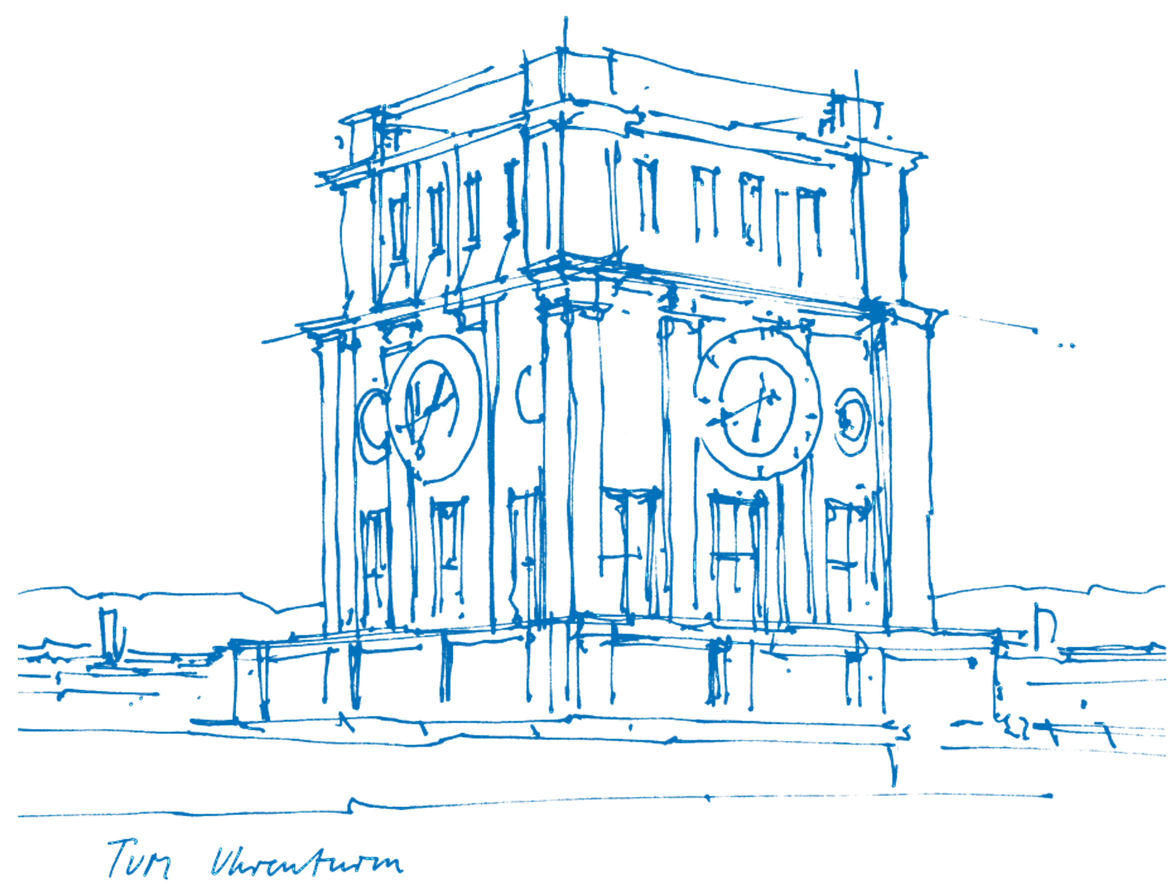
Color Constancy, Intrinsic Images, and Shape Estimation

Jonathan T. baron and Jitendra Malik
ECCV'12 Proceedings of the 12th European conference on Computer Vision - Volume Part IV

