Natural Images from the Birthplace of the Human Eye

Gašper Tkačik1,2*, Patrick Garrigan3, Charles Ratliff4, Grega Miličinski5, Jennifer M. Klein6, Lucia H. Seyfarth6, Peter Sterling6, David H. Brainard7, Vijay Balasubramanian1

1 Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 2 Institute of Science and Technology Austria, Klosterneuburg, Austria, 3 Department of Psychology, Saint Joseph’s University, Philadelphia, Pennsylvania, United States of America, 4 Department of Ophthalmology, Northwestern University, Chicago, Illinois, United States of America, 5 Sinergise d.o.o, Ljubljana, Slovenia, 6 Department of Neuroscience, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, United States of America, 7 Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Abstract

Here we introduce a database of calibrated natural images publicly available through an easy-to-use web interface. Using a Nikon D70 digital SLR camera, we acquired about 5000 six-megapixel images of Okavango Delta of Botswana, a tropical savanna habitat similar to where the human eye is thought to have evolved. Some sequences of images were captured unsystematically while following a baboon troop, while others were designed to vary a single parameter such as aperture, object distance, time of day or position on the horizon. Images are available in the raw RGB format and in grayscale. Images are also available in units relevant to the physiology of human cone photoreceptors, where pixel values represent the expected number of photoisomerizations per second for cones sensitive to long (L), medium (M) and short (S) wavelengths. This database is distributed under a Creative Commons Attribution-Noncommercial Unported license to facilitate research in computer vision, psychophysics of perception, and visual neuroscience.

Introduction

High-resolution digital cameras are now ubiquitous and affordable, and are increasingly incorporated into portable computers, mobile phones and other handheld devices. This accessibility has led to an explosion of online image databases, accessible through photo-sharing websites such as Flickr and SmugMug, social networks such as Facebook, and many other internet sites. Disciplines such as neuroscience, computer science, engineering and psychology have profited from research into the statistical properties of natural image ensembles [1,2], and it might seem that the side benefit of public photo-sharing websites is an ample supply of image data for such research. Some research, however, requires carefully calibrated images that accurately represent the light that reaches the camera sensor. For example, to address questions about the early visual system, images should not deviate systematically from the patterns of light incident onto, and encoded by, the retina. Specific limitations of uncontrolled databases that hinder their use in vision research include (i) compression by lossy algorithms, which distort image structure at fine scales; (ii) photography with different cameras, which results in unpredictable quality and incomparable pixel values; (iii) photography with different lenses or focal lengths, which may introduce unknown optical properties to the image incident on the sensor, and (iv) photography for unspecified purposes, which may bias image content toward faces, man-made structures, panoramic landscapes, etc., while under-representing sky, ground, feces, or other more mundane content.

Due to these limitations, research involving natural images typically relies on a well-calibrated database [3–5]. Examples where good image databases are essential include early visual processing in neural systems [4,6–11], computer vision algorithms [12–14], and image compression/reconstruction methods [15,16]. In each of these fields, much of the research requires accurate characterization of the statistics of natural image ensembles. Natural images exhibit characteristic luminance distributions [17] and spatial correlations [7,18]. These result from environmental regularities such as physical laws of projection and image formation, natural light sources, and the reflective properties of natural objects. Further, natural images exhibit higher-order regularities such as edges, shapes, contours and textures that are perceptually salient, but difficult to quantify [19–24]. To characterize these regularities in image patches of increasing spatial extent, the number of image samples required to collect reliable joint pixel luminance probability distributions scales exponentially with the number of pixels in the patch, ultimately limiting the reliability of our estimates. Nevertheless, with sizable databases one might push the sampling limit to patches of up to ~10 x 10 pixels discretized into a few luminance levels, and even at that restricted size interesting results have emerged [23–27].

In this paper we describe a collection of calibrated natural images, which tries to address the shortcomings of uncontrolled image databases, while still providing substantial sampling power for image ensemble research. The database is organized into themed sub-collections and is broadly annotated with keywords and tags. The raw images are linear with incident light intensity in
each of the camera’s three color channels (Red, Green, and Blue). Further, because much research is concerned with human vision, we have translated each image into physiological units relevant to human cone photoreceptors. The raw image data, demosaiced RGB images, grayscale images representing luminance, and images in the physiologically relevant cone representation are available in the database.

Results

Image calibration

We checked (i) that the Nikon D70 camera has a consistent resolution (in pixels per degree) in the vertical and horizontal directions; (ii) that its sensor responses are linear across a large range of incident light intensities; and (iii) that its shutter, aperture, and ISO settings behave regularly, so that it is possible to estimate incident photon flux given these settings. We (iv) measured the spectral response of the camera’s R, G, and B sensors by taking images of a reflectance standard illuminated by a series of 31 narrowband light sources and verified that these measured responses allowed us to predict the R, G, and B responses to spectrally broadband light. We also (v) characterized the “dark response” of the camera, i.e. the sensor output with no light incident onto the CCD (charge-coupled device) chip. Finally, we (vi) measured the spatial modulation transfer function of each of the R, G, and B sensor planes for broadband incident light. We examined two separate D70 cameras, and found that they yielded consistent results and could be used interchangeably, except for a single constant scaling factor in the CCD sensor response. Taken together, the camera measurements allow us to transform raw camera RGB values into standardized RGB response values that are proportional to incident photon fluxes seen by the three camera sensor types. Details on the camera characterization are provided in Materials and Methods: Camera response properties.

With these results in hand, we can further transform the standardized RGB data into physiologically relevant quantities (see details in Materials and Methods: Colorimetry). The photon flux arriving at the human eye is filtered by passage through the ocular media before entering a cone photoreceptor aperture. There, with some probability, each photon may cause an isomerization. There are three classes of cones, the L, M, and S cones. Each contains a different photopigment, with the spectral sensitivities of the L, M, and S cones, respectively [28]. We have transformed the RGB images into physiologically-relevant estimates of the isomerization rates that the incident light would produce in human L, M, and S cones. For this transformation, we used the Stockman-Sharpie/CIE 2-degree (foveal) estimates of the cone fundamentals [28,29], together with methods outlined by Yin et al. [30] to convert nominal cone coordinates to estimates of photopigment isomerization rates. We have also transformed the RGB images into grayscale images representing estimates of the luminance of the incident light (in units of candela per meter squared with respect to the CIE 2007 two-degree specification for photopic luminance spectral sensitivity). An important contribution of this work is to report image data from a biologically relevant environment in units of cone photopigment isomerizations, thus characterizing the information available at the first stage of the human visual processing pathways.

