Modeling the Uncertainty in Inverse Radiometric Calibration

Ying Xiong, Kate Saenko, Todd Zickler and Trevor Darrell

TR-07-11

Computer Science Group
Harvard University
Cambridge, Massachusetts

Modeling the Uncertainty in Inverse Radiometric Calibration

Ying Xiong[2], Kate Saenko[1], Todd Zickler[2] and Trevor Darrell[1]

[1]UC Berkeley, [2]Harvard University

[1]{saenko,trevor}@eecs.berkeley.edu[2]{yxiong,zickler@seas}.harvard.edu

October 3, 2011

Abstract

While the color image formats used by modern cameras provide visually pleasing images, they distort and discard a significant amount of signal that is useful for many applications. Existing methods for modeling physical world properties based on such narrow-gamut images use a deterministic, per-channel, one-to-one mapping to get back to wide-gamut physical scene colors, ignoring the uncertainty inherent in the process. Rather than fit a deterministic parametric model, we show that non-parametric Bayesian regression techniques, e.g. Gaussian Processes (GP), are well-suited to model this *de-rendering* process, and accurately capture the uncertainty in the transformation. We propose a probabilistic approach that outputs, for each low-gamut image color, a distribution over the wide-gamut scene colors that could have created it. Using a variety of different consumer camera models, we show that effective distributions can be learned by online local Gaussian process regression. Such distributions can be used to hallucinate estimates of RAW values corresponding to JPEG samples, creating "out-of-gamut" images, and also to improve robustness in related applications, e.g., when recovering three-dimensional shape via photometric stereo.

1   Introduction

Most digital images produced by consumer cameras and shared online exist in narrow-gamut, low-dynamic range formats (typically sRGB; IEC 61966-2-1:1999). This is convenient for storage, transmission, and display, but can be unfortunate for computer vision systems that seek to use this data to learn object appearance models for recognition, reconstruct scene models for virtual tourism, or achieve other forms of visual inference. Indeed, most computer vision algorithms are based, either implicitly or explicitly, on the assumption that image measurements are proportional to standardized linear trichromatic projections of spectral scene radiance (called *scene color* hereafter), but when a consumer camera renders its linear color measurements to a narrow-gamut output color space like sRGB (called *rendered color* hereafter), this proportionality is almost always destroyed.
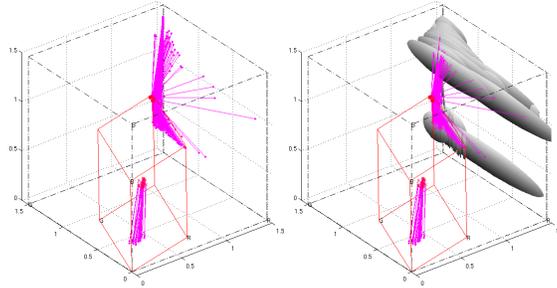
Figure 1: A probabilistic approach for color de-rendering. *Left*: Rendered colors (red dots) in small neighborhoods of $[127, 127, 127]$ and $[253, 253, 253]$ in an JPEG image are connected to the corresponding scene colors. *Right:* Predicted distributions. Near $[127, 127, 127]$ the variances are too small to be visible, and the de-rendering model behaves almost deterministically. Near $[253, 253, 253]$ the model expresses much less certainty because these values are more affected by sensor saturation and tone mapping.

From a computer vision standpoint, the most damaging step of the camera's color processing pipeline is the non-linear "color rendering" or "tone mapping" operation that reduces the original wide-gamut, high-dynamic range measurements to a narrow-gamut output. In order for computer vision systems to make effective use of these output values, they must first de-render them by converting them to estimates (up to proportionality, at least) of the scene colors that produced them.

Traditional approaches to color de-rendering employ deterministic representations of the reverse map from rendered colors to scene colors, but as we will show in Fig. 1 and Fig. 3, these representations are inappropriate for the digital color rendering pipelines that have evolved over the past two decades. In a typical consumer camera (Fig. 1), several out-of-gamut sensor measurements are mapped to the same small neighborhood of rendered in-gamut colors, and once these rendered colors are coarsely quantized (typically 8 bits per channel), this becomes a many-to-one mapping that cannot be deterministically undone.

