Color Constancy, Intrinsic Images, and Shape Estimation

Jonathan T. Barron and Jitendra Malik
{barron, malik}@eecs.berkeley.edu

UC Berkeley

Abstract. We present **SIRFS** (shape, illumination, and reflectance from shading), the first unified model for recovering shape, chromatic illumination, and reflectance from a single image. Our model is an extension of our previous work [1], which addressed the achromatic version of this problem. Dealing with color requires a modified problem formulation, novel priors on reflectance and illumination, and a new optimization scheme for dealing with the resulting inference problem. Our approach outperforms all previously published algorithms for intrinsic image decomposition and shape-from-shading on the MIT intrinsic images dataset [1, 2] and on our own "naturally" illuminated version of that dataset.

1 Introduction

In 1866, Helmholtz noted that "In visual observation we constantly aim to reach a judgment on the object colors and to eliminate differences of illumination" ([3], volume 2, p.287). This problem of color constancy — decomposing an image into illuminant color and surface color — has seen a great deal of work in the modern era, starting with Land and McCann's Retinex algorithm [4, 5]. Retinex ignores shape and attempts to recover illumination and reflectance in isolation, assumptions shared by nearly all subsequent work in color constancy [6–11]. In this paper we present the first algorithm for recovering shape in conjunction with surface color and color illumination given only a single image of an object, which we call "shape, illumination, and reflectance from shading" (SIRFS).

There are many early works regarding color constancy, such as gamut mapping techniques [6], finite dimensional models of reflectance and illumination [7], and physically based techniques for exploiting specularities [8]. More recent work uses contemporary probabilistic tools, such as modeling the correlation between colors in a scene [9], or performing inference over priors on reflectance and illumination [10]. All of this work shares the assumptions of Retinex that shape (and to a lesser extent, shading) can be ignored or abstracted away.

Color constancy can be viewed as a subset of the intrinsic images problem: decomposing a single image into its constituent "images": shape, reflectance, illumination, etc [13]. Over time, the computer vision community has reduced this task to just the decomposition of an image into shading and reflectance. Though

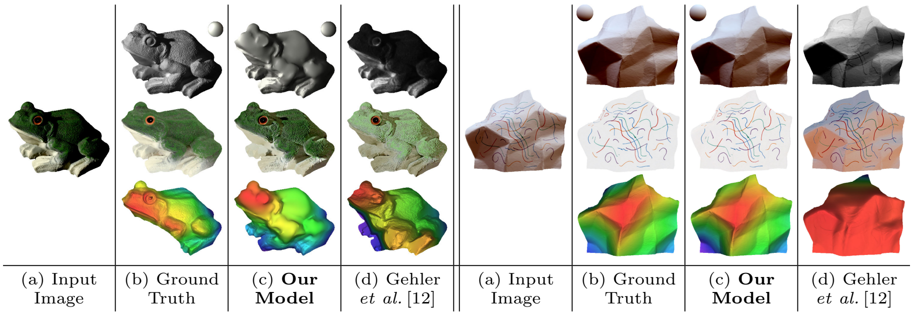


Fig. 1. Two objects from our datasets. Given just the masked input image (a), our model produces (c): a depth-map, reflectance image, shading image, and illumination model that together exactly explain the input image (illumination is rendered on a sphere, and shape is shown as a pseudocolor visualization where red is near and blue is far). Our output looks very similar to (b), the ground-truth explanation of the image — in some cases, nearly indistinguishable. The top-performing intrinsic image algorithm (d) performs much worse on our datasets, and only estimates shading and reflectance (we assume ground-truth illumination is known for (d), and run a shape-from-shading algorithm on shading to produce a shape estimate). Many more similar results can be seen in the supplementary material.

this simplified "intrinsic images" problem has seen a great deal of progress in recent years [2, 12, 14, 15] all of these techniques have critical difficulties with non-white illumination — that is, they do not address color constancy. Additionally, none of these techniques recover shape or illumination, and instead consider shading in isolation.

Another special case of intrinsic images is shape-from-shading (SFS) [16], in which reflectance and illumination are assumed to be known and shape is recovered. This problem has been studied extensively [17, 18], and very recent work has shown that accurate shape can be recovered under natural, chromatic illumination [19], but the assumptions of known illumination and uniform reflectance severely limit SFS's usefulness in practice.

Perceptual studies show that humans use spatial cues when estimating lightness and color [20, 21]. This suggests that the human visual system does not independently solve the problems of color constancy and shape estimation, in contrast to the current state of computer vision.

Clearly, these three problems of color constancy, intrinsic images, and shape from shading would benefit greatly from a unified approach, as each subproblem's strength is another's weakness. We present the first such unified approach, by building heavily on the "shape, albedo, and illumination from shading" (SAIFS) model of our previous work [1], which addresses this problem for grayscale images and white illumination. We extend this technique to color by: trivially modifying the rendering machinery to use color illumination, introducing novel priors for reflectance and illumination, and introducing a novel multiscale inference scheme for solving the resulting problem. We evaluate on the MIT intrinsic images dataset [1, 2], and on our own variant of the MIT dataset in which we have re-rendered the objects under natural, chromatic illuminations produced from real-world environment maps. This additional dataset allows us to evaluate on images produced under natural illumination, rather than the "laboratory"-style controlled illumination of the MIT dataset.