Image ensemble

We have assembled a large database of natural images acquired with a calibrated camera (see Materials and Methods), and made it available to the general public through an easy-to-use web interface. The extensive dataset covers a single environment: a rich riverine/savanna habitat in the Okavango delta, Botswana, which is home to the full panoply of vertebrate and invertebrate species, such as lion, leopard, cheetah, elephant, warthog, antelope, zebra, giraffe, various bird species etc. This environment was chosen because it is thought to be similar to the environment where the human eye, and retina in particular, have evolved. The database consists of ~5000 images, organized into about 100 folders (albums).

Tables 1 and 2 provide an album-by-album summary of the database content, along with content keywords for each album and tags that give additional information about how the images were gathered. Figure 1 shows several examples of images from the database: baboon habitat, panoramic images that include the horizon, closeups of the ground, and closeup images that include the ruler for absolute scale determination.

Figure 2 shows a simple analysis of images from album cd32b. Images of cloudless sky were taken every 10 minutes, from 6:30 until 18:30. We compute the average luminance as a function of time, and the LMS color composition of the light incident on the camera that was pointing up at the homogenous sky.

Figure 3 shows an analysis of 23 images from album cd04b, where closeups of the grass scrub on the ground were taken from various distances. We computed the pairwise luminance correlation function as a function of pixel separation, and show how it varies systematically with distance to the ground. Since the grass scrub on our images has a preferred scale, the correlation function should decay faster the farther away from the scene the camera is, and this is indeed what we see. Scale invariance in natural scenes presumably emerges because of the distribution of object sizes and distances from which the objects are viewed [31]. Our image collection can be used to study the scaling properties of the ensemble systematically.

The two examples provided above are intended to illustrate the strengths of our database: calibration and normalization into physiologically relevant units; organization into thematic (keywords) and methodological sequences (tags), that explore variations between the scenery content and variations induced by systematic changes in controllable parameters; and a thorough sampling that should suffice for the estimation of higher-order statistics or even accumulation of image patch probability distributions.

In assembling the database, we could have taken an alternative approach and built a catalogue of as many objects from the environment as possible, photographed at a chosen “standard” camera position and controlled illumination, removed from the natural context and placed on a neutral background. While there are advantages to exhaustively pursuing this approach, we regarded it as focusing on natural objects and not on natural scenes. Nevertheless we provide a limited set of albums in the collection of Table 2, where closeups of fruit, grass, bark are provided; we reasoned that some of these closeups would be of use for studying properties of natural objects such as texture.

Data access

The image database is accessible at http://tofu.psych.upenn.edu/~upennidb, or through anonymous FTP at ftp://anonymous@tofu.psych.upenn.edu/fulldb. A standard gallery program for viewing the images on the web makes browsing and downloading individual images or whole albums easy [32]. Once images have been added to the ‘cart’ and the user selects the download option, the database prompts the user to select the formats for download; the available formats are (i) NEF (raw camera sensor output, Nikon proprietary format); (ii) RGB Matlab matrix (demosaiced RGB values before dark response subtraction); (iii) LMS Matlab matrix (physiological...
Table 1. Albums 01a–61a of the Botswana dataset.

| Album  | Keywords                      | Tags                  |
|--------|-------------------------------|-----------------------|
| cd01a  | baboons, grass, tress, bushes |                       |
| cd02a  | baboons, grass, tress, bushes |                       |
| cd03a  | trees                          | sequence: vertical angle |
| cd04a  | trees, grass                   | sequence: aperture    |
| cd05a  | trees, leaves                  | sequence: aperture    |
| cd06a  | flood plain, grass, trees, sky |                       |
| cd07a  | baboons, horizon               | sequence: aperture    |
| cd08a  | sand plain, horizon            | sequence: aperture    |
| cd09a  | sand plain, bushes, horizon    | sequence: aperture    |
| cd10a  | bushes, trees                  | sequence: aperture    |
| cd11a  | forest                         | sequence: aperture    |
| cd12a  | forest, ground                 |                       |
| cd13a  | flood plain, water, horizon    |                       |
| cd14a  | horizon, sunset                | sequence: time, short +long exposure |
| cd15a  | leaves, grass, ground          |                       |
| cd16a  | road, cars, buildings, town    |                       |
| cd17a  | road, cars, buildings, town    |                       |
| cd18a  | tree, ground                   | sequence: aperture    |
| cd19a  | sand plain, ground             | closeup               |
| cd20a  | sand plain, horizon            | sequence: aperture    |
| cd21a  | sand plain, sand ground, cloudy sky |                       |
| cd22a  | baboons, sand ground, horizon  | sequence: aperture    |
| cd23a  | leaves, bushes, ground, sky    |                       |
| cd24a  | leaves                          | sequence: aperture    |
| cd25a  | plant, leaves                   | closeup               |
| cd26a  | leaves, bushes                  | sequence: aperture    |
| cd27a  | flood plain, grass, water, horizon | sequence: time, short+long exposure |
| cd28a  | termite mound, horizon          | sequence: aperture    |
| cd29a  | sand plain, horizon             | sequence: aperture    |
| cd30a  | baboons, sand plain, bushes, trees, horizon | sequence: aperture |
| cd31a  | bushes, termite mound           | sequence: scale, sequence: vertical angle |
| cd32a  | flood plain, water, grass, horizon |                       |
| cd33a  | flood plain, water, grass, horizon |                       |
| cd34a  | forrest, tree, leaves           | sequence: scale       |
| cd35a  | grass, trees                    | sequence: scale       |
| cd36a  | bark                            | closeup, sequence: aperture |
| cd37a  | leaves                          | sequence: aperture    |
| cd38a  | sand plain, trees, horizon, ground | sequence: aperture |
| cd39a  | sand plain, horizon             | sequence: aperture    |
| cd40a  | bark                            | closeup               |
| cd41a  | forrest, trees, leaves, ground  |                       |
| cd42a  | fruit, nuts                     | closeup, on table     |
| cd43a  | grass plain, baboons, water     |                       |
| cd44a  | baboons, grass, human           |                       |
| cd45a  | sand plain, rock                | sequence: scale       |
| cd46a  | plain, tree                     | sequence: scale       |

