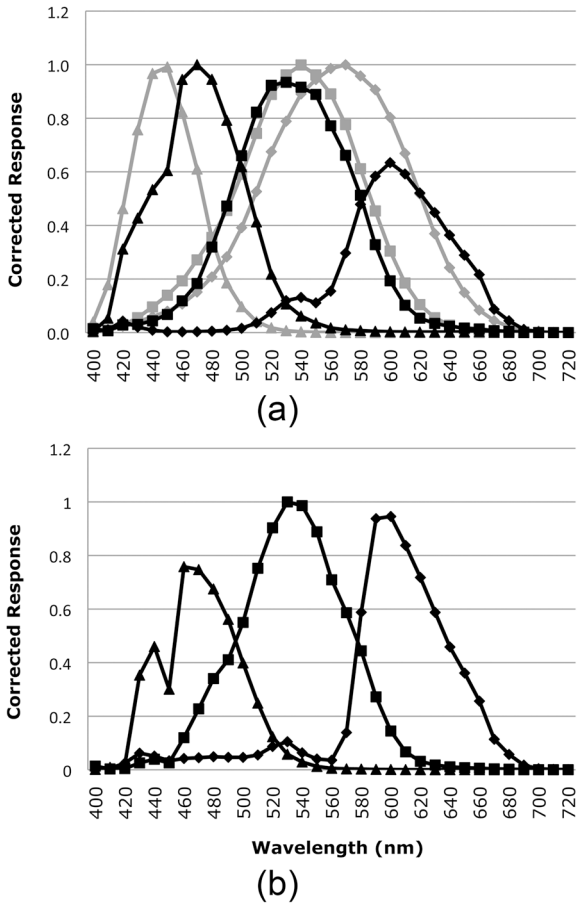


Fig. 6. (a) Canon 1Ds color matching function (black) compared to human cone fundamentals [19] (gray). Diamonds, red channel and L cone; squares, green channel and M cone; triangles, blue channel and S cone. Notice the second mode for red, under the green curve. The M cone has similar properties to the green channel, but the other two camera channels have peaks at longer wavelengths. Peak responses for the cones are normalized to 1. Camera responses are normalized for varying monochrome light energies and exposure times. (b) Color matching function for the Nikon D3 with the 55 mm lens.

Results for the Canon 1Ds and the Nikon D3 cameras are shown in Fig. 6. Responses are independent of lens and zoom setting on the 1Ds [19].

C. Predicting Luminance and Human Cone Responses

Given any natural spectrum, the response of the human visual system depends on the responses of human photoreceptors, not the camera response. Therefore, we would like to predict luminance and human cone responses from our camera's pixel responses. This requires four functions, $Lum_{pred}(R,G,B)$, $L_{pred}(R,G,B)$, $M_{pred}(R,G,B)$, and $S_{pred}(R,G,B)$, estimating the luminance, L cone, M cone, and S cone, respectively. If the functions are linear, then a single matrix transform can be used to represent all four functions.

The response of the R, G, and B channels to any radiance spectrum $P(\lambda)$ in W/sr/m^2/nm may be computed according to

$$R = \int C_R(\lambda)P(\lambda)\mathrm{d}\lambda, \qquad G = \int C_G(\lambda)P(\lambda)\mathrm{d}\lambda,$$

$$B = \int C_B(\lambda)P(\lambda)\mathrm{d}\lambda, \tag{7}$$

where C_X is the camera response for channel X and λ is the wavelength. The corresponding luminance response calculated for spectrum $P(\lambda)$ is

$$Lum_{calc} = V_{max} \int V(\lambda)P(\lambda)\mathrm{d}\lambda, \tag{8}$$

where $V(\lambda)$ is the CIE photopic luminous-efficiency function and $V_{max}=683$ cd/W is the maximum of $V(\lambda)$. Similarly, cone responses L_{calc}, M_{calc}, and S_{calc} can be computed by substituting the appropriate cone fundamentals for $V(\lambda)$. A parallel analysis can be done for scotopic vision by using the scotopic luminous-efficiency function $V'(\lambda)$.