We will show that our unified model outperforms all current techniques for the task of recovering shape, reflectance, and, optionally, illumination. By exploiting color in natural reflectance images, we do better than the grayscale technique of [1] at disambiguating between shading and reflectance. By explicitly modeling shape and illumination we are able to outperform "intrinsic image" algorithms, which only consider shading and reflectance and perform poorly as a result. By modeling chromatic illumination we are able to exploit chromatic shading information, and thereby produce improved shape estimates, as demonstrated in [19]. For these reasons, when faced with images produced under natural, non-white illumination the performance of our algorithm actually *improves*, while intrinsic algorithms perform much worse. See Figure 1 for examples of the output of our algorithm and of the best-performing intrinsic image algorithm.

In Section 2, we present a modification of the problem formulation of [1]. In Sections 3, 4, and 5 we motivate and introduce three novel priors on reflectance images: one based on local smoothness, one based on global sparsity or entropy, and one based on the absolute color of each pixel. In Section 6 we introduce a prior on illumination, and in Section 7 we present a novel multiscale optimization technique that is critical to inference. In Section 8 we show results for the MIT dataset and our own version of the MIT dataset with natural illumination, and in Section 9 we conclude.

2 Problem Formulation

Our problem formulation is an extension of the "SAIFS" problem formulation of [1], which is itself an extension of the "SAFS" formulation of [22]. We optimize over a depth map, reflectance image, and model of illumination such that cost functions on those three quantities are minimized, and such that the input image is exactly recreated by the output shape, albedo, and illumination.

More formally, let R be a log-reflectance map, Z be a depth-map, and L be a model of illumination, and $S(Z, L)$ be a "rendering engine" which produces a log-shading image given depth-map Z and illumination L. Assuming Lambertian reflectance, the log-intensity image I is equal to $R + S(Z, L)$. I is observed, and $S(\cdot)$ is defined, but Z, R, and L are unknown. We search for the most likely (or equivalently, least costly) explanation for image I, which corresponds to solving the following optimization problem:

$$\begin{aligned} \underset{Z,R,L}{\text{minimize}} \quad & g(R) + f(Z) + h(L) \\ \text{subject to} \quad & I = R + S(Z, L) \end{aligned} \tag{1}$$

where $g(R)$ is the cost of reflectance R (roughly, the negative log-likelihood of R), $f(Z)$ is the cost of shape Z, and $h(L)$ is the cost of illumination L. To optimize Equation 1, we eliminate the constraint by rewriting $R = I - S(Z, L)$, and minimize the resulting unconstrained optimization problem using multiscale L-BFGS (see Section 7) to produce depth map $\hat{Z}$ and illumination $\hat{L}$, with which we calculate reflectance image $\hat{R} = I - S(\hat{Z}, \hat{L})$. When illumination is known, L is fixed. This problem formulation differs from that of [1] in that we have a single model of illumination which we optimize over and place priors on, rather than a distribution over "memorized" illuminations. This is crucial, as the huge variety of natural chromatic illuminations makes the previous formulation intractable.

To extend the grayscale model of [1] to color, we must redefine the prior on reflectance $g(R)$ to take advantage of the additional information present in color reflectance images, and to address the additional complications that arise when illumination is allowed to be non-white. Because illumination is a free parameter in our problem formulation, we must define a prior on illumination $h(L)$. We use the same $S(Z, L)$ and a modified version of $f(Z)$ as [1] (see the supplementary material).

Our prior on reflectance will be a linear combination of three terms:

$$g(R) = \lambda_s g_s(R) + \lambda_e g_e(R) + \lambda_a g_a(R) \tag{2}$$

where the λ weights are learned using cross-validation on the training set. $g_s(R)$ and $g_e(R)$ are our priors on local smoothness and global entropy of reflectance, and can be thought of as multivariate generalizations of the grayscale model of [1]. $g_a(R)$ is a new "absolute" prior on each pixel in R that prefers some colors over others, thereby addressing color constancy.

3 Local Reflectance Smoothness

The reflectance images of natural objects tend to be piecewise smooth — or equivalently, variation in reflectance images tends to be small and sparse. This insight is fundamental to most intrinsic image algorithms [2, 4, 5, 14, 23], and is used in our previous works [1, 22]. In terms of color, variation in reflectance tends to manifest itself in both the luminance and chrominance of an image (white transitioning to blue, for example) while shading, assuming the illumination is white, affects only the luminance of an image (light blue transitioning to dark blue, for example). Past work has exploited this insight by building specialized models that condition on the chrominance variation of the input image [2, 5, 12, 14, 15]. Effectively, these algorithms use image chrominance as a substitute for reflectance chrominance, which means that they fail when faced with non-white illumination, as we will demonstrate. We instead simply place a multivariate prior over differences in reflectance, which avoids this non-white illumination problem while capturing the color-dependent nature of reflectance variation.

