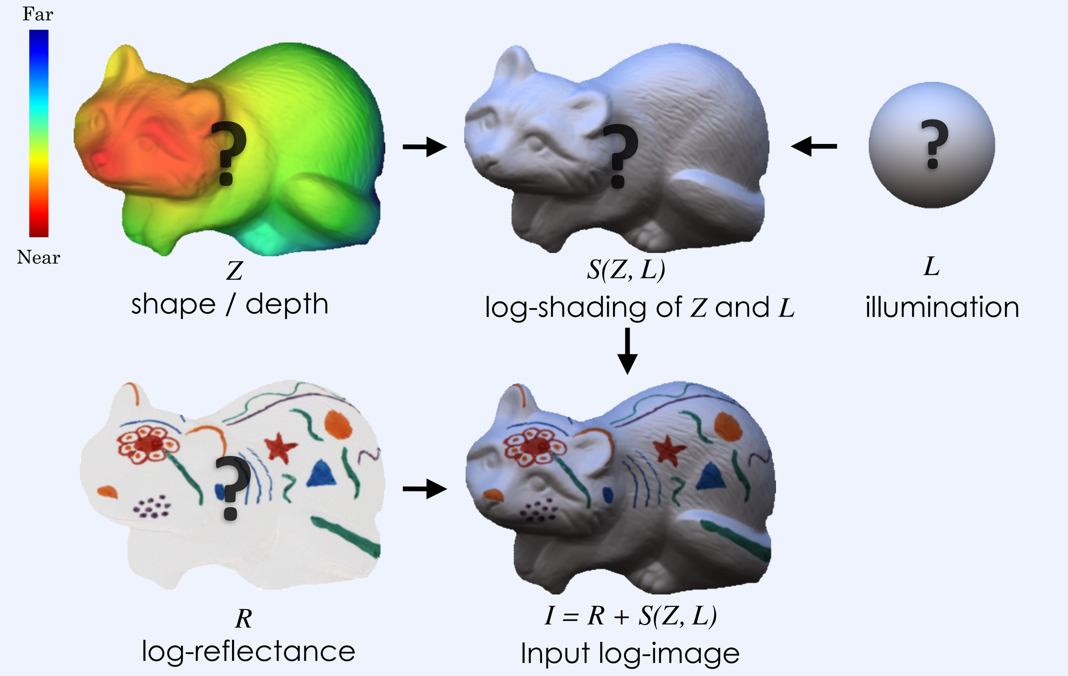
Color Constancy, Intrinsic Images, and Shape Estimation

Jonathan T Barron & Jitendra Malik
UC Berkeley

We present the first unified model for recovering shape, chromatic illumination, and reflectance from a single image.

SIRFS

shape, illumination, and reflectance from shading



$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the most likely shape, reflectance, and illumination that explains the input image"

Our model is an extension of our previous work [1], which dealt with the grayscale version of this problem. By modeling color illumination and reflectance, our model implicitly addresses color constancy.

We outperform previously published algorithms for color constancy, intrinsic images, and shape-from-shading on the MIT intrinsic images dataset [2], and on our own "naturally" illuminated version of that dataset.

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)

Reflectance

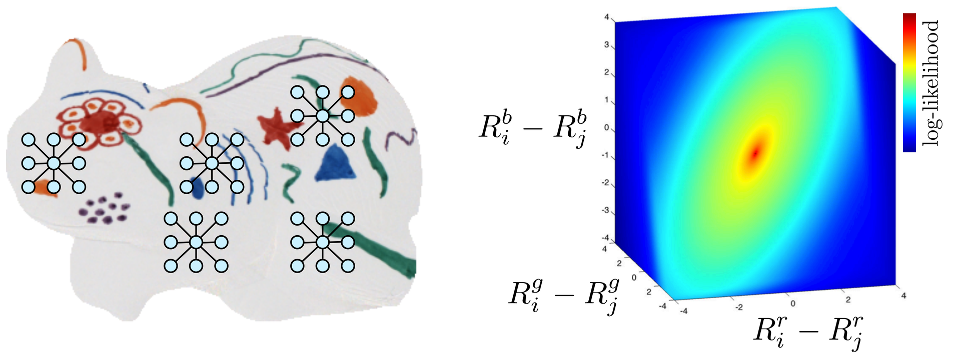
Our prior on reflectance is three components:

$$g(R) = \lambda_s g_s(R) + \lambda_e g_e(R) + \lambda_a g_a(R)$$

We assume reflectance images are locally **smooth**, that the distribution of reflectance across an image is **low-entropy**, and that some **colors** are more likely than others.

Smoothness

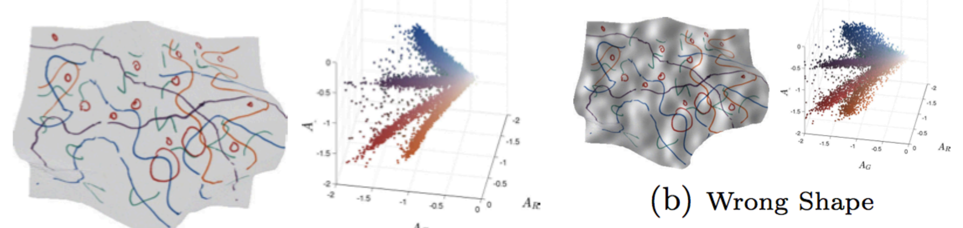
We fit a multivariate Gaussian scale mixture on the differences between nearby reflectance pixels.



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

Minimal Entropy

We minimize the multivariate Renyi entropy of the reflectance image.



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

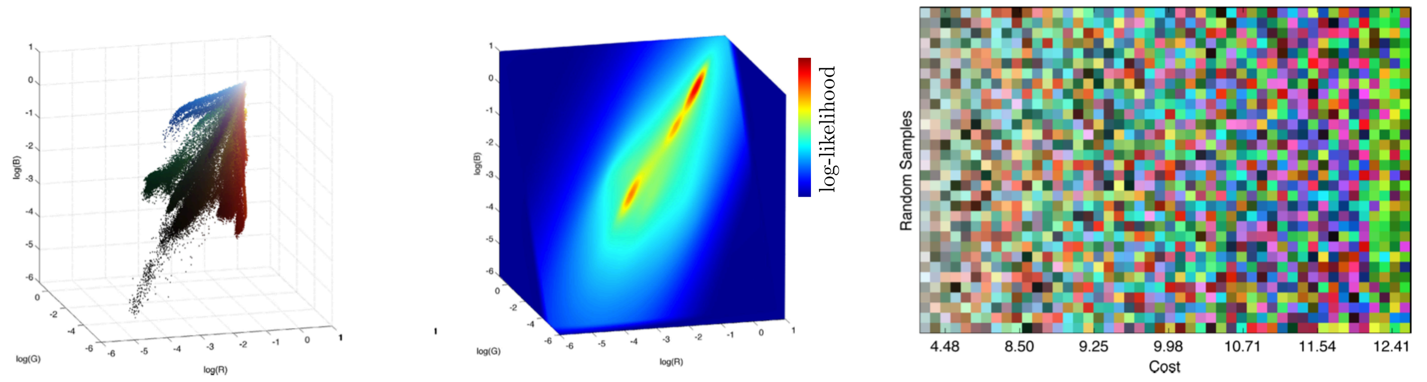
(b) Wrong Shape $g_e(R) = 1.325$

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

$$g_e(R) = -\log\left(\sum_i \sum_j \exp\left(-\frac{\|WR_i - WR_j\|_2^2}{4\sigma_e^2}\right)\right)$$

Absolute Color

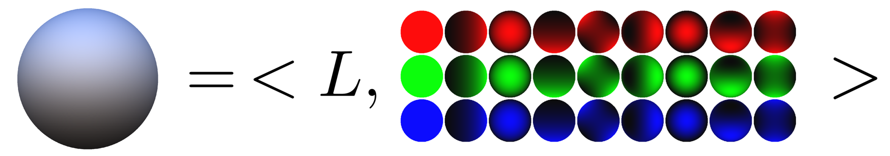
We model the distribution of reflectance pixels using a 3D thin-plate spline.