The goal of luminance and cone response prediction is to develop a function that accepts (R,G,B) vectors as an argument and use them to predict (Lum,L,M,S) vectors. In linear form, this function may be represented as a 4×3 matrix $\mathbf{M}$ such that

$$(Lum_{pred},L_{pred},M_{pred},S_{pred}) = \mathbf{M}(R,G,B)^T. \tag{9}$$

We determined $\mathbf{M}$ via minimization of the error function

$$E = \sum_{i=1}^{n} |Lum_{pred,i} - Lum_{calc,i}| + |L_{pred,i} - L_{calc,i}| + |M_{pred,i} - M_{calc,i}| + |S_{pred,i} - S_{calc,i}|, \tag{10}$$

where the sum is over n natural spectra.

Alternatively, we considered the transform

$$[\log(Lum_{pred}),\log(L_{pred}),\log(M_{pred}),\log(S_{pred})] = \mathbf{M}^{\log}[\log(R),\log(G),\log(B)]^T \tag{11}$$

with error function

$$E_{\log} = \sum_{i=1}^{n} |\log(Lum_{pred,i}) - \log(Lum_{calc,i})| + |\log(L_{pred,i}) - \log(L_{calc,i})| + |\log(M_{pred,i}) - \log(L_{calc,i})| + |\log(S_{pred,i}) - \log(S_{calc,i})|. \tag{12}$$

Optimization was achieved using a sequential quadratic programming method [20,21] because it provides good results, is readily available in commercial software [22], and is accessible through the Matlab function fmincon.

The nonlog approach has the advantage that, from a radiometric perspective, it treats all energy levels equally. Also, negative infinity is avoided when cone responses are zero, and weak cone responses, combined with small measurement errors, do not appear has huge errors.

The log approach has the advantage that negative cone responses are avoided. In addition, the error as measured by Eq. (12) is more closely related to the manner in which humans perceive differences in light energy [23].

One remaining question is, What spectra would be most appropriate? We decided to choose spectra that would be relevant to the evolution of human vision, including the environments and pelage of two trichromatic primates [24–27]. In particular, we used the spectra collected by Sumner and Mollon in Uganda and French Guiana, including 30 spectra of mandrill (*Mandrillus sphinx*) pelage, 12 spectra of the vervet monkey (*Cercopithecus aethiops*), 32 fruit spectra, and 32 foliage spectra. The pelt reflective spectra were combined with a Ugandan sunny sky illuminant (4:30 PM). The fruit and foliage reflective spectra were each combined with two of the following four illuminating spectra: the Ugandan sunny illuminant, another Ugandan illuminant with overcast conditions (9:30 AM), a Guiana illuminant with sunny conditions (12:41 PM), and a Guiana illuminant with overcast conditions (6:25 PM). All illuminants had a forest canopy surround.

These environmental spectra were augmented with 12 pelt spectra of the red-ruffed lemur (*Varecia variagata*) combined with the Ugandan sunny illuminant.

The Ugandan and Guiana spectra included few blue spectra. Therefore, we added the most common environmental source for blue spectra, skylight. In particular, we used 52 skylight spectra collected by Hernández-Andrés *et al.* in Granada [28], with color temperatures ranging from 5000 K to 100,000 K.

Finally, we augmented our spectra set with some nonenvironmental spectra, namely, the spectra of 24 MacBeth Color Checker chips as measured by Ohta [29]. We illuminated these with the global sky illuminant ASTM G173-03 [30], representing sunny conditions at an average contiguous United States latitude.

A comparison of the calculated luminance and cone responses for the 258 spectra versus the estimated responses is shown in Figs. 7 and 8. The MacBeth spectra appear most difficult to fit for M cones, perhaps due the artificial origin of these spectra. The problem is not apparent in the logarithmic domain.

The log method shows an apparently large error for some foliage and fruit spectra that lack energy in the blue wavelengths. This is expected, since very weak cone responses are susceptible to small measurement errors when using the log method. However, even for these spectra, the calculated and predicted S-cone responses (in nonlog form) are both near zero, so the absolute error is not great.

6. LIMITS TO RESOLUTION AND DYNAMIC RANGE

In any real optical system, points of light in the scene do not produce points of light in the image. Real optical systems spread point light sources into patterns of finite extent. In other words, camera resolution is limited. This is the case even if the lens is perfectly corrected for aberrations. The remaining point spread is called the point-spread function and is due to diffraction. Point spreading increases with increasing focal lengths and with decreasing apertures. For more on the theory of diffraction in camera systems, see Ray [14] and Smith [31].

