

From Pixels to Physics: Probabilistic Color De-rendering

Ying Xiong
Harvard University

yxiong@seas.harvard.edu

Kate Saenko
UC Berkeley

saenko@eecs.berkeley.edu

Trevor Darrell
UC Berkeley

trevor@eecs.berkeley.edu

Todd Zickler
Harvard University

zickler@seas.harvard.edu

Abstract

Consumer digital cameras use tone-mapping to produce compact, narrow-gamut images that are nonetheless visually pleasing. In doing so, they discard or distort substantial radiometric signal that could otherwise be used for computer vision. Existing methods attempt to undo these effects through deterministic maps that de-render the reported narrow-gamut colors back to their original wide-gamut sensor measurements. Deterministic approaches are unreliable, however, because the reverse narrow-to-wide mapping is one-to-many and has inherent uncertainty. Our solution is to use probabilistic maps, providing uncertainty estimates useful to many applications. We use a non-parametric Bayesian regression technique—local Gaussian process regression—to learn for each pixel’s narrow-gamut color a probability distribution over the scene colors that could have created it. Using a variety of consumer cameras we show that these distributions, once learned from training data, are effective in simple probabilistic adaptations of two popular applications: multi-exposure imaging and photometric stereo. Our results on these applications are better than those of corresponding deterministic approaches, especially for saturated and out-of-gamut colors.

1. Introduction

Most digital images produced by consumer cameras and shared online exist in narrow-gamut, low-dynamic range formats.¹ This is efficient for storage, transmission, and display, but it is unfortunate for computer vision systems that seek to interpret this data radiometrically when learning object appearance models for recognition, reconstructing scene models for virtual tourism, or performing other visual tasks with Internet images. Indeed, most computer vision algorithms are based, either implicitly or explicitly, on the assumption that image measurements are proportional to the spectral radiance of the scene (called *scene color* hereafter), and when a consumer camera renders its digital

¹Typically sRGB color space with JPEG encoding: IEC 10918-1:1994 and IEC 61966-2-1:1999

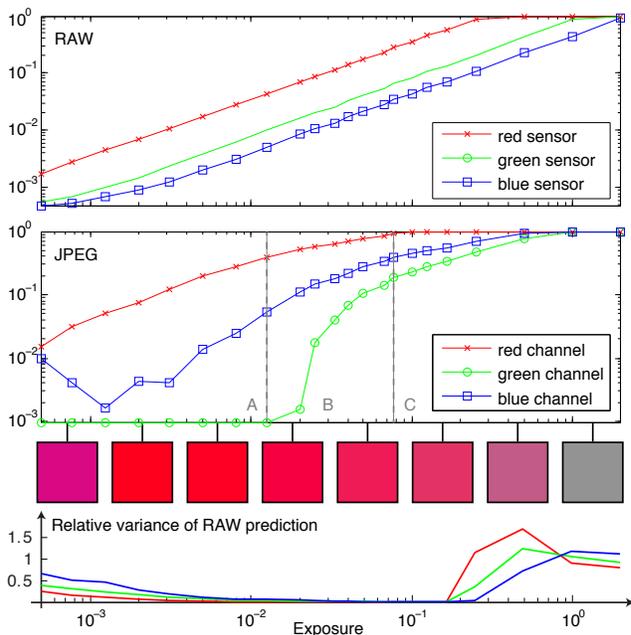


Figure 1. RAW and JPEG values for different exposures of the same spectral scene radiance collected by a consumer digital camera (DMC-LX3, Panasonic Inc.), along with normalized-RGB visualizations of the reported JPEG colors at a subset of exposures. Apart from sensor saturation, RAW values are linear in exposure and proportional to spectral irradiance; but narrow-gamut JPEG values are severely distorted by tone-mapping. Given only JPEG values, what can we say about the unknown RAW values—and thus the scene color—that induced it? How can we use all of the JPEG color information, including when some JPEG channels are saturated (regions A and C)? We answer these questions by providing a confidence level for each RAW estimate (bottom plot), which can be used for radiometry-based computer vision.

linear color measurements to a narrow-gamut output color space (called *rendered color* hereafter), this proportionality is almost always destroyed. Fig. 1 shows an example.

Existing approaches to color de-rendering attempt to undo the effects of a camera’s color processing pipeline through “radiometric calibration” [6, 20, 22], in which rendered colors (*i.e.*, those reported in a camera’s JPEG out-

put) are reverse-mapped to corresponding scene colors (*i.e.*, those that would have been reported by the same camera’s RAW output) using a learned deterministic function. This approach is unreliable, because it ignores the inherent uncertainty caused by the loss of information. A typical camera renders many distinct sensor measurements to the same small neighborhood of narrow-gamut output colors (see Fig. 2, right) and, once these output colors are quantized, the reverse mapping becomes one-to-many in some regions and cannot be deterministically undone.

How can we know which predictions are unreliable? As supported by Fig. 2, one expects the one-to-many effect to be greatest near the edges of the output gamut (*i.e.*, near zero or 255 in an 8-bit JPEG file), and practitioners try to mitigate it using heuristics such as ignoring all JPEG pixels having values above or below certain thresholds in one or more of their channels. This trick improves the reliability of deterministic radiometric calibration, but it raises the question of how to choose thresholds for a given camera. (“Should I only discard pixels with values 0 or 255, or should I be more conservative?”)² A more fundamental concern is that this heuristic works by discarding information that would otherwise be useful. Referring to Fig. 1, such a heuristic would ignore all JPEG measurements in regions A and C, even though these clearly tell us *something* about the latent scene color.