Our prior on reflectance smoothness is a multivariate Gaussian scale mixture (GSM) placed on the differences between each reflectance pixel and its neighbors. We will maximize the likelihood of R under this model, which corresponds to

(a) Our GSM smoothness prior (b) R - a proposed reflectance image (c) $g_s(R)$ - cost under our model (d) $\nabla g_s(R)$ - influence under our model

Fig. 2. Our smoothness prior is a multivariate Gaussian scale mixture on the differences between nearby reflectance pixels (Figure 2(a)). This distribution prefers nearby reflectance pixels to be similar, but its heavy tails allow for rare non-smooth discontinuities. We see this by analyzing some image R as seen by our model. Strong, colorful edges, such as those caused by reflectance variation, are very costly (have a low likelihood) while small edges, such as those caused by shading, are more likely. But in terms of *influence* — the gradient of cost with respect to each reflectance pixel — we see an inversion: because sharp edges lie in the tails of the GSM, they have little influence, while shading variation has great influence. This means that during inference our model attempts to explain shading in the image by varying shape, while ignoring sharp edges in reflectance. Additionally, because this model captures the correlation between color channels, chromatic variation has less influence than achromatic variation (because it lies further out in the tails), making it more likely to be ignored during inference.

minimizing the following cost function:

$$g_s(R) = \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \, \mathcal{N}\left(R_i - R_j \, ; \mathbf{0}, \boldsymbol{\sigma}_k \, \Sigma\right) \right) \tag{3}$$

Where $N(i)$ is the 5×5 neighborhood around pixel i, $R_i - R_j$ is a 3-vector of the log-RGB differences from pixel i to pixel j, $K = 40$ (the GSM has 40 discrete Gaussians), $\boldsymbol{\alpha}$ are mixing coefficients, $\boldsymbol{\sigma}$ are the scalings of the Gaussians in the mixture, and Σ is the covariance matrix of the entire GSM (shared among all Gaussians of the mixture). The mean is 0, as the most likely reflectance image should be flat. The GSM is learned on the reflectance images in our training set. The differences between this model and that of [1] are: 1) we have a multivariate rather than univariate GSM, to address color, 2) we're placing priors on the differences between all pairs of reflectance pixels within a window, rather than placing a prior on the magnitude of the gradient of reflectance at each pixel, as this produces better results, and 3) we have one single-scale prior, as multiscale priors no longer improve results when using our improved optimization technique. A visualization and explanation of the effect of this smoothness prior can be found in Figure 2.

4 Global Reflectance Entropy

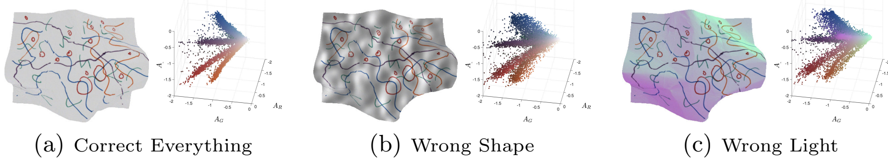
The reflectance image of a single object tends to be "clumped" in RGB space, or equivalently it can be approximated by a set of "sparse" exemplars. This motivates the second term of our model of reflectance: a measure of global entropy which we minimize. We will build upon our previous model [1], but different forms of this idea have been used in intrinsic images techniques [23, 12], photometric stereo [24], shadow removal [25], and color representation [26]. As in [1], we build upon the entropy measure of Principe and Xu [27], which is a model of quadratic entropy (or Rényi entropy) for a set of points assuming a Parzen window. This can be thought of as a "soft" and differentiable generalization of Shannon entropy, computed on a set of points rather than a histogram.

A naive extension of the one-dimensional entropy model of [1] to three dimensions is not sufficient: The RGB channels of natural reflectance images are highly correlated, causing a naive isotropic entropy measure to work poorly. To address this, we pre-compute a whitening transformation from training reflectance images and compute an isotropic entropy measure in this whitened space during inference, effectively giving us an anisotropic entropy measure. Formally, our cost function is non-normalized Rényi entropy in the space of whitened reflectance:

$$g_e(R) = -\log\left(\sum_i \sum_j \exp\left(-\frac{\|\mathrm{W}R_i - \mathrm{W}R_j\|_2^2}{4\sigma_e^2}\right)\right) \quad (4)$$

Where W is the whitening transformation learned from training reflectance images, as follows: Let X be a $3 \times n$ matrix of the pixels in the reflectance images in our training set. We compute the covariance matrix $\Sigma = \mathrm{X}\mathrm{X}^{\mathrm{T}}$ (ignoring centering), take its eigenvalue decomposition $\Sigma = \Phi\Lambda\Phi^{\mathrm{T}}$, and from that construct the whitening transformation $\mathrm{W} = \Phi\Lambda^{1/2}\Phi^{\mathrm{T}}$. σ_e is the bandwidth of the Parzen window, which determines the scale of the clusters produced by minimizing this entropy measure, and is tuned through cross-validation. See Figure 3 for a motivation of this model.

These Rényi measures of entropy are quadratically expensive to compute naively, so others have used the Fast Gauss Transform [25] and histogram-based techniques [1] to approximate it in linear time. The histogram-based technique appears to be more efficient than the FGT-based methods, and provides a way to compute the analytical gradient of entropy, which is crucial for optimization.