In this paper, we argue that the assumption that scene colors can be recovered deterministstically is a serious limitation and introduce a probabilistic approach for de-rendering. We present a method that produces from each rendered color a probability distribution over the (wide gamut, high dynamic range) scene colors that could have produced it. The method relies on a set of registered RAW and JPEG images collected by an offline calibration procedure, and it infers from these a statistical relationship between rendered colors and scene colors using local Gaussian process regression.

We evaluate our approach in three different ways. First, we assess our ability to recover wide-gamut scene colors for different consumer cameras. Next, we employ our model for *out-of-gamut imaging*, where we are given a collection of JPEG images of varying exposure and seek to merge them to produce a full-gamut result. Finally, we use our model in the context of Lambertian photometric stereo, where three dimensional shape is inferred from images captured under varying illumination.
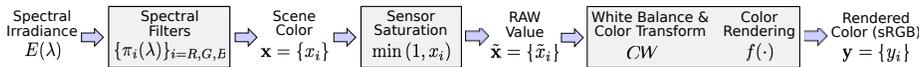
Figure 2: The forward color processing model used in this paper, along with our notation for it. Lesser effects, such as flare removal, de-mosaicking, and vignetting are ignored and treated as noise.

1.1 Related work

There is a long 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 [1, 2, 3, 4]. The benefit of this approach is that it eliminates the need for an offline calibration procedure and enables "self-calibration" through analysis of edge profiles [5] and image statistics [6, 7] or, assuming white balance is fixed or happens per-channel in the output color space [8], by making use of multiple illuminations and exposures [1, 2, 3, 9, 10, 11].

Chakrabarti et al. [12] have shown that a more accurate deterministic model can be fit to registered RAW and JPEG images captured during an offline calibration procedure, which provide corresponding measurements of scene color and rendered color. Their results suggests that a 24-parameter model can provide a reasonable fit for most cameras, but it also shows that the residual errors remain quite high at 4-6 times the camera noise level. We seek to improve this by: 1) avoiding the restriction to deterministic injective mappings; and 2) providing a model for the reverse process (*i.e.* rendered color to scene color) so that it can be used directly for computer vision.

We avoid the restriction to injective mappings by introducing a probabilistic de-rendering model based on non-parametric local regression. Local regression is desirable because the mapping from rendered colors to scene colors can be very complex and difficult to capture in a single mapping. We adopt a Bayesian non-parametric regression scheme to allow the data to determine the form of the mapping while providing an inherent representation of uncertainty. We adopt the method reported in [13], which learns a local Gaussian process for the neighborhood around a test point, in the spirit of locally-weighted regression [14] or KNN-SVM [15]. This method was developed for modeling the appearance-to-pose mapping for human body images; here we apply it to color de-rendering.

2 Probabilistic de-rendering model

We begin with a model for the forward color processing pipeline of a typical consumer digital camera; then 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 can be found elsewhere [12, 16, 17, 18].

2.1 Forward (rendering) model

Referring to Fig. 2, the forward model begins with a collection of three idealized spectral sensors with sensitivity profiles $\{\pi_i(\lambda)\}_{i=R,G,B}$ that sample the spectral irradiance incident on the sensor plane. These sensors are idealized in the sense that they do not saturate and have infinite dynamic range, and we refer to their output $\mathbf{x} = \{x_i\}_{i=R,G,B}$ as the *scene color*. Real sensors have limited dynamic range, so scene colors are clipped as they are recorded. In some consumer cameras these recorded sensor measurements $\tilde{\mathbf{x}} = \{\tilde{x}_i\}_{i=R,G,B}$ 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 [12, 16, 19]. For this reason, they provide a "relative scene-referred image" [8] 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 $\mathbf{y} = \{y_i\}_{i=R,G,B}$ 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 dis-incentive 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).