Madwaraj Hatwar
Technische Universität München
16. July 2019



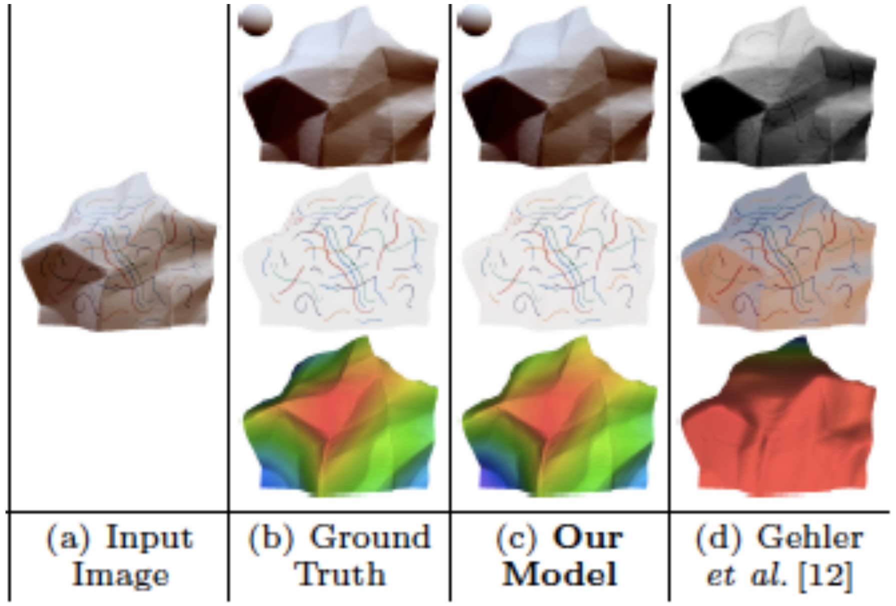
Introduction

Color constancy problem - Decomposing an image into illuminant color and surface color.

Intrinsic images problem - Decomposing a single image into its constituent shape, reflectance and illumination etc.

SIRFS(shape, illumination and reflectance from shading) - The first unified model for recovering shape, chromatic illumination, and reflectance from a single image.

Reflectance - effectiveness of the surface in reflecting the radiant energy.



Problem formulation

Assume Lambertian reflectance model, so I = R + S(Z,L). log Intensity (I) is observed and S is defined.

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

g(R) - cost of reflectance R
f(Z) - cost of shape Z
h(L) - cost of illumination L

Write R = I - S(Z,L) and minimise the unconstrained optimization problem to produce depth map Z’, and illumination L’. Then calculate R’ = I - S(Z’,L’).

Reflectance

Our prior on reflectance has three components:

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

λ are the weights learned using cross-validation

$g_s(R)$ - prior on local smoothness

$g_e(R)$ - prior on global entropy of reflectance

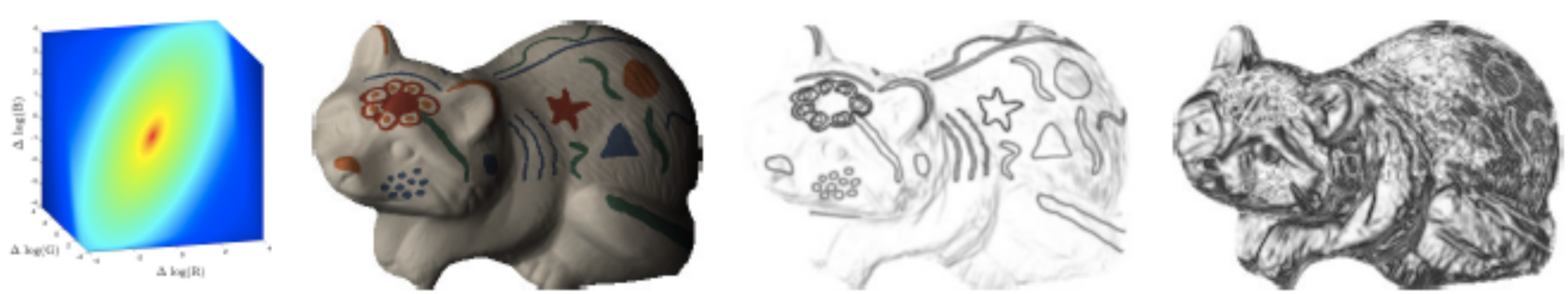
$g_a(R)$ - absolute prior on each pixel to address color constancy

Reflectance smoothness

The prior on reflectance smoothness is a multivariate Gaussian scale mixture (GSM) placed on the differences between each reflectance pixel and its neighbours.