To overcome these limitations, we introduce a probabilistic approach for de-rendering. This method produces from each rendered (JPEG) color a probability distribution over the (wide gamut, high dynamic range) scene colors that could have induced it. The method relies on an offline calibration procedure involving registered RAW and JPEG image pairs, and from these it infers a statistical relationship between rendered colors and scene colors using local Gaussian process regression. This probabilistic approach provides a measure of confidence, based on the variance of the output distribution, for every predicted scene color, thereby eliminating the need for heuristic thresholds and making better use of the scene radiance information that is embedded in an Internet image. The offline calibration procedure is required only once for each different imaging mode of each camera, thus many per-camera de-rendering models could be stored in an online database and accessed on demand using camera model and mode information embedded in the metadata of an Internet image.³

We evaluate our approach in a few different ways. First,

²Our experiments in Fig. 4 and those of [18] reveal significant variation between models and suggest the answer is often the latter.

³As has been done for lens distortion by PTLens (accessed Mar 27, 2012): <http://www.epaperpress.com/ptlens/>

⁴The boundary of the output sRGB gamut is determined automatically from image data in two steps. The edge directions of the parallelepiped are extracted from RAW metadata using ddraw[5], and then its scale is computed as a robust fit to RAW-JPEG correspondences.

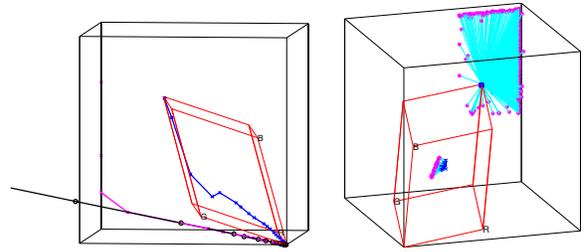


Figure 2. 3D visualization of color rendering. The black cube indicates the set of possible RAW color sensor measurements, and the red parallelepiped shows the boundary of the output sRGB gamut to which all RAW colors must be tone-mapped.⁴ *Left:* Data from Fig. 1, with black circles the scene color x at different exposure times. Corresponding RAW values \tilde{x} (magenta) are clipped due to sensor saturation, and they are tone-mapped to rendered colors y (blue) within the output sRGB gamut. *Right:* Rendered colors (blue) in small neighborhoods of $[127, 127, 127]$ and $[253, 253, 253]$ in a JPEG image, connected (through cyan lines) to their corresponding RAW measurements (magenta).

we assess our ability to recover wide-gamut scene colors from JPEG sRGB observations in four different consumer cameras. Next, we employ our probabilistic de-rendering model in relatively straightforward probabilistic adaptations of two established applications: high-dynamic range imaging with an exposure-stack of images (*e.g.*, [20]) and three-dimensional reconstruction via Lambertian photometric stereo (*e.g.*, [33]). In all cases, a probabilistic approach significantly improves our ability to infer radiometric scene structure from tone-mapped images.

1.1. Related work

There is a history of radiometric calibration for computer vision, the goal of which is to invert non-linear transformations of scene lightness and color that occur during imaging. The most common approach is to assume that the non-linearity can be described by a collection of three “radiometric response functions”, which are monotonic deterministic functions that separately affect the measurements in each output color channel [20, 22, 6, 10]. The benefit of this approach is that it enables “self-calibration” through analysis of edge profiles [19] and image statistics [7, 16] or, assuming white balance is fixed or happens per-channel in the output color space [12], by making use of multiple illuminations or exposures [20, 22, 6, 9, 28, 30]. For the case of multiple exposures, Pal et al. [24] have proposed a generalization that allows the shapes of the radiometric response functions to change between exposures, while being governed by statistical priors that give preference to smooth and monotonic functions.

A significant limitation of the monotonic per-channel model is that it cannot recover out-of-gamut chromaticities. This can be explained using Fig. 2(left), which is a three-

dimensional visualization of Fig. 1. When an out-of-gamut scene color $\mathbf{x} = (x^R, x^G, x^B)$ is rendered to a within-gamut output color $\mathbf{y} = (y^R, y^G, y^B)$, the traditional per-channel approach attempts to undo it by computing the estimate $\hat{\mathbf{x}} = \{f^c(y^c)\}_{c=R,G,B}$ using positive-valued, monotonic functions $f^c(\cdot)$. This estimate cannot always be accurate because it is restricted to lie within the cone defined by the radial extension of the output sRGB gamut.

Chakrabarti et al. [3] show that more accurate deterministic models can be fit using an offline calibration procedure involving registered RAW and JPEG sRGB images. They consider multivariate polynomial models for the forward map from scene color \mathbf{x} to output color \mathbf{y} , and while they find reasonable fits for most cameras, the residual errors remain quite high at 4-6 times most camera noise levels. Lin et al. [18] perform a thorough, larger-scale study and obtain significantly improved fits using radial basis functions, which are more flexible. Both approaches avoid the restrictions of per-channel response functions and can theoretically recover out-of-gamut chromaticities; but they remain deterministic, reporting a single color value instead of a distribution and not allowing for uncertainty prediction.