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

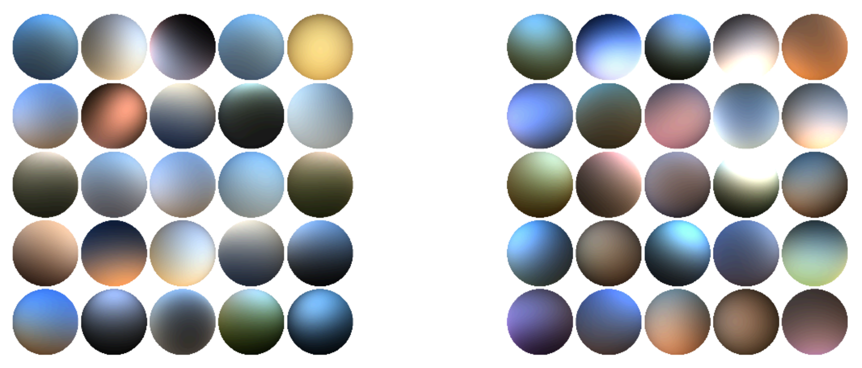
Illumination

Global Illumination is well-modeled with spherical harmonics.



Our prior on illumination is a multivariate Gaussian on spherical harmonic coefficients:

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

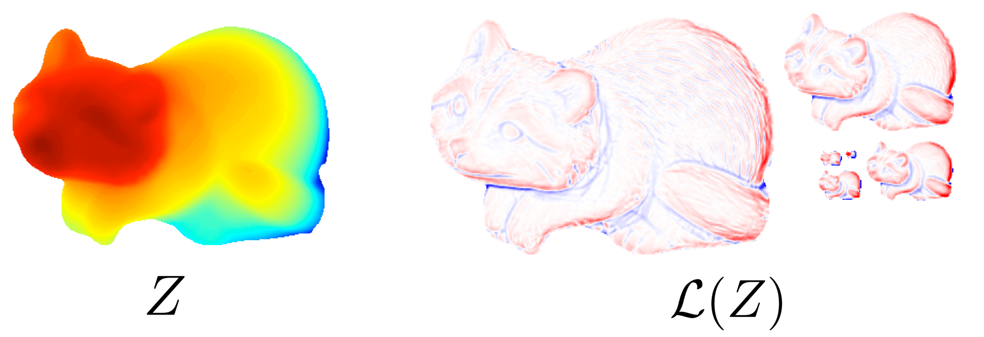


Natural Illuminations from our dataset

Samples from a Gaussian fit to the training set

Optimization

Straightforward gradient-based optimization (L-BFGS) of Z fails, and coarse-to-fine works poorly.



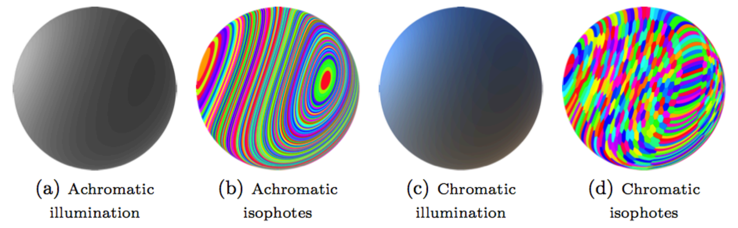
Instead, we optimize over $\mathcal{L}(Z)$, a Laplacian pyramid of Z.
Fast (5 min) and effective multiscale optimization, trivial code:

$[\ell, \nabla_Y \ell] = f'(Y)$: // Loss with respect to a pyramid:
$Z \leftarrow \mathcal{L}^{-1}(Y)$ // Flatten the pyramid to an image
$[\ell, \nabla_Z \ell] \leftarrow f(Z)$ // Compute loss with respect to the image
$\nabla_Y \ell \leftarrow \mathcal{G}(\nabla_Z \ell)$ // Backpropagate the gradient onto the pyramid

Evaluation

We evaluate on the MIT Intrinsic Images dataset [1,2], and on our own "naturally" illuminated version (based on the sIBL dataset).