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

N(i) is the 5 x5 neighbourhood 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), Σ is the covariance matrix of the entire GSM. α and σ are the coefficients of the GSM. GSM is learned using the training data.



(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

Global entropy

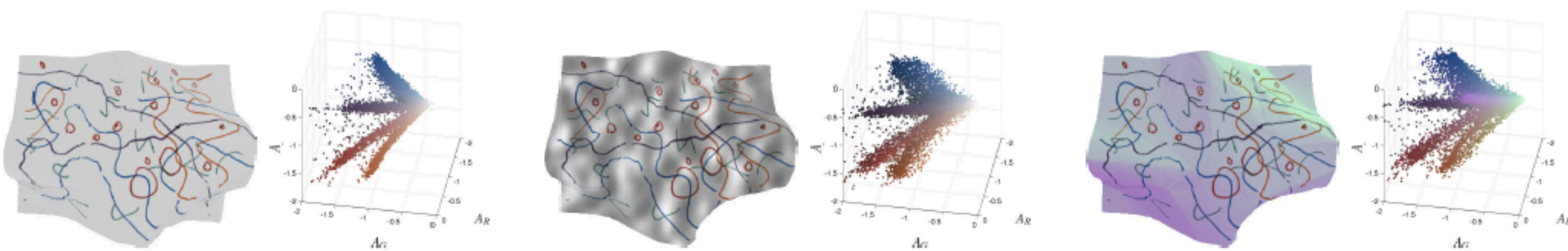
Assume reflectance image of a single object is clustered in RGB space.

We minimize the multivariate Quadratic entropy of the reflectance image.

Entropy function should be anisotropic to take care of separate RGB channels.

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

W is the whitening transformation learned from training reflectance images. σ_e is tuned through cross-validation.



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

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

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

Absolute Colour

Assume not all colours are equally likely.

Impose prior on absolute reflectance - log RGB values of each pixel.

Fit a 3D thin-plate spline (TPS) to the distribution of whitened log RGB values.

$$\underset{\mathrm{F}}{\text{minimize}} \quad \left(\sum_{i,j,k} \mathrm{F}_{i,j,k} \cdot \mathrm{N}_{i,j,k} \right) + \log \left(\sum_{i,j,k} \exp\left(-\mathrm{F}_{i,j,k}\right) \right) + \lambda \sqrt{J(\mathrm{F}) + \epsilon^2}$$

$$J(\mathrm{F}) = \mathrm{F}_{xx}^2 + \mathrm{F}_{yy}^2 + \mathrm{F}_{zz}^2 + 2\mathrm{F}_{xy}^2 + 2\mathrm{F}_{yz}^2 + 2\mathrm{F}_{xz}^2$$