Table 1. Cont.

| Album  | Keywords                          | Tags                  |
|--------|-----------------------------------|-----------------------|
| cd47a  | tree trunk, forest                | sequence: scale       |
| cd48a  | tree trunk, sand plain            | sequence: scale       |
| cd49a  | tree stump, ground                | sequence: scale       |
| cd50a  | fruit, plant                      | closeup, sequence: aperture |
| cd51a  | tree, sky                         | sequence: aperture    |
| cd52a  | palm tree                         | sequence: aperture    |
| cd53a  | trees, grass, forest              | sequence: time (all day) |
| cd54a  | sand plain, horizon               | sequence: time (all day) |
| cd55a  | grass plain, horizon              | sequence: time (all day) |
| cd56a  | baboons, grass                    | closeup               |
| cd57a  | baboons, grass, trees, bushes     |                       |
| cd58a  | baboons, grass, trees, horizon    |                       |
| cd59a  | baboons, grass, tree stumps       |                       |
| cd60a  | horizon                           | sequence: time (sunset), short+long exposure |
| cd61a  | sky, moon                         | sequence: time (night) |

Keywords provide a short description of the image content. Tags provide additional information about how the images were acquired. A “sequence” tag means that the same scene was taken several times while changing a parameter, e.g., “sequence: aperture” means that the scene was photographed while changing the aperture (and simultaneously the exposure time), “sequence: scale” means that the object is photographed at decreasing distance to the camera, “sequence: time” means that the scene is photographed at approximately equal time intervals, etc. doi:10.1371/journal.pone.0020409.0001

Materials and Methods

Camera response properties

Two D70 [Nikon, Inc., Tokyo, Japan] cameras are described. Each was equipped with an AF-S DX Zoom-Nikkor 18–70 mm f/3.5–4.5G IF-ED lens. To protect the front surface of the lens, a 52 mm DMC (Digital Multi Coated) UV skylight filter was used in all measurements. One camera, referred to in this document as the standard camera (serial number 2000a9a7), is characterized in detail. A subset of the measurements was made with a second D70 camera, which will be described in a subsequent publication.
referred to in this document as the auxiliary camera (serial number 20004b72). Its response properties matched those of the standard camera after all responses were multiplied by a single constant.

**Geometric information.** The D70’s sensor provides a raw resolution of 3040(h) | 2014(v) pixels. The angular resolution of the camera was established by acquiring an image of a meter stick at a distance of 123 cm. Both horizontal and vertical angular resolutions were 92 pixels per degree. This resolution is slightly lower than that of foveal cones, which sample the image at approximately 120 cones per degree [35].

**Image quantization.** The camera documentation indicates that the D70 has native 12-bit-per-pixel intensity resolution [36], and that the NEF compressed image format supports this resolution. Nonetheless, the D70 appears not to write the raw 12-bit data to the NEF file. Rather, some quantization/compression algorithm is applied which converts the image data from 12 bit to approximately 9.4 bit per pixel resolution [37]. The raw data extraction program, DCRAW [38], appears to correct for any pixel-wise nonlinearity introduced by this processing, but it cannot, of course, recover the full 12-bit resolution. The pixel values in the file extracted by dcraw range between 0 and 16384 for red and blue, and between 0 and 16380 for the green channel, and these are the values we use for the analysis. We do not have a direct estimate of the actual precision of this representation. We used version v5.71 of dcraw; we note that the output image representation produced by this software is highly version dependent.

**Mosaic pattern and block averaging.** The D70 employs a mosaic photosensor array to provide RGB color images. That is, each pixel in a raw image corresponds either to an R, G, or B sensor, as shown in Fig. 4. R, G, and B values can then be interpolated to each pixel location. This process is known as demosaicing. The images in the database were demosaiced by

![Figure 1. Example images from the Botswana dataset. A–D) Some natural scenes from various albums, including a tree, grass and bushes environment, the horizon with a large amount of sky, and closeups of the ground; the last image is from the image set containing a ruler than can be used to infer the absolute scale of objects. E–F) The distributions of L (red), M (green) and S (blue) channel intensities across the image for images A) and B), respectively. The large sky coverage in B) causes a peak in the S channel at high values. The horizontal axis is log base 10 of pigment photoisomerizations per cone per second. G–H) Grayscale images showing log luminance corresponding to the images in C) and D), respectively. doi:10.1371/journal.pone.0020409.g001](figure1)

![Figure 2. The color content and luminance of the sky. A) The luminance in candelas per square meter shows the rise in the morning and decay in the evening, along with the fluctuations during the day. B) The color content of the sky. To report relative changes in color content corrected for overall luminance variation, the L, M and S channels have been divided by their reference values at 12:30pm to bring the three separate curves together at the 12:30pm time point. In addition, all three curves were multiplied by the luminance at 12:30pm and then divided by the luminance at the time each measurement was taken. At sunrise and sunset, the L (redder) channel is relatively more prominent, and S channel decreases sharply. doi:10.1371/journal.pone.0020409.g002](figure2)
values that after dark subtraction yielded negative values were set to 0. Figure 5A shows the dark response that was used for subtraction. Figure 5B shows the mean response, excluding the “hot” pixels, for comparison. All measurements reported below are for RGB values after dark subtraction.
Response linearity. Fundamental to digital camera calibration is a full description of how image values obtained from the camera relate to the intensity of the light incident on the sensors. To measure the D70’s intensity-response function, a white test standard was placed 168 cm from the camera. The aperture was held constant at \( f/5.6 \) and \( f/11 \) and \( f/16 \). One picture was taken for each of the 55 possible exposure durations, which ranged from 1/8000 s to 30 s. Response values were obtained by extracting and averaging RGB values from a 100 \( \times \) 100 pixel region of the image corresponding to light from the white standard. The camera saturated in at least one channel for longer exposure durations. The camera responses were linear over the duration range 1/320 s to 1/3 s for the test light level and all apertures used, and for aperture \( f/5 \) approximate linearity extended down to smallest available exposure times, as shown in Fig. 6. Deviations from linearity are seen at short exposure durations for \( f/11 \) and \( f/16 \). These presumably reflect a nonlinearity in the sensor intensity-response function at very low dark saturated response values.