(a) Correct Everything $g_e(R) = 0.913$

(b) Wrong Shape $g_e(R) = 1.325$

(c) Wrong Light $g_e(R) = 2.366$

Fig. 3. Reflectance images and their corresponding log-RGB scatterplots. Mistakes in estimating shape or illumination produce shading-like or illumination-like artifacts in the inferred reflectance, causing the the RGB distribution of the inferred reflectance to be "smeared", and causing entropy (and therefore cost) to increase.

We therefore use a 3D generalization of the algorithm of [1] to compute our entropy measure. The resulting technique looks very similar to the bilateral grid [28] used in high-dimensional Gaussian filtering, and can be seen in the supplementary material.

5 Absolute Color

The previously described priors were imposed on *relative* properties of reflectance: the differences between adjacent or non-adjacent pixels. Though this was sufficient for past work, now that we are attempting to recover surface color and non-white illumination we must impose an additional prior on *absolute* reflectance: the raw log-RGB value of each pixel in the reflectance image. Without such a prior (and the prior on illumination presented in Section 6) our model would be equally pleased to explain a white pixel in the image as white reflectance under white illumination as it would blue reflectance under yellow illumination, for example.

This sort of prior is fundamental to color-constancy, as most basic color constancy algorithms can be viewed as minimizing a similar sort of cost: the gray-world assumption penalizes reflectance for being non-gray, the white-world assumption penalizes reflectance for being non-white, and gamut-based models penalize reflectance for lying outside of a gamut of previously-seen reflectances. We experimented with variations or combinations of these types of models, but found that a simple density model on whitened log-RGB values worked best.

Our model is a 3D thin-plate spline (TSP) fitted to the distribution of whitened log-RGB reflectance pixels in our training set. Formally, to train our model we minimize the following:

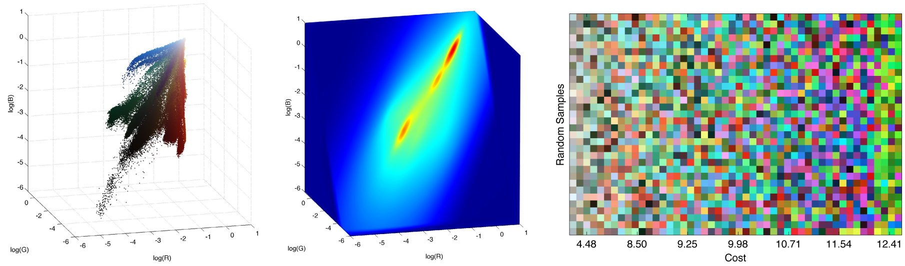
$$\underset{\mathrm{F}}{\text{minimize}} \quad \left(\sum_{i,j,k} \mathrm{F}_{i,j,k} \cdot \mathrm{N}_{i,j,k} \right) + \log \left(\sum_{i,j,k} \exp\left(-\mathrm{F}_{i,j,k}\right) \right) + \lambda \sqrt{J(\mathrm{F}) + \epsilon^2}$$
$$J(\mathrm{F}) = \mathrm{F}_{xx}^2 + \mathrm{F}_{yy}^2 + \mathrm{F}_{zz}^2 + 2\mathrm{F}_{xy}^2 + 2\mathrm{F}_{yz}^2 + 2\mathrm{F}_{xz}^2 \tag{5}$$

Where F is a 3D TSP describing cost (or non-normalized negative log-likelihood), N is a 3D histogram of the whitened log-RGB reflectance in our training data, and $J(\cdot)$ is the TSP bending energy cost (made more robust by taking its square root, with ϵ^2 added to make it differentiable everywhere). Minimizing the sum of the first two terms is equivalent to maximizing the likelihood of the training data, and minimizing the third term causes the TSP to be piece-wise smooth. The smoothness multiplier λ is tuned through cross-validation.

During inference, we maximize the likelihood of the reflectance image R by minimizing its cost under our learned model:

$$g_a(R) = \sum_i \mathrm{F}(\mathrm{W}R_i) \tag{6}$$

where $\mathrm{F}(\mathrm{W}R_i)$ is the value of F at the coordinates specified by the 3-vector $\mathrm{W}R_i$, the whitened reflectance at pixel i (W is the same as in Section 4). To



(a) Training reflectances (b) Our PDF of reflectance (c) Reflectances sorted by cost

Fig. 4. A visualization of our "absolute" prior on reflectance. On the left we have the log-RGB reflectance pixels in our training set, and a visualization of the 3D thin-plate spline PDF that we fit to that data. Our model prefers reflectances that are close to white or gray, and that lie within gamut of previously seen colors. Though our prior is learned in whitened log-RGB space, here it is shown in unwhitened coordinates, hence its anisotropy. On the right we have randomly generated reflectances, sorted by their cost (negative log-likelihood) under our model. Our model prefers less saturated, more subdued colors, and abhors brightly lit neon-like colors. The low-cost reflectances look like a tasteful selection of paint colors, while high-cost reflectances don't even look like paint at all, but instead appear almost glowing and luminescent.

make this function differentiable, we compute $\mathrm{F}(\cdot)$ using trilinear interpolation. A visualization of our model and of the colors it prefers can be seen in Figure 4.