Fig. 3 show signal values at various stages of this forward model for a consumer camera (Powershot S90, Canon Inc.). In these graphs, the black box represents the range of possible RAW values $\tilde{\mathbf{x}}$, and the red parallelepiped marks the boundary of the output sRGB gamut to which all RAW values must be mapped.[1] Each graph shows the signals acquired for a single scene color over multiple exposures. The scene colors $\mathbf{x}$ (black) lie along lines that extends well beyond the cube. In both examples, the scene colors are outside the sRGB gamut, and while the RAW values $\tilde{\mathbf{x}}$ (magenta) are very close to these scene colors for low exposures, they are clipped when the intensity grows large. Finally, the rendered colors $\mathbf{y} = f(CW\tilde{\mathbf{x}})$ (blue) lie within the output gamut, and especially in the middle case, they are significantly affected by the combined effects of sensor saturation, white balance, and the color space transform.

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 motivated above, a monotonic function is insufficient; furthermore, as information is lost in the forward process, exact recovery is not possible and thus any deterministic function that predicts a single point estimate is bound to be wrong much

---

[1]The boundary of the output gamut is determined automatically in two steps. The edge directions are extracted from RAW metadata using dcraw[20], and then the scale of the parallelepiped is computed as a robust fit to RAW-JPEG correspondences.
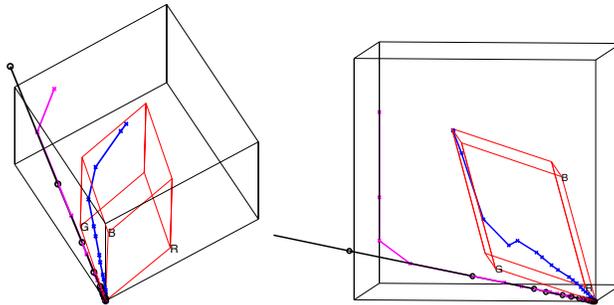
Figure 3: Empirical signal examples of the forward process (Fig. 2) for one consumer camera (Powershot S90, Canon Inc.). Each plot shows a single scene color observed with increasing exposure levels (black circles). The corresponding RAW values $\tilde{\mathbf{x}}$ (magenta) are clipped due to saturation, and they are tone-mapped to create rendered colors $\mathbf{y}$ (blue) within the output sRGB gamut.

of the time. For that reason, we propose to estimate a *distribution* over the space of possible scene colors. Specifially, 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 [21]. Given a training set $\{\mathcal{D} = (\mathbf{y}_i, \mathbf{x}_i), i = 1, \cdots, N\}$, composed of inputs $\mathbf{y}_i$ and noisy outputs $\mathbf{x}_i$, we model the outputs x^c in each channel $c = 1, 2, 3$ separately as coming from a latent function z that has a prior distribution described by a GP, and corrupted by additive noise ϵ_i: $x^c = z(\mathbf{y}_i) + \epsilon_i, \epsilon_i \propto \mathcal{N}(0, \sigma_n^2)$. We can consider z to be the inverse of the mapping containing the color rendering function, color transform, and white balance operations depicted in Fig. 2; practically, we learn it on images for which the white balance has been fixed to remove scene-dependence. Our model assumes a spatially invariant de-rendering function.

The classic GP regression paradigm uses a single set of parameters defining the smoothness of the inferred function. However, our analysis of the camera data has revealed that such globally stationary smoothness is inadequate for our problem, as shown in Fig. 1. The variance of z must vary over local neighborhoods in the input space to model this phenomenon. We therefore exploit a local GP regression model, which exploits the observation that, for compact radial covariance functions, only the points close to a test point will have significant influence on the results [13]. Given a training dataset and a test point, the method identifies the set of nearest neighbors to the test point, and learns a Gaussian Process on the fly using those nearest neighbors as training data, varying the covariance parameters locally.[2] Given a JPEG pixel observation $\mathbf{y}$, we infer a test distribution of RAW values conditioned on $\mathbf{y}$ as follows: we find the nearest neighbors to $\mathbf{y}$ in $\mathcal{D}$, $\mathcal{D}_{N(\mathbf{y})}$, and then obtain an estimate of the cor-