ISO linearity. To examine the effect of the ISO setting on the raw camera response, we acquired a series of 10 images of the white standard with the ISO setting changing from ISO 200 to ISO 1600 at exposures of 1/125 s and 1/250 s. In the regime where the sensors did not saturate, the camera response was linear in ISO for both exposure times, as shown in Fig. 7.

Aperture test. The aperture size (f-number) of a lens affects the amount of light reaching the camera’s sensors. For the same light source, the intensity per unit area for f-number \( x \), \( I(x) \), relative to the intensity per unit area for f-number \( y \), \( I(y) \) should be given by:

\[
I(x) = (y/x)^2 I(y).
\]

We tested this by measuring the sensor response as a function of exposure duration for all color channels at three different aperture sizes \( f/5, f/11, \) and \( f/16 \). We measured the sensor responses to the white test standard at various aperture sizes to confirm that the response was inversely proportional to the square of the f-number. The camera was positioned facing the white test standard illuminated by a slide projector with a tungsten bulb in an otherwise dark room. The camera exposure time was held at 1/250 s and the ISO setting was 1000. Images of the white test standard (primary image region) were taken, one for each aperture setting between \( f/1.8 \) and \( f/22 \). Because image values were saturated at the largest apertures \( f/1.8 \) and \( f/2 \), we also extracted and analyzed image values from a less intense region in the same image series (secondary image region). The measurements, shown in Fig. 8A for the primary region and Fig. 8B for the secondary region, confirm that the aperture is operating correctly. There is some scatter of the measured points around the theoretical lines. This may be due to mechanical imprecision for each aperture. We did not pursue this effect in detail, nor attempt to correct for it. Nor did we investigate whether the slightly steeper slopes found for the blue channel represent a slight systematic deviation from expectations for the responses of that channel.

Standardized RGB values. After having verified that the sensor response scales linearly with ISO setting and the exposure, and as \( f^{-2} \) with the aperture number \( f \) across most of the camera’s dynamic range, we define the standardized RGB values as dark-subtracted raw camera RGB values, scaled to the reference value of ISO 1000, reference aperture of \( f/1.8 \) and reference exposure time of 1 s:

\[
\text{standardized_RGB} = (\text{raw_RGB} - \text{dark_response}) \times \frac{(1000/\text{ISO}) \times (f/1.8)^2 \times (1/\text{exposure})}{1}.
\]

It is the standardized RGB values that provide estimates of the light incident on the camera across our image database. Our
for $f$ measured in units of $W$ predicted standardized RGB values, the input spectral radiance sensor response to arbitrary spectral light sources. To compute Fig. 9. These may be used to generate predictions of the camera sensor dark noise. The measured spectral sensitivities are shown in account for any stray broadband light in the room as well as for occluding the light projected onto the white test standard, so that they response images were acquired. These dark images were taken by performed with dark images acquired at the same time as the spectral and treated it as concentrated at the nominal center wavelength of sensitivities. For this purpose, we summed power over wavelength read by the spectroradiometer to estimate the three sensor’s spectral of the camera sensors. A slide projector (Kodak Carousel 440 [39]), the Nikon D70 Camera, and a spectroradiometer [40] were positioned in front of the white test standard. Light from the projector was passed through one of 31 monochromatic filters, which transmitted narrowband light between 400 nm and 700 nm, at 10 nm intervals. For each filter, a digital picture ($f$, 1.8, ISO 1000, and varying exposure duration) and a spectroradiometer reading were taken. The RGB data were dark subtracted and converted to standardized RGB values, and then compared to the radiant power read by the spectroradiometer to estimate the three sensor’s spectral sensitivities. For this purpose, we summed power over wavelength and treated it as concentrated at the nominal center wavelength of each narrowband filter. For these images, dark subtraction was performed with dark images acquired at the same time as the spectral response images were acquired. These dark images were taken by occluding the light projected onto the white test standard, so that they accounted for any stray broadband light in the room as well as for sensor dark noise. The measured spectral sensitivities are shown in Fig. 9. These may be used to generate predictions of the camera sensor response to arbitrary spectral light sources. To compute predicted standardized RGB values, the input spectral radiance measured in units of $W/(nm\cdot sr\cdot m^2)$ should be integrated against each spectral sensitivity. To check the end-to-end accuracy of our camera RGB calibration, we acquired an image of the Macbeth color checker chart, extracted the raw RGB values for each of its 24 swatches, and converted these to the standardized RGB representation. We then compared these measured values against predictions obtained from direct measurements of the spectral radiance reflected from each swatch. The comparison in Fig. 10A shows an excellent fit between predicted and measured values. calibration measurements indicate that these values provide good estimates over the multiple decades of light intensity encountered in natural viewing. Scatter of measurements around the fit lines shown in the figures above does, however, introduce uncertainty of a few tenths of a log (base 10) unit in intensity estimates. Within image, optical factors are likely to introduce systematic variation in sensitivity from the center of the image to the edge, an effect that we did not characterize or correct for.

**Spectral response.** Next, we measured the spectral sensitivities of the camera sensors. A slide projector (Kodak Carousel 440 [39]), the Nikon D70 Camera, and a spectroradiometer [40] were positioned in front of the white test standard. Light from the projector was passed through one of 31 monochromatic filters, which transmitted narrowband light between 400 nm and 700 nm, at 10 nm intervals. For each filter, a digital picture ($f$, 1.8, ISO 1000, and varying exposure duration) and a spectroradiometer reading were taken. The RGB data were dark subtracted and converted to standardized RGB values, and then compared to the radiant power read by the spectroradiometer to estimate the three sensor’s spectral sensitivities. For this purpose, we summed power over wavelength and treated it as concentrated at the nominal center wavelength of each narrowband filter. For these images, dark subtraction was performed with dark images acquired at the same time as the spectral response images were acquired. These dark images were taken by occluding the light projected onto the white test standard, so that they accounted for any stray broadband light in the room as well as for sensor dark noise. The measured spectral sensitivities are shown in Fig. 9. These may be used to generate predictions of the camera sensor response to arbitrary spectral light sources. To compute predicted standardized RGB values, the input spectral radiance measured in units of $W/(nm\cdot sr\cdot m^2)$ should be integrated against each spectral sensitivity. To check the end-to-end accuracy of our camera RGB calibration, we acquired an image of the Macbeth color checker chart, extracted the raw RGB values for each of its 24 swatches, and converted these to the standardized RGB representation. We then compared these measured values against predictions obtained from direct measurements of the spectral radiance reflected from each swatch. The comparison in Fig. 10A shows an excellent fit between predicted and measured values.