We represent uncertainty by employing a Bayesian non-parametric regression scheme, which allows the data to determine the form of the mapping. Specifically, we adapt the method of Urtasun and Darrell [32], which learns a local Gaussian process for the neighborhood around each test point, in the spirit of locally-weighted regression [4].

2. A probabilistic de-rendering model

We begin with a model for the forward color processing pipeline of a typical consumer digital camera; then we describe our representation for the reverse mapping. Both models ignore de-mosaicking, flare removal, noise removal, and sharpening since these have significantly less impact on the output than non-linear tone-mapping. More details on these secondary issues can be found elsewhere [3, 2, 25, 14].

An important assumption underlying our model is that the forward rendering operation is spatially-uniform, meaning that its effect on a RAW color vector is the same regardless of where it occurs on the image plane. This assumption is shared by almost all de-rendering techniques and is reasonable at present; but if spatially-varying tone-mapping operators become more common, relaxing this assumption may become a useful direction for future work.

2.1. Forward (rendering) model

Referring to Fig. 3, the forward model begins with three idealized spectral sensors with sensitivity profiles $\{\pi^c(\lambda)\}_{c=R,G,B}$ that sample the spectral irradiance incident on the sensor plane. These sensors are idealized in that they do not saturate and have infinite dynamic range, and we refer to their output $\mathbf{x} = \{x^c\}_{c=R,G,B}$ as the *scene*

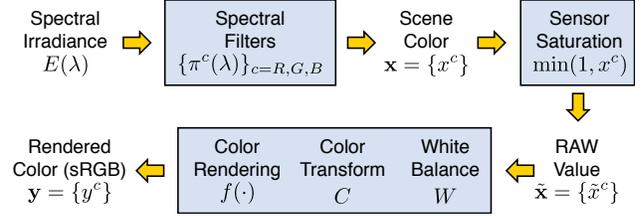


Figure 3. The forward color processing model used in this paper, along with our notation for it. Lesser effects, such as flare removal, de-mosaicking, and vignetting are ignored and treated as noise.

color. Practical sensors have limited dynamic range, so scene colors are clipped as they are recorded. In some consumer cameras these recorded sensor measurements $\tilde{\mathbf{x}} = \{\tilde{x}^c\}_{c=R,G,B}$ are made available through a RAW output format, and in others they only exist internally. Empirical studies suggest that the RAW values (in the absence of clipping) are proportional to incident irradiance and related by a linear transform to measurements that would be obtained by the CIE standard observer [3, 2, 15] (also see Fig. 1). For this reason, they provide a “relative scene-referred image” [12] and can be used directly by computer vision systems to reason about spectral irradiance.

Two linear transforms are applied to the sensor measurements. The first (W) is scene-dependent and induces white balance, and the second (C) is a fixed transformation to an internal working color space. Then, most importantly, the linearly transformed RAW values $CW\tilde{\mathbf{x}}$ are rendered to colors $\mathbf{y} = \{y^c\}_{c=R,G,B}$ in the narrow-gamut output sRGB color space through a non-linear map $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$. This map has evolved to produce visually-pleasing results at the expense of physical accuracy, and since the quality of a camera’s color rendering process plays a significant role in determining its commercial value, there is a disincentive for manufacturers to share its details. In our model, the map f includes the per-channel non-linearity (approximately a gamma of 2.2) that is part of the sRGB standard (IEC 61966-2-1:1999).

The left of Fig. 2 shows signal values at various stages of this forward model for a consumer camera (DMC-LX3, Panasonic Inc.). Recall that the black box in this plot represents the range of possible RAW values $\tilde{\mathbf{x}}$, and the red parallelepiped marks the boundary of the output sRGB gamut. The plot shows color signals produced using different exposure times for a simple static scene consisting of a uniform planar patch under constant illumination, with spatial-averaging over all patch pixels to thoroughly suppress the effects of noise, demosaicking, and JPEG compression. The scene colors \mathbf{x} (black) lie a line that extends well beyond the cube as the exposure time grows large, and the chromaticity of the patch is such that all scene colors lie outside the sRGB gamut. The wide-gamut RAW values $\tilde{\mathbf{x}}$ (magenta) are very close to these scene colors for low exposures, but

they are clipped for longer exposures when the intensity grows large. The rendered colors $\mathbf{y} = f(CW\tilde{\mathbf{x}})$ (blue) lie within the output gamut, and are significantly affected by the combined effects of sensor saturation, white balance, and the color space transform. Interestingly, these rendered colors are relatively far inside the boundary of the sRGB gamut, so the conventional wisdom in radiometric calibration that one should discard pixels with very small or very large JPEG values as being “clipped” is unlikely to detect and properly treat them.

2.2. Inverse (de-rendering) model

Our goal is to infer, for each possible rendered color \mathbf{y} , the original scene color \mathbf{x} that created it. As information is lost in the forward rendering process, exact recovery is not possible and thus any deterministic function that predicts a single point estimate is bound to be wrong much of the time. For that reason, we propose to estimate a *distribution* over the space of possible scene colors. Specifically, we seek a representation of $p(\mathbf{x}|\mathbf{y})$ from which we can either obtain a MAP estimate of \mathbf{x} or directly employ Bayesian inference as desired for a given application (see Sec. 3.1 and Sec. 3.2).