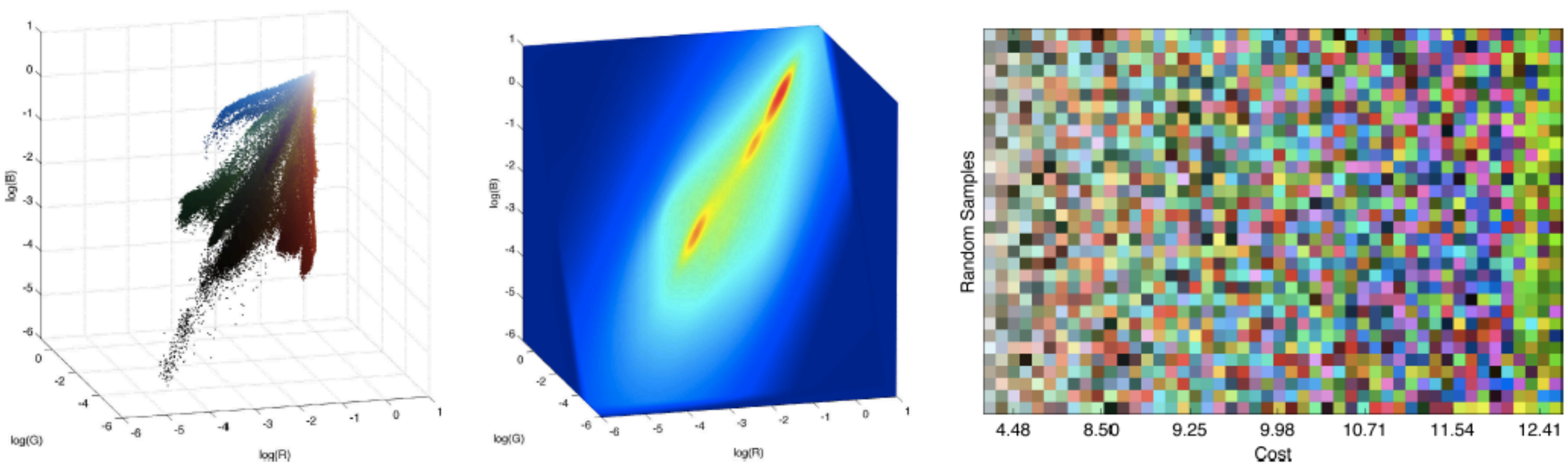
F is a 3D TPS describing cost, N is a 3D histogram of the whitened log-RGB reflectance in our training data, and J() is the TSP bending energy cost.

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

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

where F(WR) is the value of F at the coordinates specified by the 3-vector WR, the whitened reflectance at pixel i.



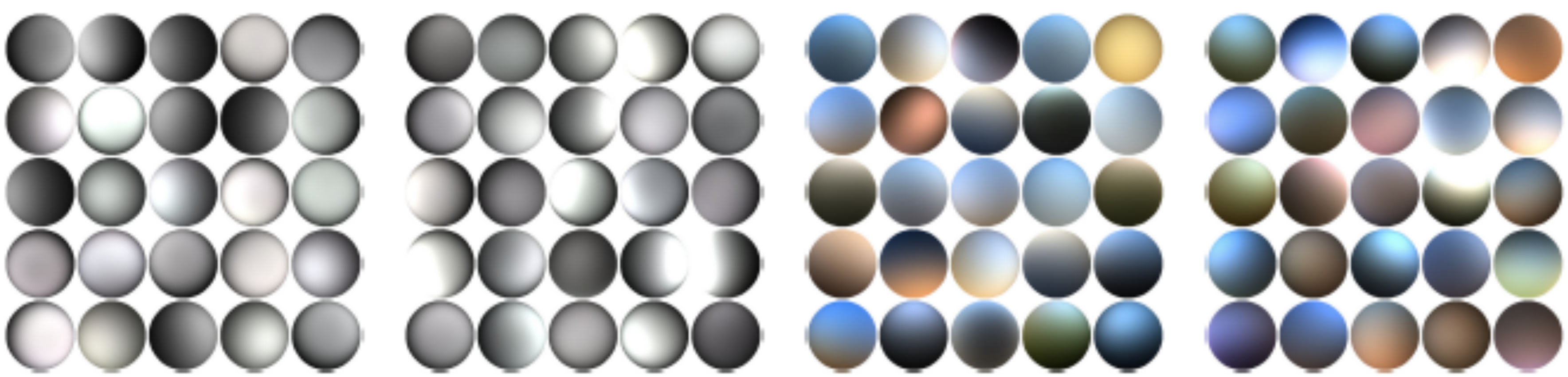
(a) Training reflectances (b) Our PDF of reflectance (c) Reflectances sorted by cost

Illumination

Global Illumination is modelled with spherical harmonics.
Cost function of our model:

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

where μ_L and Σ_L are the Gaussian parameters which we learned and λ is learned through cross-validation.