**Spatial Modulation Transfer Function.** The modulation transfer function (MTF) describes how well contrast information is transmitted through an optical system. We estimated the MTF of our camera for the R, G, and B image planes individually, to compare the transmission of contrast information at various spatial

---

Figure 6. Linearity of the camera in exposure time. The mean raw RGB response after dark subtraction of the three color channels (red, green, blue; shown in corresponding colors) is plotted against the exposure time in seconds. The values are extracted from images of a white test standard at $f/5$ (A), $f/11$ (B) and $f/16$ (C) and ISO 200 settings. Full plot symbols indicate raw dark subtracted values between 50 and 16100 raw units; these data points were used to fit linear slopes to each color channel and aperture separately. The fit slopes are 1.01 (R), 1.00 (G), 1.02 (B) for $f/5$, 1.00, 0.99, 1.02 for $f/11$, and 1.01, 1.00, 1.02 for $f/16$.

Figure 7. Linearity of the camera in ISO setting. The mean raw RGB response of the three color channels (red, green, blue; shown in corresponding colors) is plotted against the ISO setting after dark subtraction, for two values of exposure time (solid line, circles = 1/125 s; dashed line, squares = 1/250 s), and $f/2.8$ aperture. The lines are linear regressions through non-saturated data points (solid squares or circles; raw dark subtracted values between 50 and 16100); the slopes are 0.99 (R), 0.98 (G), 0.99 (B) for 1/125 s exposure and 1.03, 1.02, 1.04 for 1/250 s exposure. The camera saturated in the red channel at longer exposure; the corresponding data points (empty red circles) are not included into the linear fit.

---

Spatial Modulation Transfer Function. The modulation transfer function (MTF) describes how well contrast information is transmitted through an optical system. We estimated the MTF of our camera for the R, G, and B image planes individually, to compare the transmission of contrast information at various spatial
frequencies for light intensity measured in each of the three types of camera photosensors. To do this, we first imaged a high-contrast black and white square-wave grating at a set of distances relative to the camera and extracted a horizontal grating patch from each image. This was summed over columns to produce a one-dimensional signal for each image plane. We then used the fast-fourier transform to find the amplitude and spatial frequency of the fundamental component of the square wave-grating. In this procedure, we tried various croppings of the grating and chose the one that yielded maximum amplitude at the fundamental. This minimized spread of energy in the frequency domain caused by spatial sampling. We then reconstructed a one-dimensional spatial domain image by filtering out all frequencies except for the fundamental, and computed the image contrast as the difference between minimum and maximum intensity divided by the sum of minimum and maximum intensity. Figure 11 plots image contrast as a function of spatial frequency (cycles / pixel) for each color channel. At high spatial frequencies, the blue channel is blurred least and the red channel is blurred most, presumably due to chromatic aberration in the lens. We did not use these data in our

**Figure 8. Camera response as a function of aperture size.** The raw dark subtracted response of the camera exposed to a white test standard (A) and a darker secondary image region (B), in three color channels (red, green, blue, shown in corresponding colors), as a function of the aperture (f-value), with exposure held constant to 1/250 s and ISO set to 1000. In the regime where the sensors are not saturated and responses are not very small (solid circles, raw dark subtracted values between 50 and 16100), the lines show a linear fit on a double logarithmic scale constrained to have a slope of $-2$ (i.e. $y = k x^{-2}$). Leaving the slopes as free fit parameters yields slopes of $-2.02$ (R), $-2.04$ (G), $-2.10$ (B) for the primary region (white standard) in panel A, and $-2.00$ (R), $-2.03$ (G), $-2.06$ (B) for the secondary region in panel B. Data points in the saturated or low response regime (empty circles) were not used in the fit. The maximum absolute log base 10 deviation of the measurements from the fit lines is 0.1.

doi:10.1371/journal.pone.0020409.g008

**Figure 9. Spectral response of the camera.** A) The spectral sensitivity curves plotted here convert spectral radiance into standardized camera RGB values. B) The LMS cone fundamentals [29,45] for L (red), M (green) and S (blue) cones. Note that the fundamentals are normalized to have a maximum of 1. C) A linear transformation can be found that transforms R,G,B readings from the camera with sensitivities plotted in (A) into reconstructed fundamentals L’M’S’ shown here, such that L’M’S’ fundamentals are as close as possible (in mean-squared-error sense) to the true LMS fundamentals shown in (B).

doi:10.1371/journal.pone.0020409.g009
image processing chain, but fits to the data are available (see caption of Fig. 11) either to correct for camera blurring for applications where that is desirable or to estimate the effect of camera blur on any particular image analysis. It should be noted that the MTF will depend on a number of factors, including f-stop, exactly how the image is focused, and the exact spectral composition of the incident light. The measurements were made for f/5.6 and the camera’s auto focus procedure. In addition, the measurements do not account for affects of lateral chromatic aberrations, which can produce magnification differences across the images seen by the three color channels. For these reasons, the MTF data should be viewed as an approximation.