---

[2]To handle multimodality in the mapping, [13] 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 the inverse map does not have multimodal structure, and we found that a single local regressor provided adequate results as described in Section 4.

responding raw value using $p_x(\mathbf{x}|\mathbf{y}) = \prod_c p_{GP}(x^c|\mathcal{D}_{N(\mathbf{y})}, \mathbf{y})$, where $p_{GP}(x|D, \mathbf{y})$ is the conditional GP likelihood of x using training data D for $\mathbf{y}$.

3 Inferring scene properties under photometric uncertainty

Linear measurement of the scene irradiance is a crucial requirement for many computer vision algorithms (e.g. shape from shading, photometric stereo, image-based rendereing, etc.), and the output of our de-rendering model can be readily used in such tasks. In this section, we describe two new methods enabled by the proposed probabilistic derendering model, showing how photometric uncertainty modeling is critical to obtain robust results.

3.1 Probabilistic out-of-gamut imaging

In a wide gamut imaging application, we are given a sequence of JPEG sRGB vectors captured at shutter speeds of $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ seconds. Represent these by $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. We would like to predict the RAW image that would have been obtained with a shutter speed of α_0 seconds. Let's call this $\mathbf{x}_0$. Note that α_0 need not be one of the shutter speeds used to capture the JPEG input.

Given a training set $\mathcal{D}$ as described below, 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.

To obtain a prediction for $\mathbf{x}_0$ we combine these as follows:

$$p_{x_0}(\mathbf{x}_0|\mathbf{y}_1, \ldots, \mathbf{y}_N) = \prod_i p_{x_0}(\mathbf{x}_0|\mathbf{y}_i) = \prod_i p_{x_i}\left(\frac{\alpha_i}{\alpha_0}\mathbf{x}_0|\mathbf{y}_i\right) \tag{1}$$

Since each channel $p_{x_i}(\mathbf{x}_i|\mathbf{y}_i)$ is modeled by a Gaussian process, then $p_{x_i}\left(\frac{\alpha_i}{\alpha_0}\mathbf{x}_0|\mathbf{y}_i\right)$ will have a Gaussian distribution, and so their product, the conditional distribution $p_{x_0}(\mathbf{x}_0|\mathbf{y}_1, \ldots, \mathbf{y}_N) = \prod_i p_{x_0}(\mathbf{x}_0|\mathbf{y}_i)$ will be Gaussian as well. Therefore, our output for $\mathbf{x}_0$ also provides both a mean and a variance.

From this application, we can see the power of our probablistic model: it gives an distribution estimate rather than a point estimate, which can be made use of when combining different estimate results, put more weights on the accurate estimates with small variance and less weights on the ones with large variance.

3.2 Probabilistic 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 [22] and a fixed viewpoint. Suppose there are N different directional lighting conditions, and $\mathbf{l}_i$ is the direction of ith directional light source. Consider a single color channel of single pixel in the image plane, denote by I_i the *linear* intensity recorded under the ith

light direction, and let $\mathbf{n}$ and ρ be the normal direction and the albedo of the surface patch at the back-projection of this pixel. Under the Lambertian model, we can write $\rho \langle \mathbf{l}_i, \mathbf{n} \rangle = I_i$, and the goal of photometric stereo is to infer ρ and $\mathbf{n}$ given the set $\{ \mathbf{l}_i, I_i \}$.