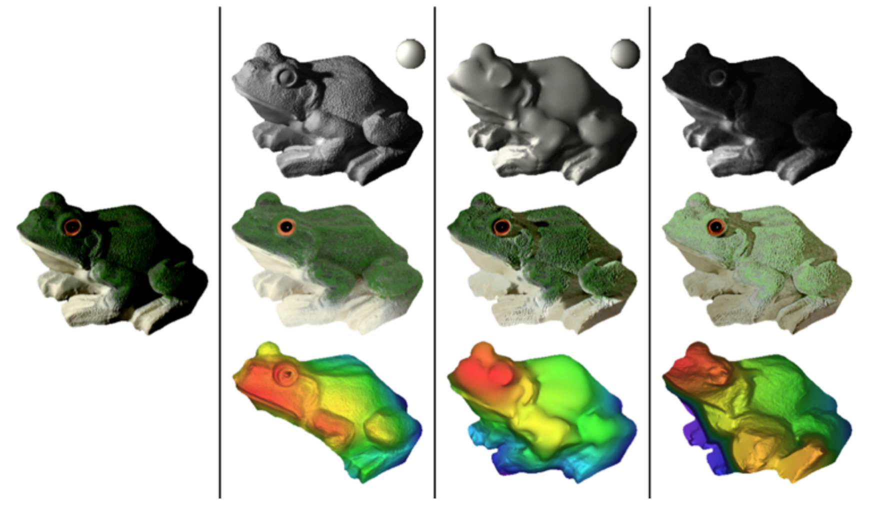
Natural illumination is colorful, which is a powerful cue for shape:



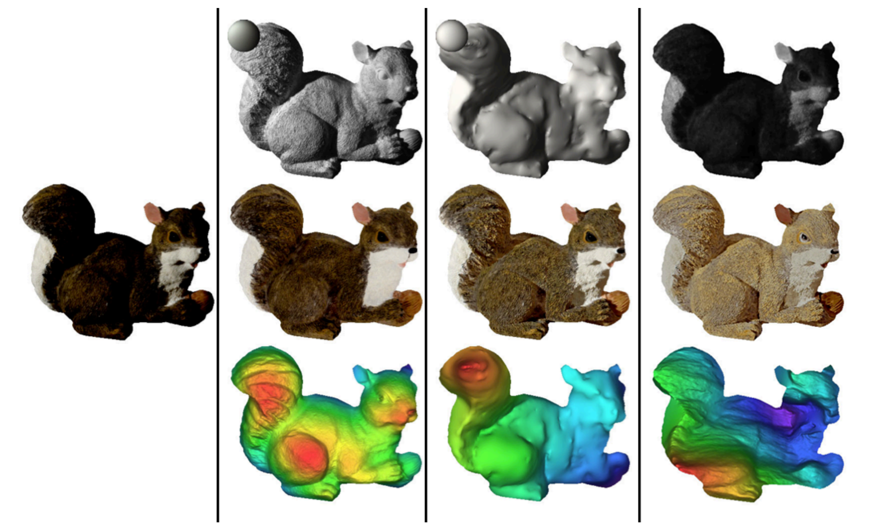
But intrinsic image algorithms assume white light, and therefore break on natural images (and color constancy algorithms do not fix this problem). Our algorithm, in contrast, works **better** in the presence of natural illumination.