We model the underlying de-rendering function, denoted z , using Gaussian process (GP) regression [27]. Given a training set $\{\mathcal{D} = (\mathbf{y}_i, \mathbf{x}_i), i = 1, \dots, N\}$, composed of inputs \mathbf{y}_i and noisy outputs \mathbf{x}_i , we model the outputs $\{x_i^c\}_{c=R,G,B}$ in each channel separately as coming from a latent function z^c that has a prior distribution described by a GP, and is corrupted by additive noise ϵ_i :

$$x_i^c = z^c(\mathbf{y}_i) + \epsilon_i, \quad \epsilon_i \propto \mathcal{N}(0, \sigma_n^2). \quad (1)$$

The latent function z serves as the inverse of the forward rendering map composed of the color rendering function, color transform, and white balance operations depicted in Fig. 3. We will learn it using images in which the white balance has been fixed to remove scene-dependence.

The classic GP regression paradigm uses a single set of hyper-parameters controlling the smoothness of the inferred function. However, our analysis of camera data has revealed that such globally-defined (*i.e.*, stationary) smoothness is inadequate because there is significantly different behavior in different regions of the sRGB gamut (see right of Fig. 2.) Instead, the variance of z should be allowed to vary over local neighborhoods of the sRGB color space.

Several extensions to the classic GP have been proposed to model input-varying noise [26, 32, 21]. Here, we employ a *local* GP regression model, which exploits the observation that, for compact radial covariance functions, only the points close to a test point have significant influence on the results [32]. Given a training dataset and a test point, the method identifies a local neighborhood of the test point, and performs prediction with the model either pre-trained based on some local cluster (“offline local GP”), or learned on the

fly using neighbor points just detected (“online local GP”).⁵ More precisely, given training set \mathcal{D} and a test sRGB color \mathbf{y} , we infer a test distribution of RAW values \mathbf{x} conditioned on \mathbf{y} by identifying a local neighborhood of \mathbf{y} in \mathcal{D} , denoted $\mathcal{D}_{N(\mathbf{y})}$, and computing

$$p_x(\mathbf{x}|\mathbf{y}) = \prod_c p_{GP}(x^c|\mathcal{D}_{N(\mathbf{y})}, \mathbf{y}), \quad (2)$$

where $p_{GP}(x|D, \mathbf{y})$ is the conditional GP likelihood of x using training data D for sRGB colors \mathbf{y} .

3. Working with photometric uncertainty

Linear measurements of scene radiance are crucial for many computer vision tasks (shape from shading, image-based rendering, deblurring, color constancy, intrinsic images, *etc.*), and the output of our de-rendering model can be readily used in probabilistic approaches to these tasks. Here we describe two such tasks and show how modeling photometric uncertainty leads to more robust results.

3.1. Probabilistic wide-gamut imaging

Many applications that use Internet images operate by inferring radiometric scene properties from multiple observations of the same scene point. For example, multiple observations under different illuminations can be exploited for inferring diffuse object color [23] or more general BRDFs [11]. To explore the benefits of modeling photometric uncertainty in such cases, we consider an example scenario motivated by traditional HDR imaging with exposure stacks [20, 6]. Given as input multiple exposures of the same stationary scene, we seek to combine them into one floating-point, HDR, and wide-gamut image.

Assume we are given a sequence of sRGB vectors captured at shutter speeds of $\{\alpha_1, \alpha_2, \dots, \alpha_N\}$ seconds. Represent these by $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. We would like to predict the RAW color, \mathbf{x}_0 say, that would have been obtained with a shutter speed of α_0 seconds. Note that α_0 need not be one of the shutter speeds used to capture the sRGB input.

Given a training set \mathcal{D} , for each sRGB value \mathbf{y} we estimate the conditional distributions $p_{x_i}(\mathbf{x}_i|\mathbf{y}_i)$ for the RAW value \mathbf{x}_i that would have been obtained with shutter speed α_i . Then, to obtain \mathbf{x}_0 , we combine them using

$$\begin{aligned} p_{x_0}(\mathbf{x}_0|\mathbf{y}_1, \dots, \mathbf{y}_N) &= \prod_i p_{x_0}(\mathbf{x}_0|\mathbf{y}_i) \\ &= \prod_i \frac{\alpha_i}{\alpha_0} p_{x_i} \left(\frac{\alpha_i}{\alpha_0} \mathbf{x}_0 | \mathbf{y}_i \right). \end{aligned} \quad (3)$$

⁵To handle multimodality in the mapping, [32] shows how clustering may be performed in both input and output spaces for the training data, and a set of local regressors returned. However we believe that our inverse map does not have multimodal structure, and we found that a single local regressor provided adequate results. Implementation details with regard to online and offline models are described in Sec. 4.

Since each channel $p_{x_i}(\mathbf{x}_i|\mathbf{y}_i)$ is modeled by a Gaussian process, this expression represents the product of Gaussian distributions, so the conditional distribution $p_{x_0}(\mathbf{x}_0|\mathbf{y}_1, \dots, \mathbf{y}_N) = \prod_i p_{x_0}(\mathbf{x}_0|\mathbf{y}_i)$ will be Gaussian as well. Our output for \mathbf{x}_0 , therefore, is the mean and variance of this Gaussian distribution.

This application reveals the power of a probabilistic model: it provides a distribution rather than a point estimate. For applications that combine multiple independent measurements, this provides a natural way to assign more weight to the estimates that have smaller variance.