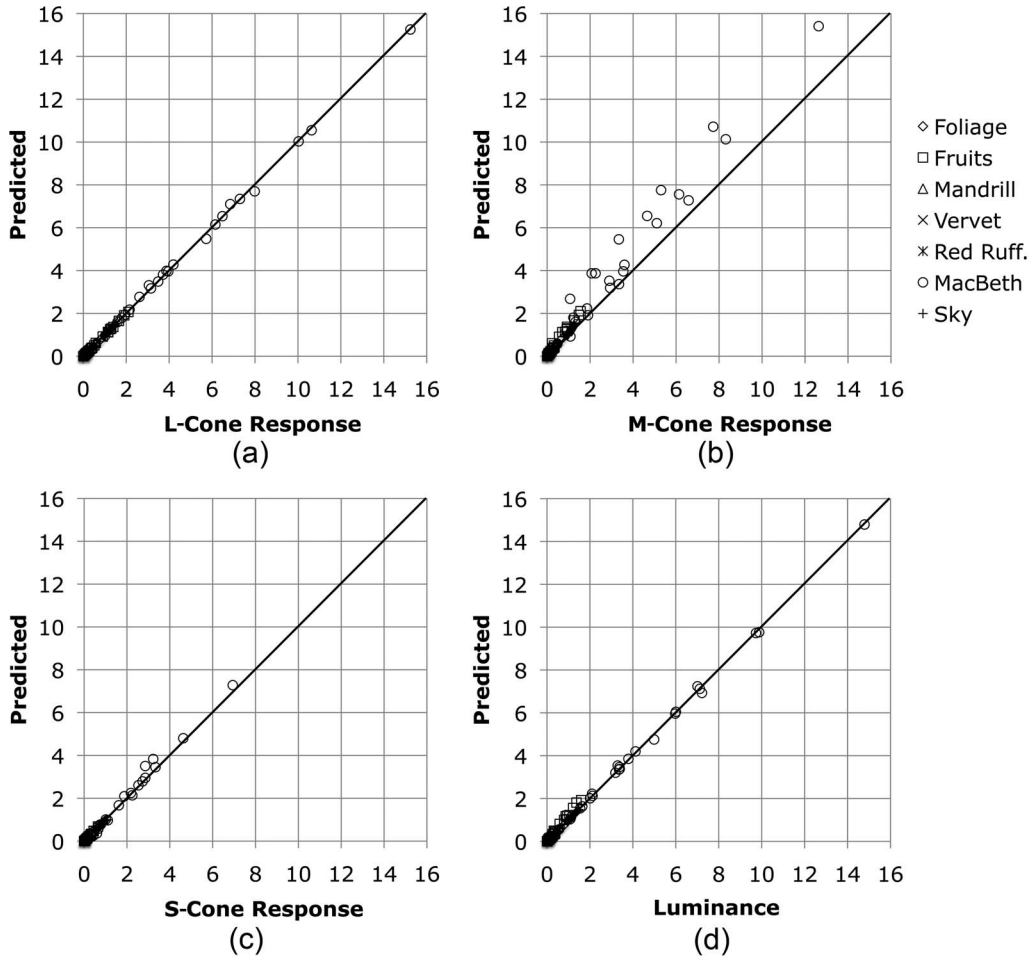


Fig. 7. Calculated and predicted cone responses to 258 spectra using the nonlog method. The points lie very close to the identity function line with the exception of M-cone responses to the spectra of artificial pigments.

A small amount of diffraction-generated point spread can benefit discrete sensor systems, which would otherwise suffer from aliasing. However, the best place to limit resolution is at the sensor, where microlenses provide just enough blurring to prevent aliasing. Blur elsewhere in the system will vary with camera settings.

Unfortunately, perfect point light sources for measures do not exist and are therefore difficult to study. However, it turns out that camera resolution can also be measured in terms of peak-to-peak attenuation of spatial sine wave patterns in an image as a function of wavelength. The lower the resolution, the faster the sine wave amplitude drops with spatial frequency. This decreasing function is called the modulation transfer function or MTF.