(a) "Lab" Data (b) "Lab" Samples (c) "Natural" Data (d) "Natural" Samples

Prior on shape

Prior on shape is a linear combination of the following terms

$$f(Z) = \lambda_f f_f(Z) + \lambda_c f_c(Z) + \lambda_k f_k(Z)$$

$f_f(Z)$ is a flatness term,

$f_c(Z)$ encourages shapes to face outwards at the boundaries

$f_k(Z)$ is a smoothness term.

λ are learned through cross-validation.

Flatness

The prior prefers a flat shape, by minimising the slant of Z.

$$f_f(Z) = -\sum_{x,y} \log\left(2N^z_{x,y}(Z)\right)$$

Where $N^z_{x,y}$ (Z) is the z-component of the surface normal of Z at position (x, y).

If we have observed a surface in space, it is more likely that it faces the observer($N^z_{x,y}$=1) than that it is perpendicular to the observer ($N^z_{x,y}$ = 0).

Occluding boundary

We minimise the following cost function.

$$f_c(Z) = \sum_{i \in C} \sqrt{\left(N_i^x(Z) - n_i^x\right)^2 + \left(N_i^y(Z) - n_i^y\right)^2}$$

N is the surface normal on the depth map Z, and n is local normal to the occluding boundary in the image plane.

Variation of mean curvature

The prior penalises change in mean curvature.

$$f_k(Z) = \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \alpha_k \mathcal{N}\left(H(Z)_i - H(Z)_j ; 0, \sigma_k\right) \right)$$

where N(i) is the 5x5 neighborhood around pixel i, $H(Z)_i$ is the mean curvature of shape Z, $H(Z)_i - H(Z)_j$ is the difference between the mean curvature at pixel i and pixel j, k = 40 (the GSM has 40 discrete Gaussians), α and σ are the coefficients. The GSM is learned using the training set.

Optimization

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

Instead, we optimize over , a Laplacian pyramid L(X,h) where X is the variable to be optimised and h is the filter.

$[\ell, \nabla_Y \ell] = f'(Y)$ // compute loss with respect to pyramid

$X \leftarrow \mathcal{L}^{-1}(Y, h)$ // reconstruct the signal from the pyramid

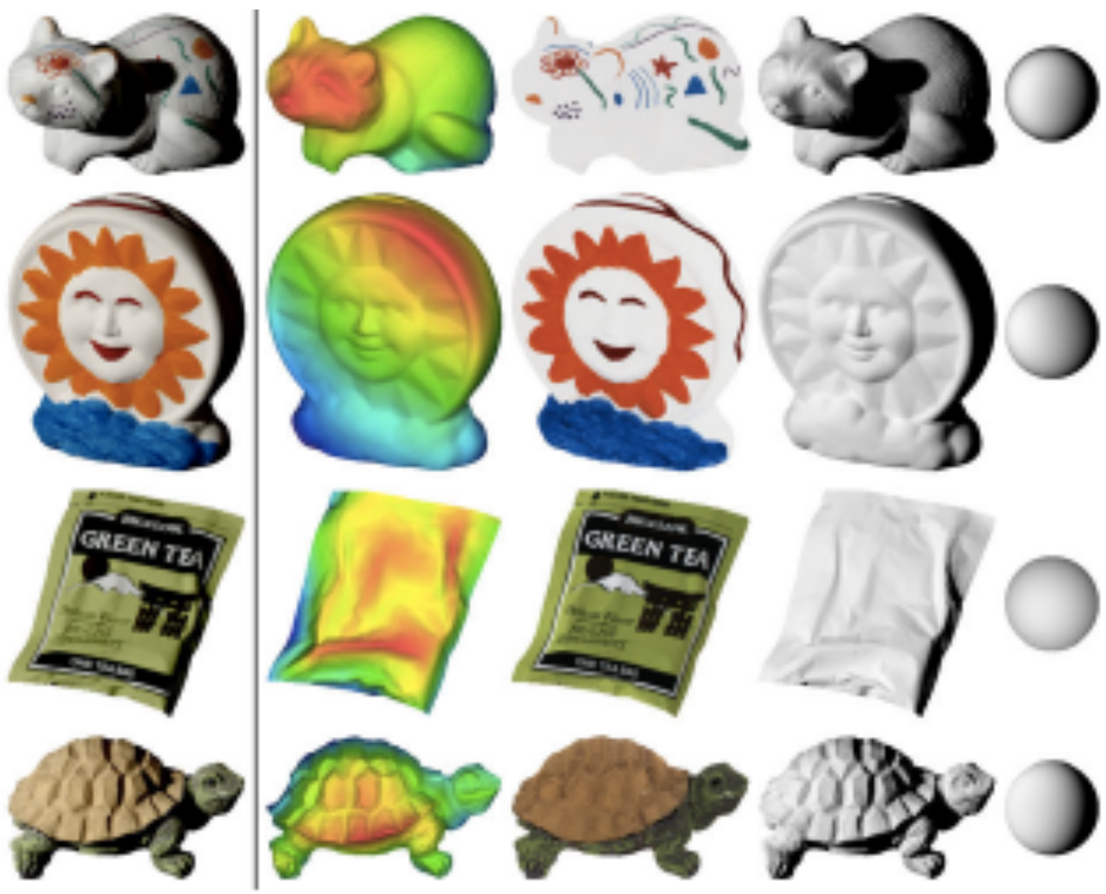
$[\ell, \nabla_X \ell] \leftarrow f(X)$ // compute the loss and gradient with respect to the signal