3.2. Probabilistic Lambertian photometric stereo

When illumination varies, another way that multiple observations of the same scene can be used is to recover lighting information and/or scene geometry. This may be useful when using Internet images for weather recovery [29], geometric camera calibration [17], or 3D reconstruction [1]. To quantitatively assess the utility of uncertainty modeling in these types of applications we consider the toy problem of recovering from JPEG images three-dimensional scene shape using Lambertian photometric stereo.

Lambertian photometric stereo is a technique for estimating the surface normals of a Lambertian object by observing that object under different lighting conditions and a fixed viewpoint [33]. Suppose there are N different directional lighting conditions, with $\mathbf{l}_i \in \mathbb{R}^3$ the direction and strength of the i th source. Consider a single color channel of single pixel in the image plane; denote by I_i the linear intensity recorded in that channel under the i th light direction; and let $\mathbf{n} \in \mathbb{S}^2$ and $\rho \in \mathbb{R}^+$ be the normal direction and the albedo of the surface patch at the back-projection of this pixel. The Lambertian reflectance model provides the relation $\rho\langle \mathbf{l}_i, \mathbf{n} \rangle = I_i$, and the goal of photometric stereo is to infer the material ρ and shape \mathbf{n} given the set $\{\mathbf{l}_i, I_i\}$.

Defining a pseudo-normal $\mathbf{b} \triangleq \rho\mathbf{n}$, the relation between the observed intensity and the scene parameters becomes

$$\mathbf{l}_i^T \mathbf{b} = I_i. \quad (4)$$

Given three or more $\{\mathbf{l}_i, I_i\}$ -pairs, the traditional Lambertian photometric stereo estimates pseudo-normal \mathbf{b} (and thus ρ and \mathbf{n}) in a least-squares sense:

$$\mathbf{b} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{I}, \quad (5)$$

where \mathbf{L} and \mathbf{I} are the matrix and vector formed by stacking the light directions \mathbf{l}_i and measurements I_i , respectively.

The linear relation between intensity \mathbf{I} and scene radiance is crucial in photometric stereo. One can use RAW measurements when they are available, but for Internet-based vision tasks that rely on sRGB images, one must first de-render the colors to achieve this linearity. In our case, the

de-rendering result for each pixel is described as a Gaussian random variable $I_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, and Eq. (4) can be re-written as

$$\mathbf{l}_i^T \mathbf{b} = \mu_i + \sigma_i \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1). \quad (6)$$

From this it follows (e.g., [13]) that the maximum likelihood estimate of the pseudo-normal \mathbf{b} is obtained through weighted least-squares, with weights given by the reciprocal of the variance. That is,

$$\mathbf{b} = (\mathbf{L}^T \mathbf{W} \mathbf{L})^{-1} \mathbf{L}^T \mathbf{W} \boldsymbol{\mu}, \quad (7)$$

with $\mathbf{W} = \text{diag}\{\sigma_i^{-2}\}_{i=1}^N$.

Once again we see that distributions provided by a probabilistic de-rendering system can be employed very naturally to selectively weight measurements for improved accuracy and robustness.

4. Evaluation

For training, we collect for each camera model densely sampled corresponding measurements of scene color and rendered color. We obtain these by capturing a set of registered RAW and JPEG images of a standard color chart (140-patch Digital ColorChecker SG, X-Rite Inc.). To obtain a complete coverage of the RAW space, we use various camera exposure settings (from all-black to all-white) and various illumination spectra (a tungsten lamp sequentially filtered by 16 distinct gels). This provides a very dense set of RAW/JPEG pairs and more observations of saturated colors than is available in existing databases [3, 18]. We average the RAW and JPEG pixel values within each of the 140 color patches in each image to thoroughly suppress the effects of demosaicking, noise, and compression, and all in all, we obtain between 30,000 and 50,000 RAW/JPEG color pairs $\{\tilde{\mathbf{x}}_i, \mathbf{y}_i\}$ for each camera.

Scene colors \mathbf{x} are obtained from RAW values $\tilde{\mathbf{x}}$ using ddraw [5] for demosaicking without white balance or a color space transform, which produces 16-bit uncompressed color images in the color space defined by the camera's spectral filters. RAW values corresponding to saturated sensor measurements are discarded and replaced by estimates of scene color \mathbf{x} extrapolated from RAW measurements by the same camera under the same illuminant but with lower camera exposure settings.

Three of the cameras—two point-and-shoot models (Canon Powershot S90; Panasonic DMC-LX3) and a digital SLR (Canon EOS 40D)—provide simultaneous RAW and JPEG output, allowing training from each of these camera's data on its own. We also evaluate a fourth camera (Fuji FinePix J10) that provides only JPEG output, and for this we use one of the RAW-capable cameras (the Panasonic) as a proxy to collect the registered RAW images.

For GP regression we use the GPML toolkit.⁶ We im-

⁶Available online at <http://www.gaussianprocess.org/gpml/>.