The importance of MTF to natural image statistics is rather direct. If one is collecting data on the contrast energy as a function of frequency in an image, knowing the MTF permits one to attribute part of the energy decrease to the camera. Whatever relation between spatial frequency and energy remains after subtracting the camera's MTF is due to properties of the scene.

MTF is defined mathematically as

$$M = \frac{M_i}{M_o}, \quad \text{where} \quad M_i = \frac{E_{i\,max} - E_{i\,min}}{E_{i\,max} + E_{i\,min}} \quad \text{and}$$

$$M_o = \frac{E_{o\,max} - E_{o\,min}}{E_{o\,max} + E_{o\,min}}. \tag{13}$$

Here $E_{i\,max}$ is the peak energy in the image pattern, $E_{i\,min}$ is the minimum energy in the image pattern, $E_{o\,max}$ is the peak energy reflected from the test object surface, and $E_{o\,min}$ is the minimum energy reflected from the test object surface; M_i is the image modulation, and M_o is the object reflectance modulation. We use the sine wave pattern chart from Sine Pattern, LLC [32], with its multiple sine wave frequencies as our object. Then M can be calculated for each spatial frequency, thus defining the MTF for the camera being calibrated.

Any actual measurement of MTF involves a confound among aberrations, diffraction, sensor element sampling, stray light, and an unknown demosaicing algorithm. In this section we show how to sort out the relative influences of these factors.

By using fully raw data, one can dispense with the problem of proprietary demosaicing algorithms. In this analysis we measure MTF for each color channel individually. This is done by generating a single channel pixel from each Bayer pattern [33] element as follows: the red element in the Bayer pattern represents the red channel,