6 Priors over Illumination

In our previous work, inference with unknown illumination involved maximizing an expected complete log-likelihood with respect to a memorized set of $\sim$100 illuminations taken from the training set. That framework was an effective way of both optimizing with respect to illumination (as the posterior distribution over illuminations was re-evaluated at each step in optimization, effectively "moving" the light around) and of regularizing illumination in a non-parametric way (as only previously seen illuminations were considered). However, that framework requires an extremely expensive marginalization over a set of illuminations, which causes inference to be extremely slow — hours per image. That framework also scales linearly with the complexity of the illumination, so modeling the variety of natural, colorful illuminations makes inference impossibly slow. For these reasons, in this paper we adopt a simplified model (Equation 1) in which we explicitly optimize over a single model of illumination in conjunction with shape. This allows us to model and recover a very wide variety of natural illuminations (see Figure 5), while making inference effectively as fast as if illumination were known — around 5 minutes per image. Unfortunately, this model also requires us to explicitly define $h(L)$, our prior on illumination.

We use a spherical-harmonic (SH) model of illumination, so L is a 27 dimensional vector (9 dimensions per RGB channel). In contrast to traditional SH

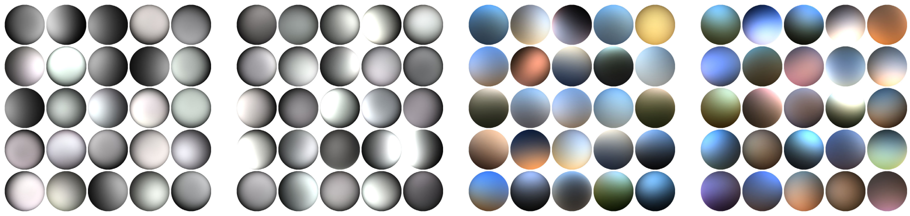


Fig. 5. We use two datasets: the "laboratory"-style illuminations of the MIT intrinsic images dataset [2, 1] which are harsh, mostly-white, and well-approximated by point sources, and a new dataset of "natural" illuminations, which are softer and much more colorful. We model illumination using just a multivariate Gaussian on spherical harmonic illumination. Shown here are some example illuminations from our datasets and samples from our models, all rendered on Lambertian spheres. The samples looks superficially similar to the data, suggesting that our model is reasonable.

illumination, we parametrize log-shading rather than shading. This choice makes optimization easier as we don't have to deal with "clamping" illumination at 0, and it allows for easier regularization as the space of log-shading SH illuminations is surprisingly well-modeled by a simple multivariate Gaussian. Training our model is extremely simple: we fit a multivariate Gaussian to the SH illuminations in our training set. During inference, the cost we impose is the negative log-likelihood under that model:

$$h(L) = \lambda_L (L - \boldsymbol{\mu}_L)^{\mathrm{T}} \Sigma_L^{-1} (L - \boldsymbol{\mu}_L) \tag{7}$$

where $\boldsymbol{\mu}_L$ and Σ_L are the parameters of the Gaussian we learned, and λ_L is the multiplier on this prior (learned through cross-validation). Separate Gaussians and multipliers are learned from the illuminations in our two different datasets (see Section 8). See Figure 5 for a visualization of our training data and of samples from our learned models.

The Gaussians we learn for illumination mostly describe a low-rank subspace of SH coefficients. For this reason, it is important that we optimize in the space of whitened illumination. Whitened illumination is used as the internal representation of illumination during optimization, but is transformed to un-whitened space when calculating the loss function.

7 Multiscale Optimization

Here we present a novel multi-scale optimization method that is simpler, faster, and finds better local optima than the previous coarse-to-fine techniques we have presented [1, 22]. Our technique seems similar to multigrid methods [29], though it is extremely general and simple to implement. We will describe our technique in terms of optimizing $f(X)$, where f is some loss function and X is some n-dimensional signal.

Let us define $\mathcal{L}(X, h)$, which constructs a Laplacian pyramid from a signal, $\mathcal{L}^{-1}(Y, h)$, which reconstructs a signal from a Laplacian pyramid, and $\mathcal{G}(X, h)$, which constructs a Gaussian pyramid from a signal. Let h be the filter used in constructing and reconstructing these pyramids. Instead of minimizing $f(X)$ directly, we reparameterize X as $Y = \mathcal{L}(X, h)$, and minimize $f'(Y)$:

$$\begin{aligned}
&[\ell, \nabla_Y \ell] = f'(Y): \qquad (8)\\
&\quad X \leftarrow \mathcal{L}^{-1}(Y, h) \text{ // reconstruct the signal from the pyramid}\\
&\quad [\ell, \nabla_X \ell] \leftarrow f(X) \text{ // compute the loss and gradient with respect to the signal}\\
&\quad \nabla_Y \ell \leftarrow \mathcal{G}(\nabla_X \ell, h) \text{ // backpropagate the gradient onto the pyramid}
\end{aligned}$$

We then solve for $\hat{X} = \mathcal{L}^{-1}(\arg\min_Y f'(Y), h)$ using L-BFGS. Other gradient-based techniques could be used, but L-BFGS worked best in our experience.