	RMSE		Rel. RMSE	
	All	Out-of-gamut	All	Out-of-gamut
CANON 40D-baseline	.05	.09	.31	.36
CANON 40D-ours	.02	.03	.07	.09
CANON S90-baseline	.08	.14	.32	.49
CANON S90-ours	.03	.04	.13	.14
PANASONIC-baseline	.14	.09	.64	.56
PANASONIC-ours	.04	.03	.13	.16
FUJI-baseline	.24	n/a	1.46	n/a
FUJI-ours	.13	n/a	.39	n/a

Table 1. Accuracy of single-image RAW prediction. Root-mean-squared error (RMSE) and relative RMSE of the mean values output by our GP model compared to those of a polynomial baseline [3]. We separately show errors over all RAW test colors, and those only over RAW colors that are outside of the sRGB gamut.

plemented and tested both online and offline methods. For the offline local GP, we cluster the training data inputs into exemplars using k -means and learn a local regressor per exemplar. At test time, we use the prediction of the model from the test point’s closest exemplar. The performance of both methods are about the same, but the complexity of the latter is significantly lower. In all experiments described in the following, we use offline local GP with $k = 10$ clusters and output the nearest cluster as the $\mathcal{D}_{N(y)}$ neighborhood for a test point.

We also tested linear and squared exponential (SE) kernels and found the latter to provide superior performance, perhaps because of the nonlinear nature of the rendering operation. The parameters of the SE kernel, as well as the parameters of the additive noise covariance on the outputs, were estimated via maximum likelihood for each local GP.

4.1. De-rendering

To begin, we evaluate our ability to hallucinate scene colors from a single narrow-gamut sRGB image. We use as a baseline a deterministic representation based on [3], which suggests a forward rendering model composed of a linear map C followed by per-channel polynomials. Since our aim is to recover the reverse mapping we invert this model numerically. We only consider the best point estimate (Gaussian mean, and MAP) and ignore for now the uncertainty output of our model. In this scenario, the derendering results of the proposed GP are similar to an RBF model like that of Lin et al. [18].⁷ (The benefit over [18] of providing confidence levels will be evaluated subsequently.)

For each camera, we split the data points into training and testing sets at random, training on 5000 pairs $\{\mathbf{x}_i, \mathbf{y}_i\}$ and testing on the rest. This experiment is designed to provide insight into the predictive power of our model, as compared to the baseline. We report in Table 1 both root-mean-squared error (RMSE) and relative RMSE between

⁷While Lin et al. fit a per-channel nonlinear function followed by a linear kernel RBF, our approach models both effects simultaneously using a squared exponential kernel.

the ground truth scene color and each model’s prediction. Because our dataset is dominated by lower-valued RGB colors, relative RMSE provides a more meaningful measure of the error by accounting for the total brightness of the RGB vectors. We separately report the errors corresponding to RAW test points that are outside of the sRGB gamut (*e.g.*, 29% of our RAW colors captured with the Canon 40D are outside the sRGB gamut) because, as suggested by Fig. 2, these are more affected by color rendering.

Based on these results we can say the following: 1) our model achieves significantly lower mean errors than the deterministic baseline on all four cameras; 2) overall the errors are higher for the Fuji camera, perhaps due to differences between its spectral filters and those of our (Panasonic) RAW proxy; and 3) our model performs equally well for scene colors that are inside and outside of the sRGB gamut (note that we cannot identify them for the Fuji).

4.2. Wide gamut imaging

Here we follow a different experimental paradigm. We hold out all 22 images of an exposure sequence taken under a single illumination, and we train on a randomly sampled subset of 5000 points from the rest. We repeat this for all 16 illuminants and average the results. Comparisons are made to a deterministic HDR algorithm similar to [6] but with offline pre-calibration using either the polynomial model of [3] or the RBF model similar to [18].

Results are shown in Fig. 4, where we see that the GP model consistently outperform both HDR baselines, especially for out-of-gamut colors. As discussed earlier, practitioners often seek to improve the performance of deterministic HDR by applying thresholds to discard JPEG measurements that are near the boundaries of the sRGB gamut. We evaluate this approach by systematically reducing the interval of 8-bit JPEG values that are used as input, starting with all of them ($[0, 255]$), then discarding the lowest and highest graylevels (*i.e.*, using only values in $[1, 254]$), and so on. The performance of deterministic approaches improves dramatically as the thresholds are tightened, but the optimal thresholds seem to be different for different cameras. In contrast, the performance obtained with our probabilistic approach is small and uniform over all test intervals, a property that follows from its proper accounting of uncertainty.

The advantage of the GP model becomes more clear when we separately consider the errors for out-of-gamut chromaticities. These scene colors tend to be poorly estimated by deterministic approaches, which are constrained to providing a single point estimate instead of distribution that fits the one-to-many map. By explicitly modeling these distributions, the GP model provides predictions for out-of-gamut chromaticities that are almost as accurate as those within the sRGB gamut.