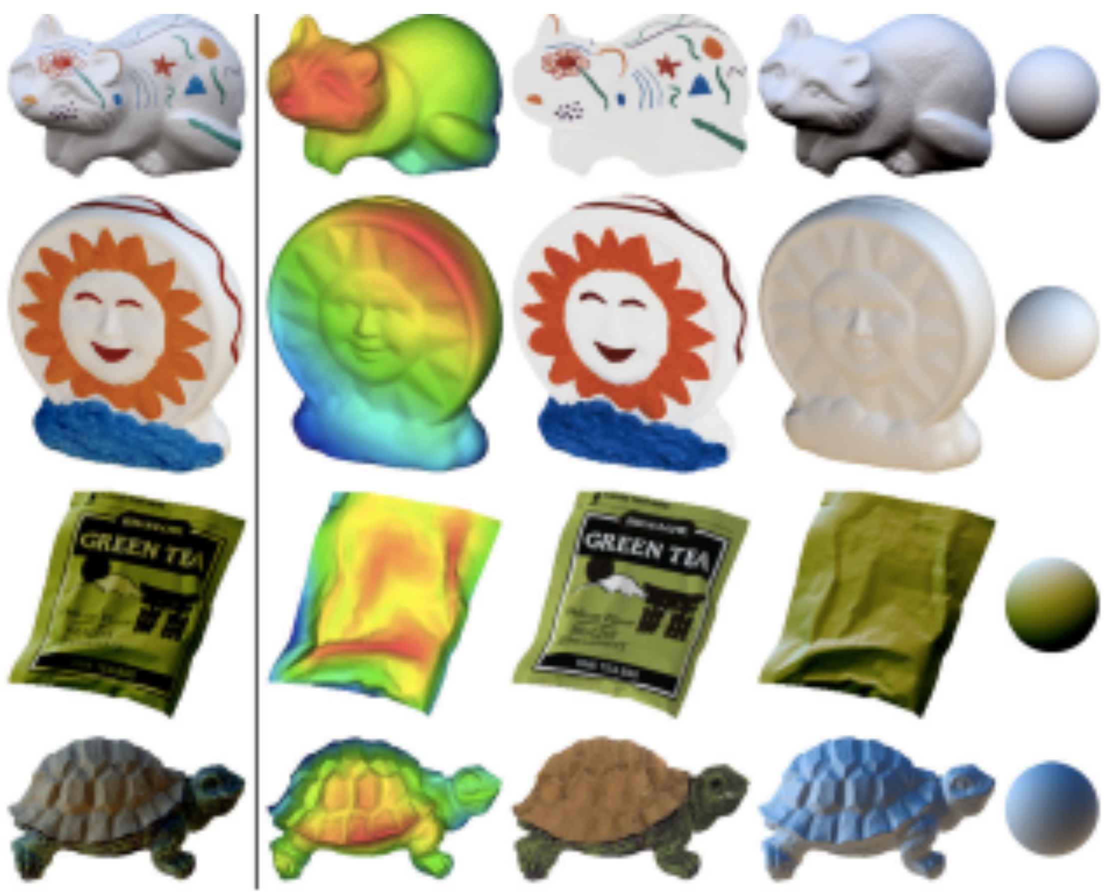
$\nabla_Y \ell \leftarrow \mathcal{G}(\nabla_X \ell, h)$ // backpropagate the gradient onto the pyramid