The choice of h, the filter used for our Laplacian and Gaussian pyramids, is crucial. We found that 5-tap binomial filters work well, and that the choice of the magnitude of the filter dramatically affects multiscale optimization. If $\|h\|_1$ is small, then the coefficients of the upper levels of the Laplacian pyramid are so small that they are effectively ignored, and optimization fails. If $\|h\|_1$ is large, then the coarse scales of the pyramid are optimized and the fine scales are ignored. The filter that we found worked best is: $h = \frac{1}{4\sqrt{2}}[1, 4, 6, 4, 1]$, which has twice the magnitude of the filter that would normally be used for Laplacian pyramids. This increased magnitude biases optimization towards adjusting coarse scales before fine scales, without preventing optimization from eventually optimizing fine scales.

Note that this technique is substantially different from standard coarse-to-fine optimization, in that *all* scales are optimized simultaneously. As a result, we find much lower minima than standard coarse-to-fine techniques, which tend to keep coarse scales fixed when optimizing over fine scales. Our improved optimization also lets us use simple single-scale priors instead of multiscale priors, as was necessary in our previous work [1].

This optimization technique is used to solve Equations 1 and 5. When optimizing Equation 1 we initialize Z to 0 and L to $\boldsymbol{\mu}_L$, and optimize with respect to a vector that is a concatenation of $\mathcal{L}(Z, h)$ and a whitened version of L. For both problems, naive single-scale optimization fails badly.

8 Results

We evaluate our algorithm using the MIT intrinsic images dataset [1, 2]. The MIT dataset has very "laboratory"-like illumination — lights are white, and are placed at only a few locations relative to the object. Natural illuminations display much more color and variety (see Figures 5 and 6).

We therefore present an additional pseudo-synthetic dataset, in which we have rendered the objects in the MIT dataset using natural, colorful illuminations taken from the real world. We took all of the environment maps from the sIBL Archive[1], expanded that set of environment maps by shifting and mirroring them, and varying their contrast and saturation (saturation was only decreased, never increased), and produced spherical harmonic illuminations from the resulting environment maps. After removing similar illuminations, the illuminations were split into training and test sets. Each object in the MIT dataset was randomly assigned an illumination (such that training illuminations were assigned to training objects, etc), and each object was re-rendered under its new illumination, using that object's ground-truth shape and reflectance.

Our experiments can be seen in Table 1, in Figure 1, and in the supplementary material. We present four sets of experiments, with either the "laboratory" illumination of the basic MIT dataset or our "natural" illumination dataset, and with the illumination either known or unknown. We use the same training and test split as in [1], with our hyperparameters tuned to the training set, and with the same parameters used in all experiments and all figures.

For the known-lighting case our baselines are a "flat" baseline of $Z = 0$, four intrinsic image algorithms (these produce shading and reflectance images, and we then run the SFS algorithm of [1] using the recovered shading and known illumination to recover shape), the achromatic technique of our previous work [1], and the shape-from-contour algorithm of [1]. For unknown illumination, the only existing baseline is our previous work [1]. We present two simplifications of our model in which we apply the smoothness and entropy albedo priors of [1] to the RGB or YUV channels of color reflectance (while still using our absolute color and illumination priors), to demonstrate the importance of our multivari-

[1] http://www.hdrlabs.com/sibl/archive.html

Laboratory Illumination Dataset

Algorithm	N-MSE	s-MSE	r-MSE	rs-MSE	L-MSE	Avg.
Known Illumination						
Flat Baseline	0.6141	0.0572	0.0452	0.0354	-	0.0866
Retinex [2, 5] + SFS [1]	0.8412	0.0204	0.0186	0.0163	-	0.0477
Tappen *et al.* 2005 [14] + SFS [1]	0.7052	0.0361	0.0379	0.0347	-	0.0760
Shen *et al.* 2011 [15] + SFS [1]	0.9232	0.0528	0.0458	0.0398	-	0.0971
Gehler *et al.* 2011 [12] + SFS [1]	0.6342	0.0106	0.0101	0.0131	-	0.0307
Barron & Malik 2012A [1]	0.2032	0.0142	0.0160	0.0181	-	0.0302
Shape from Contour [1]	0.2464	0.0296	0.0412	0.0309	-	0.0552
Our Model (Complete)	0.2151	0.0066	0.0115	0.0133	-	0.0215
Unknown Illumination						
Barron & Malik 2012A [1]	0.1975	0.0194	0.0224	0.0190	0.0247	0.0332
Our Model (RGB)	0.2818	0.0090	0.0118	0.0149	0.0098	0.0213
Our Model (YUV)	0.2906	0.0110	0.0171	0.0182	0.0126	0.0263
Our Model (No Light Priors)	0.5215	0.0301	0.0273	0.0285	0.2059	0.0758
Our Model (No Absolute Prior)	0.3261	0.0124	0.0195	0.0189	0.0166	0.0301
Our Model (No Smoothness Prior)	0.2727	0.0105	0.0179	0.0223	0.0125	0.0270
Our Model (No Entropy Model)	0.2865	0.0109	0.0161	0.0152	0.0141	0.0255
Our Model (White Light)	0.2221	0.0082	0.0112	0.0136	0.0085	0.0188
Our Model (Complete)	0.2793	0.0075	0.0118	0.0144	0.0100	0.0205