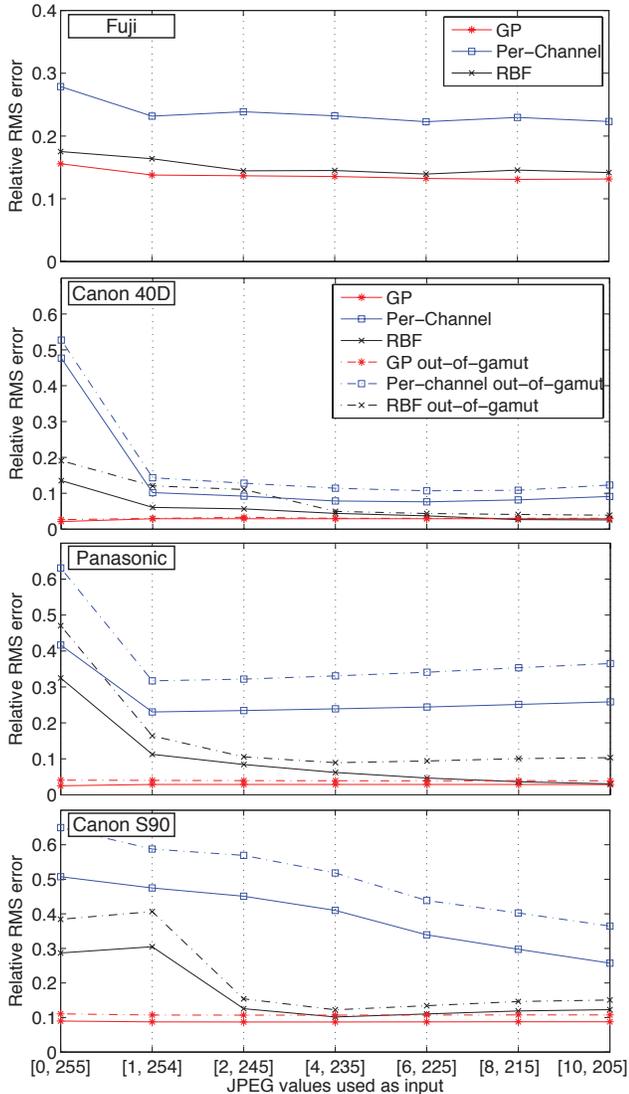


Figure 4. *Wide-gamut imaging*: Estimating wide-gamut linear scene colors from a 22-image exposure sequences of sRGB JPEG images. Plots show relative RMSE in predicted scene colors averaged over 16 runs with input sequences of the same scene under distinct illuminations. Horizontal axis reveals performance change when reduced input JPEG intervals are selected using different thresholds. Label GP (red) refers to our algorithm; Per-channel (blue) refers to use of [3]; and RBF (black) is similar to [18].

4.3. Photometric Stereo

Finally, we evaluate our model in the context of probabilistic Lambertian photometric stereo. We use the Canon 40D to collect JPEG images of a wooden sphere from a fixed (approximately orthographic) viewpoint under directional lighting from twenty different known directions. We apply the algorithm from Sec. 3.2 to estimate a surface normal for each pixel that back-projects to the sphere’s surface. Since the shape of the surface can be determined from its

occluding contour in the orthographic image plane we can compare our results directly to ground truth.

The angular error (degrees) in the estimated surface normal at each pixel is displayed in the left of Fig. 5. The maximum likelihood estimates obtained with the proposed GP model are more accurate than those estimated by the baseline, in which JPEG values are deterministically de-rendered via [3] prior to least-squares estimation of the surface normals. The baseline method yields inaccurate estimates of the surface normals when the JPEG images contain near-saturated values. The third column shows the error that results from a second baseline using gamma-inversed JPEG values (a gamma parameter of 2.2 is assumed), and such errors are much larger, as expected. Quantitatively, the average angular error of the proposed GP model is 3.41° , for baseline model the error is 4.54° , and for gamma-inversed JPEG the error is 8.92° . The improved accuracy of the probabilistic approach is also apparent in the right of Fig. 5, which shows horizontal cross-sections of the depth-maps obtained by integrating the normal fields using [8].

5. Conclusion

Most images captured and shared online are not in linear (RAW) formats, but are instead in narrow-gamut (sRGB) formats with colors that are severely distorted by cameras’ color rendering processes. In order for computer vision systems to maximally exploit the color information in these images, they must first undo the color distortions as much as possible. This paper advocates a probabilistic approach to color de-rendering, one that embraces the multivalued nature of the de-rendered map by providing for each rendered sRGB color a distribution over the latent linear scene colors that could have induced it. An advantage of this approach is that it does not require discarding any image data using ad-hoc thresholds. Instead, it allows making use of all rendered color information by providing for each de-rendered color a measure of its uncertainty.

Our experimental results suggest that a probabilistic representation can be useful when combining per-image estimates of linear scene color, and when recovering the shape of Lambertian surfaces via photometry. The output of our approach—a mean and variance over scene colors for each sRGB image color—may have a practical impact for probabilistic adaptations of other computer vision tasks as well (deblurring, dehazing, matching and stitching, color constancy, image-based modeling, object recognition, *etc.*). One direction worth exploring is the use of spatial structure in the input sRGB image(s), such as edges and textures, to further constrain the de-rendered scene colors. This is in the spirit of [31], and it begs the question of how well a full-gamut linear scene color image can be recovered from a single tone-mapped sRGB one.

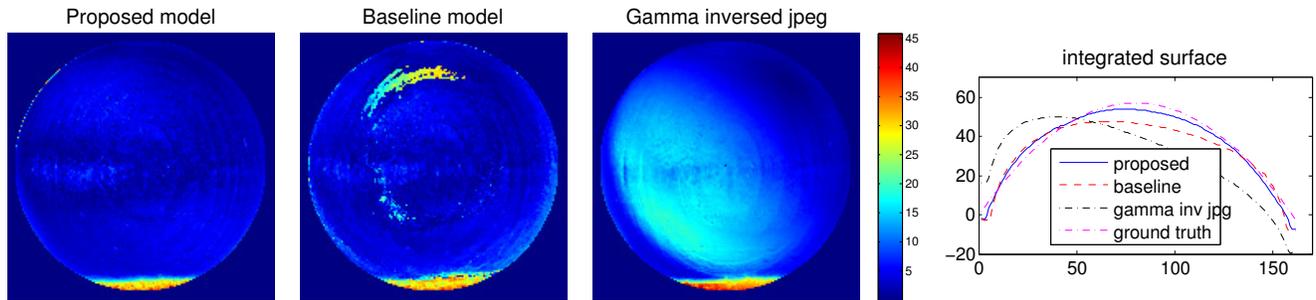


Figure 5. *Photometric stereo*: The left three figures show the angular errors, in degrees, in the per-pixel surface normals obtained using the proposed method, the deterministic baseline, and the gamma-inversed JPEG values. The right figure shows one-dimensional cross sections through surfaces obtained by integrating each set of surface normals, as compared to the ground truth shape.