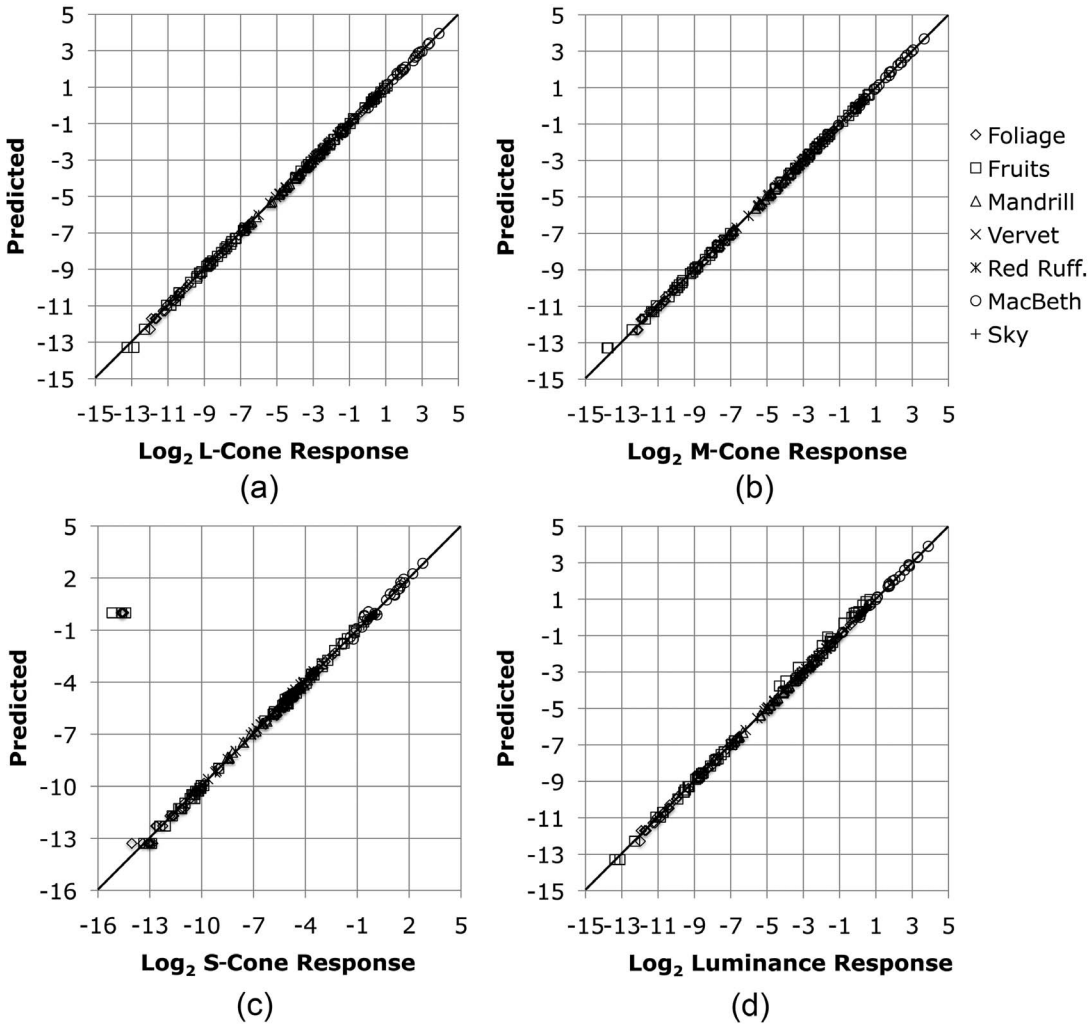


Fig. 8. Log-calculated and log-predicted cone responses to 258 spectra using the log method. The points lie very close to the identity function line with the exception of the weak *S*-cone responses to certain fruits and foliage. Even in this case, the nonlog difference is not great.

similarly with blue, and the two green elements are averaged to form a single pixel. Alternatively, the green elements can be used without averaging to form an image with twice the sampling resolution of the red and blue. If demosaicing is desired in the application of the camera, a known demosaicing algorithm can be applied and the resulting MTF compared to the no-demosaicing MTF, thus determining the contribution of demosaicing to MTF.

Demosaicing algorithms of all sorts depend upon interpolation to fill in color X (say, red) values at elements with Y colored (say, blue) filters. These interpolations depend in turn upon the phase difference between the colored spatial-frequency patterns in the optical image and the phase of the sensor array. Such colored patterns, with spatial frequencies near that of the sensor element array, will invariably introduce small amounts of chromatic aliasing.

Of the remaining factors, only one is expected to be independent of cycles per degree, and that is stray light or flare. Therefore, the effect of flare appears as an asymptote, as shown in Fig. 9(a). Normally, flare is measured as the optical flare factor (F),

$$F = \frac{E_{i\,max} - E_{i\,min}}{E_{o\,max} - E_{o\,min}}, \tag{14}$$

which is related to the asymptote of the MTF (M_a) by

$$F = M_a \frac{E_{i\,max} + E_{i\,min}}{E_{o\,max} + E_{o\,min}}. \tag{15}$$

For the Canon D30 with a 28–105 mm lens, we find $M_a = 0.89$, 0.93, and 0.97 for red, green, and blue, respectively. Characterizing flare in this way is efficient in that it is accomplished in conjunction with MTF measurements. However, stray light is generally structured, in that points of high luminance in a scene may be projected nonuniformly throughout the image. Therefore, some means of measuring flare distribution must be used in order to fully characterize and correct for the effects of flare. However, a complete description of such methods is beyond the scope of this paper. A method for measuring the structure of flare distribution has been described by Martin [34].