**Camera comparisons.** We compared the response properties of two Nikon D70 cameras – the standard camera (used for the analysis detailed above), and a second auxiliary camera. We found that, under identical conditions, the sensor response of the auxiliary camera was approximately 80%±5% of the response of the standard camera, across all channels, response times, and apertures (see e.g. Fig. 12). We compared the two cameras imaging the white test standard with varying exposure durations, and for color swatches of the Macbeth color checker. We found that the ratio of camera responses was relatively constant throughout the linear range of exposure durations. To ensure that the response difference was not a result of the slightly different positions of the tripods on which each camera was mounted, we also took pictures of the white standard with the cameras on the same tripod (sequentially), then with the auxiliary camera 25 cm to the left of the standard camera and visa versa. The difference in camera responses could not be accounted for by small positional differences. Therefore, the different camera response magnitudes probably indicate a characteristic of the cameras themselves. At very fast or very slow exposure durations, the responses of both cameras were dominated by dark-response and sensor saturation, respectively, and the 80% response ratio did not apply. This analysis suggested that a complete calibration of the auxiliary camera was unnecessary. Instead, the spectral sensitivities of R, G, B sensors for the auxiliary camera were defined to be 0.80 times the spectral sensitivities of the standard camera; by using this simple multiplicative conversion to bring the auxiliary and standard cameras into accordance, the remaining camera; by using this simple multiplicative conversion to bring the auxiliary and standard cameras into accordance, the remaining image transformation steps remain unchanged between both cameras.

**Colorimetry.** For applications to human vision it is useful to convert the camera RGB representation to one that characterizes

---

**Figure 10. Checking the camera calibration.** Digital images and direct measurements of spectral radiance were obtained for the 24 color swatches of the Macbeth color checker chart. A) Raw standardized RGB values were obtained from the camera images as described in Materials and Methods. RGB response was also estimated directly from the radiometric readings via the camera spectral sensitivities shown in Fig. 9A. Plotted is the comparison of the corresponding 24 × 3 RGB values; black line denotes equality. B) The luminance in cd/m² measured directly by the radiometer compared to the luminance values obtained from the standardized camera RGB values. C) This plot shows the correspondence between the Stockman-Sharpe/CIE 2-degree LMS cone coordinates estimated from the camera and those obtained from the measured spectra. Plot symbols red, green and blue indicate L,M,S values respectively, and the data are for the 24 MCC squares.

doi:10.1371/journal.pone.0020409.g010

**Figure 11. Camera spatial MTF.** Estimated MTF is plotted as a function of spatial frequency for the red, green, and blue image planes (shown in corresponding colors). Solid lines show empirical fits to y = a exp{−(b(x−d)²) + (1−a) exp(−cx²)}, where for all x ≤ d, y is set to 1 and where any fit values y greater than 1 were also set to 1. The fit parameters are a = 1.00, b = 0.96, c = 0.50, d = 0.10 for red channel, a = 0.52, b = 12.37, c = 0.02, d = 0.11 for green channel, and a = 1.00, b = 2.24, c = 0.50, d = 0.12 for blue channel. MTF values at ~0.33 cycles/pixel (empty plot symbols) systematically deviated from the rest and were excluded from the fit.

doi:10.1371/journal.pone.0020409.g011
Figure 12. Comparison between standard and auxiliary camera. Raw dark subtracted values for three color channels (red, green, blue; shown in corresponding colors) of the same standard taken at different exposures using the standard and auxiliary cameras. Black line is a linear fit on non-saturated points (raw dark subtracted pixel values between 50 and 16100), with fit slope of 0.84. Images were acquired at f/5.6 and ISO 400. Similar slopes (0.8 ± 0.05) were found when positions of cameras imaging the white standard were slightly changed (see text), and when the camera readouts were compared on color swatches of the Macbeth color checker (f/2.8, ISO 400).

doi:10.1371/journal.pone.0020409.g012

Table 3. Cone Parameters for LMS Images.

| Cone Parameter | Value  | Source          |
|----------------|--------|-----------------|
| Outer Segment Length | 33 μm | [61] Appendix B |
| Inner Segment Length | 2.3 μm | [61] Appendix B |
| Specific Density | 0.5 (axial optical density) | [61] Appendix B |
| Lens Transmittance | See CVRL database | [62] |
| Macular Transmittance | See CVRL database | [63] |
| Pupil Diameter | See [64, Eq.1] [64] |
| Eye Length | 16.1 mm | [61] Appendix B |
| Photoreceptor Nomogram | See CVRL database | [28] |
| Photoreceptor Quantal Efficiency | 0.667 | [61] page 472 |

CVRL database is accessible at http://www.cvrl.org/.
doi:10.1371/journal.pone.0020409.t003

To determine the constant of proportionality required to obtain LMS isomerization rates from the LMS cone coordinates for each cone class, we estimated photopigment isomerization rates directly from the measured spectral radiance of the light reflected from the 24 color swatches of the Macbeth color checker. This was done using software available as part of the Psychophysics Toolbox [30,46] along with the parameter values for human vision provided in Table 3. We also computed the response of the LMS cones from these same measured spectra using the cone response functions plotted in Fig. 9B. For each cone class, we then regressed the 24 isomerization rates against the 24 corresponding cone coordinates to obtain the scalar required to transform cone coordinate to isomerization rate. For the L, M, and S cones respectively, the scalar values are 1.75×10^3,1.60×10^3,3.49×10^4, with units that yield isomerizations per human cone per second.

For certain analyses, it is useful to provide grayscale versions of the images. We thus computed estimates of the luminance Y (in units of candelas per meter squared with respect to the CIE 2007 two-degree specification for photopic luminance spectral sensitivity [29]). These were obtained as a weighted sum of the estimated Stockman-Sharpe/CIE 2-degree LMS cone coordinates:

Y = 433.94L + 275.82M − 0.09S;

these weights yield luminance in units of cd/m².

Figure 10B shows the comparison between the luminance for the Macbeth color checker and the luminance measured directly using the radiometer. Figure 10C shows the comparison between the Stockman-Sharpe/CIE 2-degree LMS cone coordinates estimated from the camera and those computed directly from the radiometrically measured spectra. Overall the agreement is good, with the exception of the S-cone coordinates for a few color checker squares. These deviations occur because the camera spectral sensitivities are not an exact linear transformation of the cone fundamentals.

Image extraction and data formats. For each image in every database album (folder), the following operations are carried out:

1. Raw image .NEF to .PPM conversion. Images with name pattern DSC_#####.NEF (where # are image serial numbers) were first extracted from the camera as proprietary Nikon Electronic Format (NEF) files (approximately 6 Mb in size), which record “raw” sensor values. These files were converted to PPM files using dcraw v5.71 [38] in Document Mode (no color interpolation between RGB sensors) by using the −d flag, and written out as 48 bits-per-pixel (16 bits per color channel) PPM files by using the −f flag; we note that the behavior of dcraw is highly version dependent. In addition, we extracted the following image meta-data from the NEF file: the camera serial number (to determine whether an image was taken using the standard or auxiliary camera), exposure duration, f-value, and the ISO setting. We also extracted JPG formatted images of the NEF images; both JPG and NEF formats are available from the database as DSC_#####.NEF and DSC_#####.JPG.