Acknowledgments

The authors thank Ayan Chakrabarti for many helpful discussions and for providing code from [3]. Authors Ying Xiong and Todd Zickler were supported by NSF awards IIS-0905243, CRI-0708895, and IIS-0546408. Authors Kate Saenko and Trevor Darrell were supported by DARPA contract W911NF-10-2-0059, by NSF awards IIS-0905647 and IIS-0819984, and by Toyota and Google.

References

- [1] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from photometric stereo by example. In *Proc. Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments*, 2010.
- [2] M. Brady and G. Legge. Camera calibration for natural image studies and vision research. *JOSA A*, 26(1):30–42, 2009.
- [3] A. Chakrabarti, D. Scharstein, and T. Zickler. An empirical camera model for internet color vision. In *BMVC*, 2009.
- [4] W. Cleveland, S. Devlin, and E. Grosse. Regression by local fitting. In *Journal of Econometrics*, volume 37, pages 87–114, 1988.
- [5] Decoding raw digital photos in linux. <http://www.cybercom.net/~dcoffin/dcrw/>, Last accessed: January 10, 2011.
- [6] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pages 369–378, 1997.
- [7] H. Farid. Blind inverse gamma correction. *Image Processing, IEEE Transactions on*, 10(10):1428–1433, 2002.
- [8] R. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *PAMI*, 10(4):439–451, 1988.
- [9] M. Grossberg and S. Nayar. Determining the camera response from images: what is knowable? *PAMI*, pages 1455–1467, 2003.
- [10] M. Grossberg and S. Nayar. Modeling the space of camera response functions. *PAMI*, 26(10):1272–1282, 2004.
- [11] T. Haber, C. Fuchs, P. Bekaer, H. Seidel, M. Goesele, and H. Lensch. Relighting objects from image collections. In *CVPR*, 2009.
- [12] D. Hasler and S. Süsstrunk. Mapping colour in image stitching applications. *Journal of Visual Communication and Image Representation*, 15(1):65–90, 2004.
- [13] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.
- [14] J. Holm, I. Tastil, L. Hanlon, and P. Hubel. Color processing for digital photography. In P. Green and L. MacDonald, editors, *Colour Engineering: Achieving Device Independent Colour*, pages 179–220. Wiley, 2002.
- [15] M. Kim and J. Kautz. Characterization for high dynamic range imaging. *Computer Graphics Forum (Proc. EGSR)*, 27(2):691–697, 2008.
- [16] S. Kuthirummal, A. Agarwala, D. Goldman, and S. Nayar. Priors for large photo collections and what they reveal about cameras. In *ECCV*, 2008.
- [17] J. Lalonde, S. Narasimhan, and A. Efros. What do the sun and the sky tell us about the camera? *IJCV*, 88(1):24–51, 2010.
- [18] H. Lin, S. J. Kim, S. Susstrunk, and M. Brown. Revisiting radiometric calibration for color computer vision. In *ICCV*, 2011.
- [19] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum. Radiometric calibration from a single image. In *CVPR*, 2004.
- [20] S. Mann and R. Picard. Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T Annual Conf.*, pages 422–428, 1995.
- [21] A. McHutchon and C. E. Rasmussen. Gaussian process training with input noise. In *NIPS*, 2011.
- [22] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *CVPR*, 1999.
- [23] T. Owens, K. Saenko, A. Chakrabarti, Y. Xiong, T. Zickler, and T. Darrell. Learning object color models from multi-view constraints. In *CVPR*, pages 169–176, 2011.
- [24] C. Pal, R. Szeliski, M. Uyttendaele, and N. Jojic. Probability models for high dynamic range imaging. In *CVPR*, 2004.
- [25] R. Ramanath, W. Snyder, Y. Yoo, and M. Drew. Color image processing pipeline. *IEEE Signal Processing Magazine*, 22(1):34–43, 2005.
- [26] C. E. Rasmussen and Z. Ghahramani. Infinite mixtures of gaussian process experts. In *NIPS*, 2002.
- [27] C. R. Rasmussen and C. K. Williams. *Gaussian Process for Machine Learning*. MIT Press, 2006.
- [28] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High dynamic range imaging*. Elsevier, 2006.
- [29] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *CVPR*, pages 1850–1857, 2009.
- [30] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *CVPR*, 2010.
- [31] K. E. Spaulding, A. C. Gallagher, E. B. Gindele, and R. W. Ptucha. Constructing extended color gamut images from limited color gamut digital images. U.S. Patent No. 7,308,135, 2007.
- [32] R. Urtaşun and T. Darrell. Sparse probabilistic regression for activity-independent human pose inference. In *CVPR*, 2008.
- [33] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980.