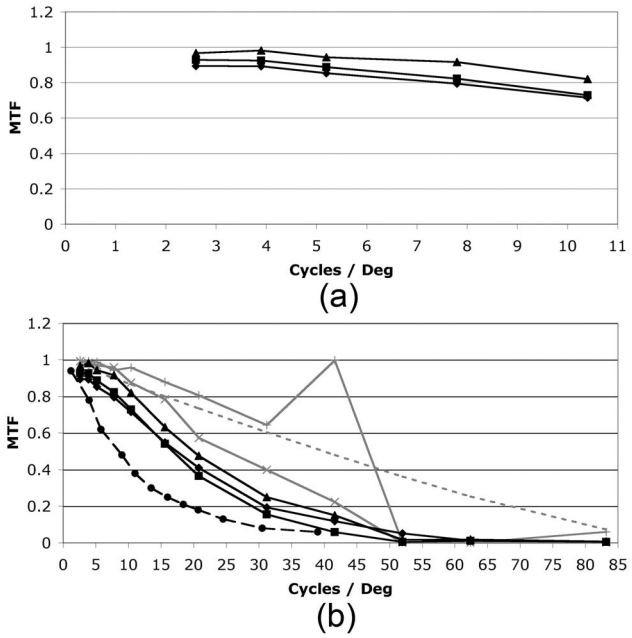


Fig. 9. MTF for the D30 camera with the 28–105 mm lens, a focal length of 105 mm, and an aperture of $f/22$. (a) Detail of MTF for low cycles per degree, showing asymptotic behavior. (b) MTF for red (diamonds), green (squares), and blue channels (triangles) compared to theoretical point sampling (upper gray), area sampling (lower gray), diffraction only (dotted), and human MTF at the 2.5 mm pupil (dashed with circles). Human data adapted from Artal and Navarro [35].

In a system free of flare, diffraction, and aberrations, the expected MTF due to sampling alone can be computed. We consider two cases: the theoretical case of point sampling and the case of sensor filling area sampling. In order to compute the MTF for the point-sampled case, the mean local $E_{i\,max}$ and $E_{i\,min}$ are taken from a point-sampled $\sin(x)$ object function, with sampling at the frequency determined by the Bayer pattern frequency. To compute the space filling MTF, $E_{i\,max}$ and $E_{i\,min}$ are taken from a step function derived from $\sin(x)$, where intervals of $\sin(x)$ are averaged over the Bayer pattern width to form the steps.

Since each sensor element occupies $\leq 1/4$ of a Bayer pattern, the actual sensor array is somewhere between these two extremes. However, the sensor array of the D30 is covered by a low-pass filter (as are the sensors of the 1Ds and D30), which means that each element is integrating light from an area greater than its 1/4 Bayer region. As a result of the low-pass sensor filter, the empirical MTFs of the three color channels follow the area sampling curve most closely, as shown in Fig. 9(b). This correspondence also indicates that the MTF of this camera–lens combination is nearly sampling limited. Points of increasing modulation in the point-sampling case are due to aliasing.

Given a design free of all factors except for diffraction, the resulting MTF can be computed as

$$\mathrm{MTF}(c) = \frac{2}{\pi}\,\mathrm{acos}(\lambda cf) - (\lambda cf)\sin[\mathrm{acos}(\lambda cf)], \qquad (16)$$

where c is cycles/mm on the sensor, λ is the wavelength (we use a middle wavelength of 530 nm), and f is the f-stop of the lens [29]. The resulting function for the D30 at f-stop=22 is shown in Fig. 9(b). It is obvious that this setup is not diffraction limited, which helps prevent aliasing.

Finally, we can compare the camera MTF to human MTF. Human MTF varies with a pupil size. Comparing our MTF to human MTF with near optimal pupil diameter of 2.5 mm [32], Fig. 9(b) shows that our setup can meet or exceed the MTF of the human eye [35].

7. TESTING THE CAMERA AS A PHOTOMETER

We tested the 1Ds camera, calibrated as a photometer, using a dozen spectra. These spectra were chosen on the basis of variety and avoiding any artificial pigments. The luminance calculations were compared with readings from a Minolta LS-110 photometer for both the $\mathbf{M}$ and the $\mathbf{M}^{\log}$ transforms. See Fig. 10. The test spectra were selected in a quasi-random fashion and included foliage (cherry in shade and oak in shade), rusted iron (shade), yellow squash (sun and shade), bone (shade), cottontail rabbit pelage (shade), flower (peach colored in sun), stone (red in sun), human skin (Asian, sun and shade), and human lip (Caucasian, shade). Images were collected at 3:30 PM, August 1 at a latitude of 46.9, under a cloudless sky. Many of these spectral samples differed significantly in origin from those used to calibrate the camera. Nevertheless, both the log and nonlog methods performed well.