Defining $\mathbf{b} = \rho \mathbf{n}$, (since $\mathbf{n}$ is a unit vector, $\mathbf{b}$ uniquely determines ρ and $\mathbf{n}$, and vice versa), the relation between intensity and light direction can be written as

$$\mathbf{l}_i^T \mathbf{b} = I_i \tag{2}$$

Given three or more $\{ \mathbf{l}_i, I_i \}$-pairs, the traditional approach to Lambertian photometric stereo estimates $\mathbf{b}$ in a least square sense:

$$\mathbf{b} = (\boldsymbol{L}^T \boldsymbol{L})^{-1} \boldsymbol{L}^T \mathbf{I}, \tag{3}$$

where $\boldsymbol{L}$ 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 $\mathbf{I}$ and scene irradiance is crucial in photometric stereo, and therefore a RAW measurement is required. However, if we only recorded the JPEG images when doing this experiment, we can still recover the linear measurement using our GP model. In this case, the linear measurement of each pixel is described as a Gaussian random variable $I_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, and Equation (2) can be written as

$$\mathbf{l}_i^T \mathbf{b} = \mu_i + \sigma_i \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, 1) \tag{4}$$

As noted in [23], when each measurement has different uncertainty, the maximum likelihood estimator for Equation (4) is a weighted least square, using the reciprocal of variance as weight. In this case, the solution is given by

$$\mathbf{b} = (\boldsymbol{L}^T \boldsymbol{W} \boldsymbol{L})^{-1} \boldsymbol{L}^T \boldsymbol{W} \boldsymbol{\mu}, \qquad \text{where} \quad \boldsymbol{W} = \mathrm{diag} \{ \sigma_i^{-2} \}_{i=1}^N \tag{5}$$

This application shows again how we can incorporate the uncertainty measurement produced by the model and get more robust results. The performance of our algorithm is shown in Section 4.3.

4 Evaluation

For training, we require for each camera model many 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.) with various camera exposure settings (from all-black to all-white) and various illumination spectra (Lowel Pro tungsten lamp sequentially filtered by 16 distinct gels). This provides a much more dense set of RAW/JPEG matches than is available in any existing database, such as the Middlebury database [12], as required for our method. We average the RAW and JPEG pixel values within each of the 140 color patches in each image to 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 dcraw [20] 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 that we evaluate—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.[3] We 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 low-gamut images. Since we are not aware of any existing methods that attempt to regress from JPEG to RAW, we use as a baseline the deterministic representation proposed by Chakrabarti et al. [12]. In that paper, the authors analyzed a variety of consumer cameras, and suggested that a deterministic model consisting of a linear map C followed by a per-channel polynomial is adequate for the forward rendering process of most cameras. Here we aim to recover the reverse mapping, so for our deterministic baseline, we invert their model numerically.

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 means that in this experiment, for each particular patch number and illumination, we may be training on some of the exposures, 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 both root mean squared error (RMSE) and relative RMSE between the ground truth scene color and each model's prediction. Because our dataset is dominated by lower-valued RGB colors, relative RMSE gives a better picture of the error as it accounts for the total brightness of the RGB vectors. Finally, we report separately the errors corresponding to datapoints that are outside of the sRGB gamut (29% of RAW colors captured by CANON are outside the sRGB gamut) because, as suggested by Fig. 3, these are more affected by color rendering.

The results are shown in Table 1, from which we can say the following: 1) our model achieves significantly lower mean errors than the deterministic baseline on all three cameras; 2) overall the errors are higher for the Fuji camera, which is not surprising since the original RAW values were not available and the Panasonic RAW was

---

[3] Available online at http://www.gaussianprocess.org/gpml/

8

Table 1: Derendering results

	rmse all	rmse out-of-gamut	rrmse all	rrmse 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

used; 3) for Canon and Panasonic, our model performs equally well for scene colors that are inside and outside of the sRGB gamut (we cannot identify them for the Fuji).

4.2 Wide gamut imaging

Here we show results for *Wide Gamut imaging*, one of many possible applications of our model. The application is similar to High Dynamic Range (HDR) imaging, in that it combines a sequence of images taken at different exposures, and reconstructs an estimate of scene irradiance. The difference is that while traditional HDR is limited to the sRGB gamut, with our model we hope to be better able to reconstruct those scene colors that are outside the gamut.