"Laboratory"-style Illumination

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 (Complete)	0.2793	0.0075	0.0112	0.0136	0.0085	0.0188



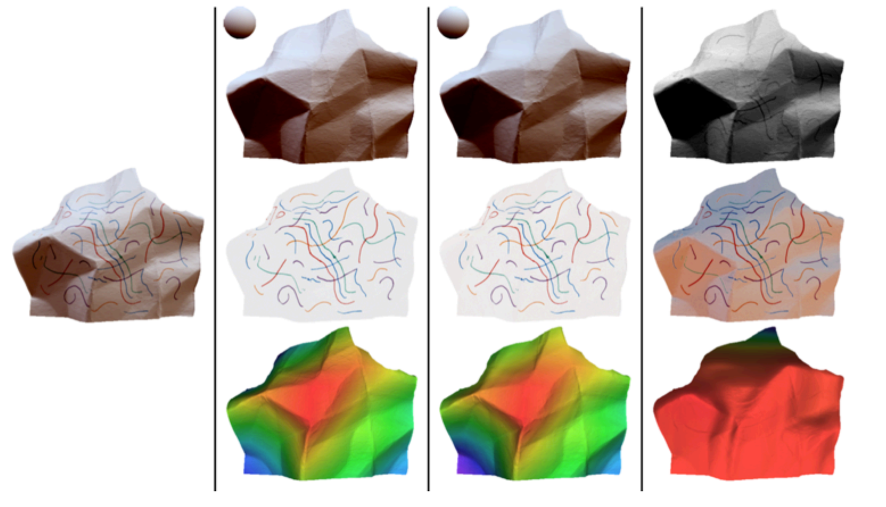
Input Image Ground Truth **Our Model** Gehler *et al*



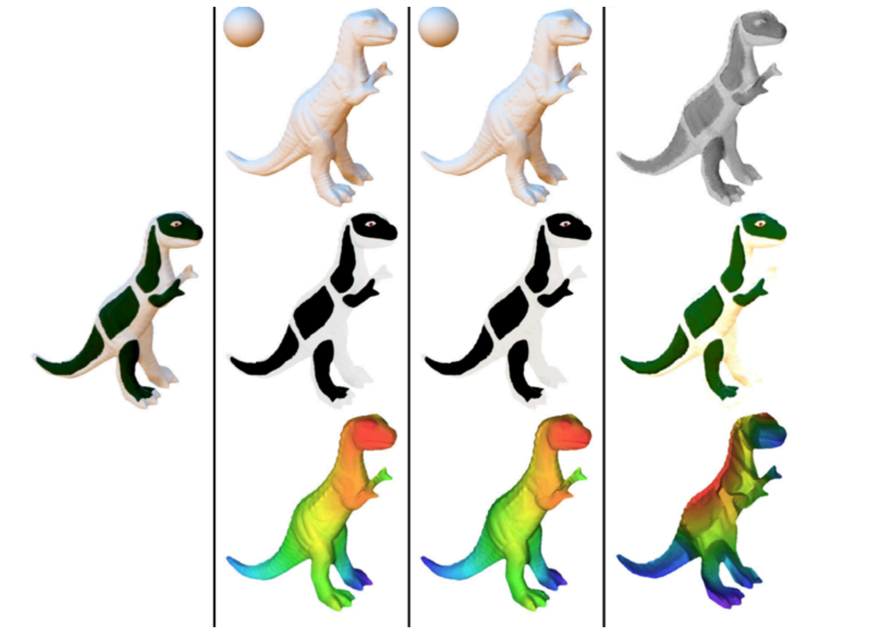
Input Image Ground Truth **Our Model** Gehler *et al*

"Natural" Illumination

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 (Complete)	0.2348	0.0060	0.0049	0.0042	0.0084	0.0119



Input Image Ground Truth **Our Model** Gehler *et al*



Input Image Ground Truth **Our Model** Gehler *et al*