In order to use the camera as a photometer, or to predict human cone responses, using exposures, apertures, and gains different from those used during calibration,

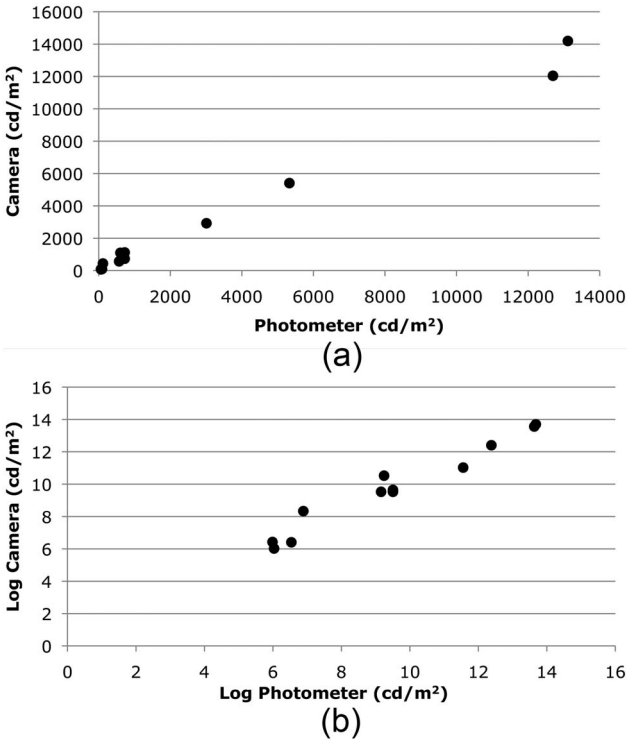


Fig. 10. The camera as a photometer. Camera response as a function of photometer reading for 12 test spectra using (a) the nonlog method matrix $\mathbf{M}$ and (b) the log method matrix $\mathbf{M}^{\log}$. The overall performance is similar for the two methods.

Table 1. Effects of Camera Configurations on Calibration Procedures[a]

	Camera Configuration Parameter					
Calibration Type	Lens	Zoom	Exposure	Aperture	ISO	Object Distance
Lateral chromatic aberration	D	D	I	N	I	N
Longitudinal chromatic aberration	D	I	I	N	I	N
Distortion	D	D	I	N	I	D
Vignetting	D	D	I	D	I	D
Dark current	I	I	D	I	D	I
OECF	I:Canon D: Nikon	I:Canon N:Nikon	A	A	A	I
Color matching function	I:Canon N:Nikon	I:Canon N:Nikon	A	A	A	I
Predicted cone and luminance response	I: Canon N:Nikon	I: Canon N:Nikon	A	A	A	I
MTF	D	D	I	D	I	D
Flare	D	D	I	D	I	I

[a]D, procedure depends on parameter; I, procedure is independent of parameter; A, procedure is dependent but a mathematical adjustment is given in the paper; N, not specifically tested or unknown; Canon, Canon 1Ds with test lenses; and Nikon, Nikon D3 with test lenses.

adjustments must be made to the (R,G,B) measures as described in Subsection 5.A. The photographer should verify that these adjustments give constant results when various camera adjustments are applied to a constant stimulus.

8. SUMMARY

In this paper, we have described a number of camera calibration procedures. Each resulting calibration is valid under a range of camera parameter settings. However, some settings will affect the completed calibration. These relationships are mentioned throughout this paper and are summarized in Table 1.

Digital cameras have long been used in vision research and computer vision. However, the issue of how specific properties of camera images relate to the biological observer has been largely ignored. In this article, we have explained how to correct a number of camera-specific artifacts and how to transform some of the remaining camera properties to better match those of the human observer. Although, more remains to be done in this critical area, the methods described here provide an essential foundation for future development.

ACKNOWLEDGMENTS

The authors thank Dwight Burkhardt and Jonathon George for their valuable assistance. This publication was made possible by grant P20-RR-020-151 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).