Here we follow a different experimental paradigm: we hold out all 22 multiple-exposure images taken under a single illumination as our test sequence, and train on a randomly sampled subset of 5K points from the rest. We repeat this for all 16 illuminants and average the results. Comparisons are made with a traditional HDR algorithm implemented in [3]. Results are shown in Figure 4.

The results show that our GP model consistently outperform the HDR baseline, especially in the out-of-gamut region. Since our probablistic model takes account of confidence level (variance) of the estimate, the prediction error is small and constant in all test intervals; while the traditional deterministic algorithm would be affected by saturation and out-of-gamut colors, and therefore its performance gets better when the testing interval shrinks. The advantage of the GP model is more clear when only considering the out-of-gamut colors. As the traditional HDR algorithm limits its operation inside the sRGB gamut, it is unable to accurately infer those colors that are outside and therefore the prediction performance is poor. But the GP model has the advantage of calibrating the camera on the whole wide-gamut space, which gives its a much better performance. From the results, we see that our prediction on out-of-gamut colors are almost as good as those in-gamut.

4.3 Photometric Stereo

Finally, we evaluate our model for probabilistic Lambertian photometric stereo. For this we use the Canon EOS 40D to collect JPEG images of a wooden sphere from
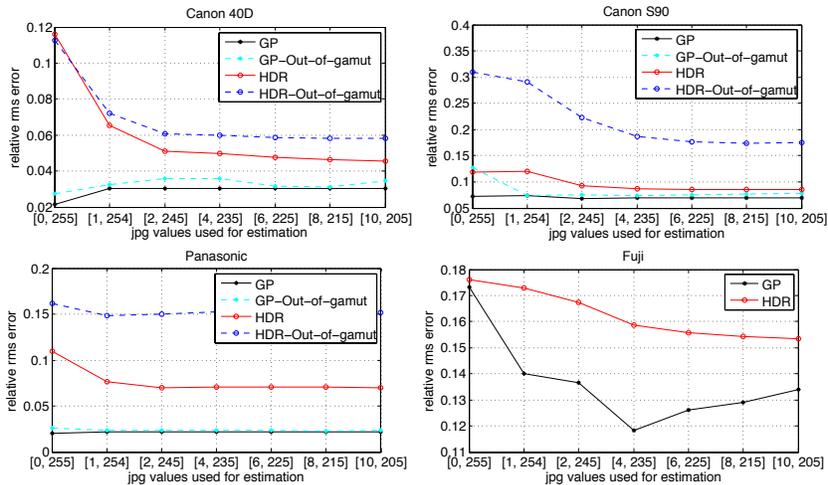
Figure 4: *Wide-gamut imaging:* Results of estimating wide-gamut linear scene colors from an exposure-sequence of sRGB JPEG images captured with 22 different exposures. Plots show relative RMSE in the predicted scene colors averaged over sixteen runs with exposure sequences of the same scene under distinct illuminations.

a fixed (approximately orthographic) viewpoint under directional illumination from twenty different known lighting directions. We apply the algorithms from Sec 3.2 to estimate the surface normal for each pixel that back-projects to the surface of the ball. Since the shape of the surface is known (i.e., it is defined by its occluding contour in the orthographic image plane) we can compare our results directly to ground truth.

The angular error (radians) 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 [12] prior to least-squares estimation of the surface normals. The baseline method yields very poor estimates of the surface normals when the JPEG images contain very large values. The third column shows the error that results from using the JPEG values directly without any de-rendering, and these are much larger, as expected. Quantitively, the average angular error of the proposed GP model is 3.41, for baseline model the error is 4.54, and for JPEG the error is 11.43 (all the errors are mesured in unit of degree).

As an additional comparison, we integrate each of the three normal vector fields to obtain a heigh field using [24], and show a one-dimensional cross section of each height field, corresponding to the horizontal scanline through the middle of each sphere. These are drawn in the right of Fig. 5 along with the ground truth shape, and we see that the proposed approach provides a more accurate result.