Natural Illumination Dataset

Algorithm	N-MSE	s-MSE	r-MSE	rs-MSE	L-MSE	Avg.
Known Illumination						
Flat Baseline	0.6141	0.0246	0.0243	0.0125	-	0.0463
Retinex [2, 5] + SFS [1]	0.4258	0.0174	0.0174	0.0083	-	0.0322
Tappen *et al.* 2005 [14] + SFS [1]	0.6707	0.0255	0.0280	0.0268	-	0.0599
Gehler *et al.* 2011 [12] + SFS [1]	0.5549	0.0162	0.0150	0.0105	-	0.0346
Gehler *et al.* 2011 [12] + [11] + SFS [1]	0.6282	0.0163	0.0164	0.0106	-	0.0365
Barron & Malik 2012A [1]	0.2044	0.0092	0.0094	0.0081	-	0.0195
Shape from Contour [1]	0.2502	0.0126	0.0163	0.0106	-	0.0271
Our Model (Complete)	0.0867	0.0022	0.0017	0.0026	-	0.0054
Unknown Illumination						
Barron & Malik 2012A [1]	0.2172	0.0193	0.0188	0.0094	0.0206	0.0273
Our Model (RGB)	0.2373	0.0086	0.0072	0.0065	0.0104	0.0159
Our Model (YUV)	0.3064	0.0095	0.0088	0.0072	0.0110	0.0183
Our Model (No Light Priors)	0.3722	0.0141	0.0149	0.0118	0.1491	0.0424
Our Model (No Absolute Prior)	0.1914	0.0124	0.0106	0.0036	0.0136	0.0165
Our Model (No Smoothness Prior)	0.2700	0.0084	0.0071	0.0065	0.0090	0.0157
Our Model (No Entropy Prior)	0.2911	0.0080	0.0067	0.0054	0.0109	0.0155
Our Model (White Light)	0.6268	0.0211	0.0207	0.0089	0.0647	0.0437
Our Model (Complete)	0.2348	0.0060	0.0049	0.0042	0.0084	0.0119

Table 1. A comparison of our model against others, on the "laboratory" MIT intrinsic images dataset [1, 2] and our own "natural" illumination variant, with the illumination either known or unknown. Shown are the geometric means of five error metrics (excluding L-MSE when illumination is known) across the test set, and an "average" error (the geometric mean of the other mean errors). N-MSE, L-MSE, s-MSE, and r-MSE measure shape, illumination, shading, and reflectance errors, respectively, and rs-MSE is the error metric of [2], (where it is called "LMSE") which measures shading and reflectance errors. These metrics are explained in detail in the supplementary material.

ate models. We also present an ablation study in which priors on reflectance or illumination are removed, and in which illumination is forced to be white (achromatic) during inference.

For our "natural" illumination dataset, we use the same baselines (except for [15], as their code was not available). We also evaluate against the intrinsic image algorithm of Gehler *et al.* [12] after having run a contemporary white-balancing algorithm [11] on the input image, which shows that a "color constancy" algorithm does not fully address natural illumination for this task.

For the "laboratory" case, our algorithm is the best-performing algorithm whether or not illumination is known. Surprisingly, performance is slightly better when illumination is unknown, possibly because optimization is able to find more accurate shapes and reflectances when illumination is allowed to vary. The shading and reflectances produced by Gehler *et al.* [12] seem equivalent to ours with regards to rs-MSE, s-MSE, and r-MSE (the metrics that consider shading and reflectance). However, when SFS is performed on their shading, the resulting shapes are much worse than ours in terms of N-MSE (the metric that consider shape). This appears to happen because, though this algorithm produces very accurate-looking shading images, that shading is often inconsistent with the known illumination or inconsistent with itself, causing SFS to produce a contorted shape. We see that treating color intelligently works better than a naive RGB or YUV model, and much better using only grayscale images (Barron and Malik 2012A [1]). The ablation study shows that all priors contribute positively: removing any reflectance prior hurts performance by 30-50%, and removing the illumination prior completely cripples the algorithm. Constraining the illumination to be white helps performance on this dataset, but would presumably make our model generalize worse on real-world images.

For the "natural" illumination case, we outperform all other algorithms by a very large margin — our error is less than 40% of the best-performing intrinsic image algorithm (20% if illumination is known). This shows the necessity of explicitly modeling chromatic illumination. While our complete model outperforms all other models, the "white light" case often underperforms many other models, even the achromatic model of [1]. This shows that attempting to use color information in the presence of non-white illumination without taking into consideration the color of illumination can actually *hurt* performance. For example, in the "laboratory" MIT dataset, our model performs equivalently to Gehler *et al.* in some error metrics, but in the "natural" illumination case, Gehler *et al.* and the other intrinsic image algorithms all perform significantly worse than our model. Because these intrinsic image algorithms rely heavily on color cues and assume illumination to be white, they suffer greatly when faced with colorful "natural" illuminations. In contrast, our model actually performs as well or better in the "natural" illumination case, as it can exploit color illumination to better disambiguate between shading and illumination (Figure 2), and produce higher-quality shape reconstructions (Figure 6). See the supplementary material for many examples of the output of our model and others, for all four experiments.