2. PPM to raw RGB conversion. PPM images were loaded by our Matlab script as 2014×3038 matrices. Because of the
CCD mosaic, pixels in this matrix that have no corresponding CCD sensor contain zeros. Subsequently, each color plane of the image was block-averaged in 2 × 2 blocks into a resulting 1007 × 1519 × 3 raw RGB format, where the last dimension indicates the color channel (R = 1, G = 2, B = 3); red and blue pixels are weighted by 4 and green pixels by 2, reflecting the number of sensors for each color in every 2 × 2 block. The resulting matrix has raw units spanning the range from 0 to 16384 for red and blue channels, and from 0 to 16380 for the green channel. It is available as a Matlab matrix with name pattern DSC_######_RGB.mat from the image database.

3. Raw RGB to standardized, dark-subtracted RGB. We next took the resulting block-averaged RGB images and subtracted the dark response, characteristic for each color channel and for the image exposure duration, using the dark response values shown in Fig. 5A. After subtracting the dark response, each image was multiplied by the scale factor, to yield a RGB dark-subtracted image standardized to one second exposure duration, ISO 1000 setting, and f-value of 1.8, as prescribed by Eq (1).

4. Standardized, dark-subtracted RGB to LMS and Luminance formats. By regressing R,G,B camera sensitivities (measured in Fig. 9) against known LMS cone sensitivities, we obtain a 3 × 3 matrix that approximately transforms from the RGB into the LMS color coordinate system. Each pixel of the image can thus be transformed into the LMS system. By multiplying the image in the LMS system with the LMS isomerization rate factors (see Colorimetry above), each image is now expressed in units of LMS isomerizations per second and saved as a Matlab 1007 × 1519 × 3 matrix with name pattern DSC_######_LUM.mat. In parallel, the image in the LMS system can be transformed into a grayscale image with pixels in units of cd/m², by summing over the three (L, M, S) color channels of each pixel with appropriate weights (see Colorimetry above). The grayscale image is saved as a 1007 × 1519 matrix with name pattern DSC_######_LUM.mat.

5. Image metadata in AUX files. For each image, the database contains a small auxiliary Matlab structure saved as DSC_######_AUX.mat that contains image meta-data: (i) the EXIF fields extracted from the JPG and NEF images, among others, the aperture, exposure settings, ISO sensitivity, camera serial number, and timestamp of the image; (ii) the identifier of the image in the master database (the image name and album); (iii) the timestamp when the image processing was done, (iv) fraction of pixels at saturation in the RGB image, (v) a warning flag indicating whether the exposure time is out of the linear range or if too many pixels are either at low (dark-response) intensities or at saturation, and (vi) several human-annotated annotations (not available for all images), e.g. the distance to the target, the tripod setting, qualitative time of day, lighting conditions etc. Most images are taken in normal, daylight viewing conditions, and should therefore have no warning flags; some images, however, especially ones that are part of the exposure or time-of-day series, can have saturated or dark pixels, and it is up to the database users to properly handle such images.