REFERENCES AND NOTES

1. E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1216 (2001).
2. M. J. Brady, "The Cottage Grove and Ruffner Mountain data bases," http://www.psych.ndsu.nodak.edu/brady/.
3. P. Askey, "Canon EOS 5D: dynamic range," http://www.dpreview.com/reviews/canoneos5d/page23.asp.
4. R. N. Clark, "Dynamic range and transfer functions of digital images and comparison to film," http://www.clarkvision.com/imagedetail/dynamicrange2/.
5. P. Askey, "Canon EOS-1 Ds Mark II: ISO sensitivity/noise levels," http://www.dpreview.com/reviews/canoneos1dsmkii/page19.asp.
6. R. N. Clark, "The signal to noise ratio of digital cameras and comparison to film," http://clarkvision.com/imagedetail/digital.signal.to.noise/.
7. R. N. Clark, "Film versus digital information," http://www.clarkvision.com/imagedetail/film.vs.digital.1.html.
8. T. Knoll, *Photoshop CS* (Adobe, 2003).
9. D. Coffin, "dcraw," http://cybercom.net/~dcoffin/dcraw, 2008.
10. I. Vertex Video Systems, *Accu-Chart Set* (Vertex Video Systems, 1990).
11. E. Hecht and A. Zajac, *Optics* (Addison-Wesley, 1974).
12. H. Dersch, "Correcting barrel distortion," http://home.no.net/dmaurer/%7Edersch/barrel/barrel.html.
13. K. Kratzke, "LensFix," Kekus Digital, LLC, 2005.
14. S. F. Ray, *Applied Photographic Optics*, 1st ed. (Focal Press, 2002).
15. Canon, "Noise reduction technology," http://www.canon.co.jp/Imaging/cmos/technology-e/noise_reduction.html.
16. International Organization for Standardization, "Graphic technology and photography—colour characterisation of digital still cameras (DSCs)—part 1: stimuli, metrology and test procedures," ISO publ. 14524:1999(E) (ISO, 1999).
17. International Organization for Standardization, "Photography—electronic still-picture cameras—methods for measuring optoelectronic conversion functions (OECFs)," ISO publ. ISO 17321–1:2006(E) (ISO, 2006).
18. F. Martínez-Verdú, J. Pujol, and P. Capilla, "Calculation of the color matching functions of digital cameras from their complete spectral sensitivities," J. Imaging Sci. Technol. **46**, 15–25 (2002).
19. A. Stockman and L. T. Sharpe, "Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype," Vision Res. **40**, 1711–1737 (2000).
20. R. Fletcher and M. J. Powell, "A rapidly convergent descent method for minimization," Comput. J. **6**, 163–168 (1963).

21. D. Goldfarb, "A family of variable metric updates derived by variational means," Math. Comput. **24**, 23–26 (1970).
22. C. Moler, *Matlab* (The Mathworks, 2003).
23. G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. (Wiley, 1982).
24. P. Sumner and J. D. Mollon, "Catarrhine photo pigments are optimized for detecting targets against a foliage background," J. Exp. Biol. **203**, 1963–1986 (2000).
25. P. Sumner and J. D. Mollon, "Chromaticity as a signal of ripeness in fruits taken by primates," J. Exp. Biol. **203**, 1987–2000 (2000).
26. P. Sumner and J. D. Mollon, "Did primate trichromacy evolve for fruivory or folivory?" in *Normal and Defective Color Vision*, J. D. Mollon, J. Pokorny, and K. Knoblauch, eds. (Oxford U. Press, 2003), pp. 21–30.
27. P. Sumner and J. D. Mollon, "The colors of primate pelage and skin: Objective assessment of conspicuousness," Am. J. Primatol. **59**, 67–91 (2003).
28. J. Hernández-Andrés, J. Romero, and R. L. Lee, "Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear sky skylight in Granada, Spain," J. Opt. Soc. Am. A **18**, 412–420 (2001).
29. N. Ohta, *The Basis of Color Reproduction Engineering* (Corona-sha Co., 1997).
30. Subcommittee-G03.09, "ASTM G173 03e1 Standard reference tables for reference solar spectral irradiances: direct normal and hemispherical on 37° tilted surface," (American Society for Testing and Materials, 2008).
31. W. J. Smith, *Modern Optical Engineering*, 3rd ed. (McGraw-Hill, 2000).
32. Sine Patterns, "MTF Products," http://www.sinepatterns.com/i_MTFprd.htm.
33. The Bayer pattern is a 2×2 matrix with green sensor elements on the diagonal, with the remaining elements being one red and one blue.
34. S. Martin, "Glare characteristics of lenses," J. Mod. Opt. **11**, 499–513 (1972).
35. P. Artal and R. Navarro, "Monchromatic modulation transfer function of the human eye for different pupil diameters: an analytic expression," J. Opt. Soc. Am. A **11**, 246–249 (1994).