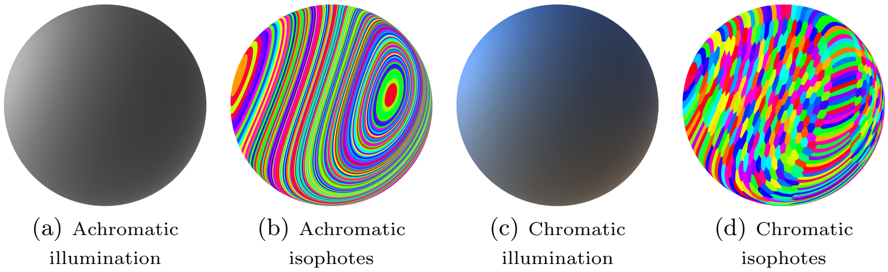


Fig. 6. Chromatic illumination dramatically helps shape estimation. Achromatic isophotes (K-means clusters of log-RGB values) are very elongated, while chromatic isophotes are usually more tightly localized. Therefore, under achromatic lighting a very wide range of surface orientations appear similar, but under chromatic lighting only similar orientations appear similar.

9 Conclusion

We have extended our previous work [1] to present the first unified model for recovering shape, reflectance, and chromatic illumination from a single image, unifying the previously disjoint problems of color constancy, intrinsic images, and shape-from-shading. We have done this by introducing novel priors on local smoothness, global entropy, and absolute color, a novel prior on illumination, and an efficient multiscale optimization framework for jointly optimizing over shape and illumination.

By solving this one unified problem, our model outperforms all previously published algorithms for intrinsic images and shape-from-shading, on both the MIT dataset and our own "naturally" illuminated variant of that dataset. When faced with images produced under natural, chromatic illumination, the performance of our algorithm improves dramatically because it can exploit color information to better disambiguate between shading and reflectance variation, and to improve shape estimation. In contrast, other intrinsic image algorithms (which incorrectly assume illumination to be achromatic) perform very poorly in the presence of natural illumination. This suggests that the "intrinsic image" problem formulation may be fundamentally limited, and that we should refocus our attention towards developing models that jointly reason about shape and illumination in addition to shading and reflectance.

Acknowledgements: J.B. was supported by NSF GRFP and ONR MURI N00014-10-10933.

References

1. Barron, J.T., Malik, J.: Shape, albedo, and illumination from a single image of an unknown object. CVPR (2012)
2. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. ICCV (2009)

3. Helmholtz, H.v.: Treatise on physiological optics, 2 vols., translated. Washington, DC: Optical Society of America. (1924)
4. Land, E.H., McCann, J.J.: Lightness and retinex theory. JOSA (1971)
5. Horn, B.K.P.: Determining lightness from an image. Computer Graphics and Image Processing (1974)
6. Forsyth, D.A.: A novel algorithm for color constancy. IJCV (1990)
7. Maloney, L.T., Wandell, B.A.: Color constancy: a method for recovering surface spectral reflectance. JOSA A (1986)
8. Klinker, G., Shafer, S., Kanade, T.: A physical approach to color image understanding. IJCV (1990)
9. Finlayson, G., Hordley, S., Hubel, P.: Color by correlation: a simple, unifying framework for color constancy. TPAMI (2001)
10. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. JOSA A (1997)
11. Gijsenij, A., Gevers, T., van de Weijer, J.: Generalized gamut mapping using image derivative structures for color constancy. IJCV (2010)
12. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schoelkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. NIPS (2011)
13. Barrow, H., Tenenbaum, J.: Recovering intrinsic scene characteristics from images. Computer Vision Systems (1978)
14. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. TPAMI (2005)
15. Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. CVPR (2011)
16. Horn, B.K.P.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, MIT (1970)
17. Brooks, M.J., Horn, B.K.P.: Shape from shading. MIT Press (1989)
18. Zhang, R., Tsai, P., Cryer, J., Shah, M.: Shape-from-shading: a survey. TPAMI (1999)
19. Johnson, M.K., Adelson, E.H.: Shape estimation in natural illumination. CVPR (2011)
20. Gilchrist, A.: Seeing in Black and White. Oxford University Press (2006)
21. Boyaci, H., Doerschner, K., Snyder, J.L., Maloney, L.T.: Surface color perception in three-dimensional scenes. Visual Neuroscience (2006)
22. Barron, J.T., Malik, J.: High-frequency shape and albedo from shading using natural image statistics. CVPR (2011)
23. Shen, L., Yeo, C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. CVPR (2011)
24. Alldrin, N., Mallick, S., Kriegman, D.: Resolving the generalized bas-relief ambiguity by entropy minimization. CVPR (2007)
25. Finlayson, G.D., Drew, M.S., Lu, C.: Entropy minimization for shadow removal. IJCV (2009)
26. Omer, I., Werman, M.: Color lines: Image specific color representation. CVPR (2004)
27. Principe, J.C., Xu, D.: Learning from examples with quadratic mutual information. Workshop on Neural Networks for Signal Processing (1998)
28. Chen, J., Paris, S., Durand, F.: Real-time edge-aware image processing with the bilateral grid. SIGGRAPH (2007)
29. Terzopoulos, D.: Image analysis using multigrid relaxation methods. TPAMI **8** (1986) 129–139