References
1. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–216.
2. Geisler WS, Raigach D (2009) Natural systems analysis. Introduction. Vis Neurosci 26: 1–3.
3. Olmos A, Kingdom FAA (2004) McGill calibrated colour image database. URL http://tabley.vision.mcgill.ca. Last accessed 2011-05-08.
4. van Hateren JH (1992) Real and optimal neural images in early vision. Nature 360: 68–70.
5. Parraga CA, Vazquez-Corral J, Varela M (2009) A new cone activation-based natural images dataset. Perception 36(Suppl): 180.
6. Atick J, Redlich A (1992) What does the retina know about natural scenes? Neural Computation 4: 196–210.
7. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A 4: 2379–94.
8. Doi E, Balcan DC, Lewicki MS (2007) Robust coding over noisy overcomplete channels. IEEE Transactions on Image Processing 16: 442–452.
9. Karklin Y, Lewicki MS (2009) Emergence of complex cell properties by learning to generalize in natural scenes. Nature 457: 38–46.
10. Ratliff CP, Borrquis BG, Kao YH, Sterling P, Balasubramanian V (2010) Retina is structured to process an excess of darkness in natural scenes. Proc Natl Acad Sci U S A 107: 17360–73.
11. Garrigan P, Ratliff CP, Klein JM, Sterling P, Brainard DH, et al. (2010) Design of a trichromatic cone array. PLoS Comput Biol 6: e1000677.
12. DARPA image understanding workshop. Available: http://openlibrary.org/b/0LJ7Z66910HM/. Accessed 2011 June 8.
13. Forsyth DA, Ponce J (2002) Computer Vision: A Modern Approach Prentice Hall (NJ, USA).
14. Torralba A, Oliva A (2003) Statistics of natural image categories. Network: Computation in Neural Systems 14: 391–412.
15. Pennebaker WB, Mitchell JL (1993) The JPEG Still Image Data Compression Standard Van Nostrand Reinhold (Norwell, MA, USA).
16. Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. International J of Comp Vis 40: 49–71.
17. Richards WA (1982) Lightness scale from image intensity distributions. Applied Optics 21: 2569–2582.
18. Barrows GJ, Moorhead IR (1987) Color and spatial structure in natural scenes. Applied Optics 26: 157–170.
19. Olhausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381: 607–609.
20. Bell AJ, Sejnowski TJ (1997) The independent components of natural scenes are edge filters. Vision Research 37: 3327–3338.
21. Vorhees H, Poggio T (1998) Computing texture boundaries from images. Vision Research 38: 1159–1171.
22. Riesenhuber M, Poggio T (2000) Models of object recognition. Nature Neuroscience 3: 1199–1204.
23. Geisler WS, Perry JS, Super RJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. Vision Research 41: 711–724.
24. Tkacik G, Prentice JS, Victor JD, Balasubramanian V (2010) Local statistics in natural scenes predict the saliency of synthetic textures. Proc Natl Acad Sci U S A 107: 18149–54.
25. Kersten D (1987) Predictability and redundancy of natural images. J Opt Soc Am A 4: 2395–400.
26. Chandler D, Field D (2007) Estimates of the information content and dimensionality of natural scenes from proximity distributions. J Opt Soc Am A Opt Image Sci Vis 24: 922–41.
27. Stephens GJ, Mora T, Tkacik G, Bialek W (2008) Thermodynamics of natural images. arXiv: q-bio/0806.2694.
28. Stockman A, Sharpe LT (2000) Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. Vision Research 40: 1711–1737.
29. CIE 2007 fundamental chromaticity diagram with physiological axes, Parts 1 and 2. Technical report 170-1. Vienna: Central Bureau of the Commission Internationale de l’Eclairage.
30. Yin L, Smith RG, Sterling P, Brainard DH (2006) Chromatic properties of horizontal and ganglion cell responses follow a dual gradient in cone opsin expression. J Neurosci 26: 12531–61.
31. Lee AB, Mumford D, Huang J (2001) Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. Int J of Comp Vis 41: 35–59.
32. Public domain internet image gallery software. Available: http://gallery.menalco.com. Last accessed 2011-05-08. Accessed 2011 June 8.
33. Public domain implementation of wget. Available: http://www.gnu.org/software/wget/. Accessed 2011 June 8.
34. Creative commons license. Available: http://creativecommons.org/licenses/by-nc/3.0/. Accessed 2011 June 8.
35. Packer O, Williams DR (2003) Light, the retinal image, and photoreceptors. In The Science of Color, 2nd ed. SK. Shevall, ed. Optical Society of America, Elsevier (Oxford, UK); pp 41–102.
36. Nikon documentation. Available: http://www.nikonusa.com/pdf/manuals/dslr/D700_en.pdf. Accessed 2011 June 8.
37. Is the Nikon D700 NEF (RAW) format truly lossless? Available: http://www.manfujinfo/mynsf/weblog/2009/05/02-1.html. Accessed 2011 June 8.
38. Public domain image extraction software DCRAW. Available: http://www.cybercom.net/~docfin/dcrw/. Accessed 2011 June 8.
39. Kodak Carousel 4400 projector. Available: http://slideprojector.kodak.com/carousel/4400.dhtml. Accessed 2011 June 8.
40. Photo Research PR-650 SpectraScan Colorimeter. Available: http://www.photoresearch.com/current/pr650.asp. Accessed 2011 June 8.
41. Horn BKP (1984) Exact reproduction of colored images. Computer Vision, Graphics and Image Processing 26: 135–167.
42. Wandell BA (1987) The synthesis and analysis of color images. IEEE Trans Pattern Analysis and Machine Intelligence 9: 2–13.
43. Finlayson GD, Drew MS (1996) The maximum ignorance assumption with positivity. Proceedings of 12th IST/SID Color Imaging Conference. pp 202–204.
44. Zhang X, Brainard DH (2004) Bayesian color correction method for non-colorimetric digital image sensors. Proceedings of 12th IST/SID Color Imaging Conference (Scotsdale, AZ); pp 308–314.
45. Sharpe LT, Stockman A, Jylä W, Jägle H (2005) A luminous efficiency function for daylight adaptation. Journal of Vision 5: 948–968.
46. Brainard DH (1997) The psychophysics toolbox. Spatial Vision 10: 433–436.
47. Barlow HB (1982) What causes trichromacy? A theoretical analysis using comb-filtered spectra. Vision Research 22: 635–643.
48. Atick JJ, Redlich AN (1992) Understanding retinal color coding from first principles. Neural Computation 4: 449–572.
49. Maloney LT (1986) Evaluation of linear models of surface spectral reflectance with small numbers of parameters. Journal of the Optical Society of America A 3: 1673–1683.
50. Regan BC, Julliot C, Simmen B, Vernet F, Charles-Dominique P, et al. (2001) Fruits, foliage and the evolution of primate colour vision. Philosophical Transaction of the Royal Society of London, B 356: 229–233.
51. Lythgoe JN, Partridge JC (1989) Visual pigments and the acquisition of visual information. Journal of Experimental Biology 146: 1–20.
52. Chittka L, Menzel R (1992) The evolutionary adaptation of ower colours and the insect pollinators’ colour vision. Journal of Comparative Physiology A 171: 171–181.
53. Osorio D, Vorobyev M (1996) Colour vision as an adaptation to frugivory in primates. Proceedings of the Royal Society of London, B 263: 583–599.
54. Levi A, Zhaoping L (2006) Are cone sensitivities determined by natural color statistics? Journal of Vision 6: 283–302.
55. Ruederman D, Bialek W (1994) Statistics of natural images: Scaling in the woods. Phys Rev Lett 73: 814–817.
56. Ruederman D (1994) The statistics of natural images. Network 5: 517–548.
57. Ruederman DL (1997) Origins of scaling in natural images. Vision Research 37: 3305–3308.
58. Manue V, Frazor RA, Bonin V, Geisler WS, Caramazza M (2005) Independence of luminance and contrast in natural scenes and in the early visual system. Nat Neurosci 8: 1690–7.
59. Sigman M, Cecchi G, Gilbert C, Magnasco MO (2001) On a common circle: natural scenes and gestalt rules. Proc Natl Acad Sci U S A 98: 1935–40.
60. Karklin Y, Lewicki MS (2003) Learning higher-order structures in natural images. Network 14: 483–99.
61. Rodieck RW (1998) The First Steps of Seeing Sinauer Associates (Sunderland, MA, USA).
62. Stockman A, Sharpe LT, Fach CC (1999) The spectral sensitivity of the human short wavelength cones. Vision Research 40: 2901–2927.
63. Bone RA, Landrum JT, Cains A (1992) Optical density spectra of the macular pigment in vivo and in vitro. Vision Research 32: 105–110.
64. Pokorny J, Smith VC (1997) The Veriest Lecture. How much light reaches the retina? In CR. Cavonius, ed. Colour Vision Deficiencies XIII. Documenta Ophthalmologica Proceedings, volume 59 Springer. pp 491–511.
65. A Calibrated Natural Image Database