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 cam-
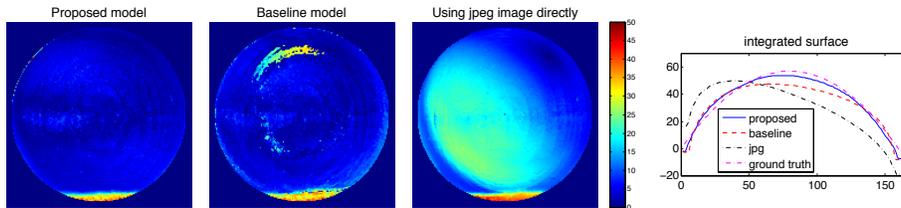
Figure 5: *Photometric stereo:* The left three figures show the angular errors in the per-pixel surface normals obtained using the proposed method, the deterministic baseline, and the JPEG values directly without derendering (unit: degree). The right figure shows one-dimensional cross sections through surfaces obtained by integrating each set of surface normals and compares them to the ground truth shape.

eras' 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-rendering 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 degree to which the output of our approach—a mean and variance over scene colors for each sRGB image color—can have a practical impact for various other computer vision tasks (image-based modeling, object recognition, etc.) remains to be determined in future research. One direction that is likely worth exploring in the short term 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 Ref. [25], and it leads one to wonder about the accuracy with which a full-gamut scene color image can be recovered from a single sRGB one.

References

[1] 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.

[2] T. Mitsunaga and S. Nayar. Radiometric self calibration. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1999.

[3] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proc. Conf. Computer Graphics*, pages 369–378, 1997.

[4] M. Grossberg and S. Nayar. Modeling the space of camera response functions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, 2004.

[5] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum. Radiometric calibration from a single image. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004.

[6] H. Farid. Blind inverse gamma correction. *Image Processing, IEEE Transactions on*, 10(10):1428–1433, 2002.

[7] S. Kuthirummal, A. Agarwala, D. Goldman, and S. Nayar. Priors for large photo collections and what they reveal about cameras. In *Proc. European Conf. Computer Vision*, 2008.

[8] D. Hasler and S. Süsstrunk. Mapping colour in image stitching applications. *Journal of Visual Communication and Image Representation*, 15(1):65–90, 2004.

[9] M.D. Grossberg and S.K. Nayar. Determining the camera response from images: what is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1455–1467, 2003.

[10] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High dynamic range imaging*. Elsevier, 2006.

[11] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Proc. CVPR*, 2010.

[12] A. Chakrabarti, D. Scharstein, and T. Zickler. An empirical camera model for internet color vision. In *British Machine Vision Conference*, 2009.

[13] R. Urtasun and T. Darrell. Sparse probabilistic regression for activity-independent human pose inference. In *CVPR08*, 2008.

[14] W. Cleveland, S. Devlin, and E. Grosse. Regression by local fitting. In *Journal of Econometrics*, volume 37, pages 87–114, 1988.

[15] H. Zhang, A.C. Berg, M. Maire, and J. Malik. Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *CVPR 2006*, 2006.

[16] M. Brady and G.E. Legge. Camera calibration for natural image studies and vision research. *Journal of the Optical Society of America A*, 26(1):30–42, 2009.

[17] R. Ramanath, W. Snyder, Y. Yoo, and M. Drew. Color image processing pipeline. *IEEE Signal Processing Magazine*, 22(1):34–43, 2005.

[18] J. Holm, I. Tastl, 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.

[19] M.H. Kim and J. Kautz. Characterization for high dynamic range imaging. *Computer Graphics Forum (Proc. EGSR)*, 27(2):691–697, 2008.

[20] Decoding raw digital photos in linux. http://www.cybercom.net/ dcoffin/dcraw/, Last accessed: Janurary 10, 2011.

[21] C. R. Rasmussen and C. K. Williams. *Gaussian Process for Machine Learning*. MIT Press, 2006.

[22] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980.

[23] Jerome Friedman Trevor Hastie, Robert Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[24] R.T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):439–451, 1988.

[25] 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.