solve for $\hat{X} = \mathcal{L}^{-1}(\arg\min_Y f'(Y), h)$ using L-BFGS.

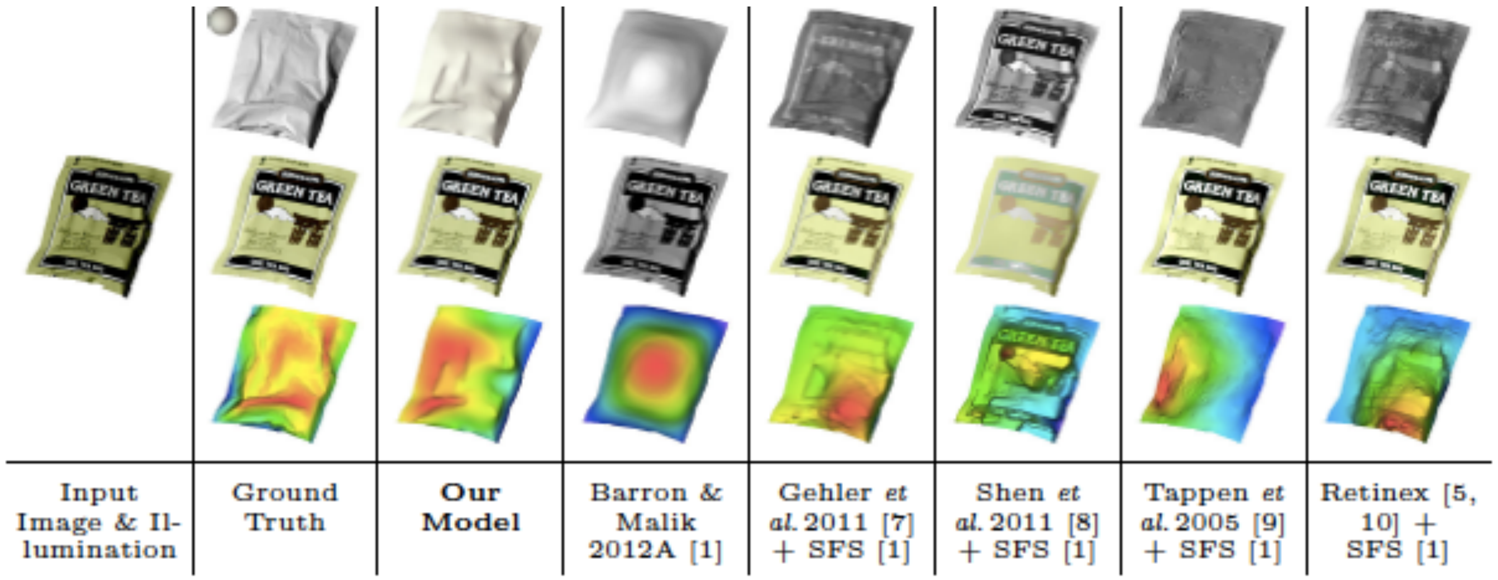
h is a binomial filter of length 5 and the filter that worked best is h = [1, 4, 6, 4, 1]/(4$\sqrt{2}$).

Results

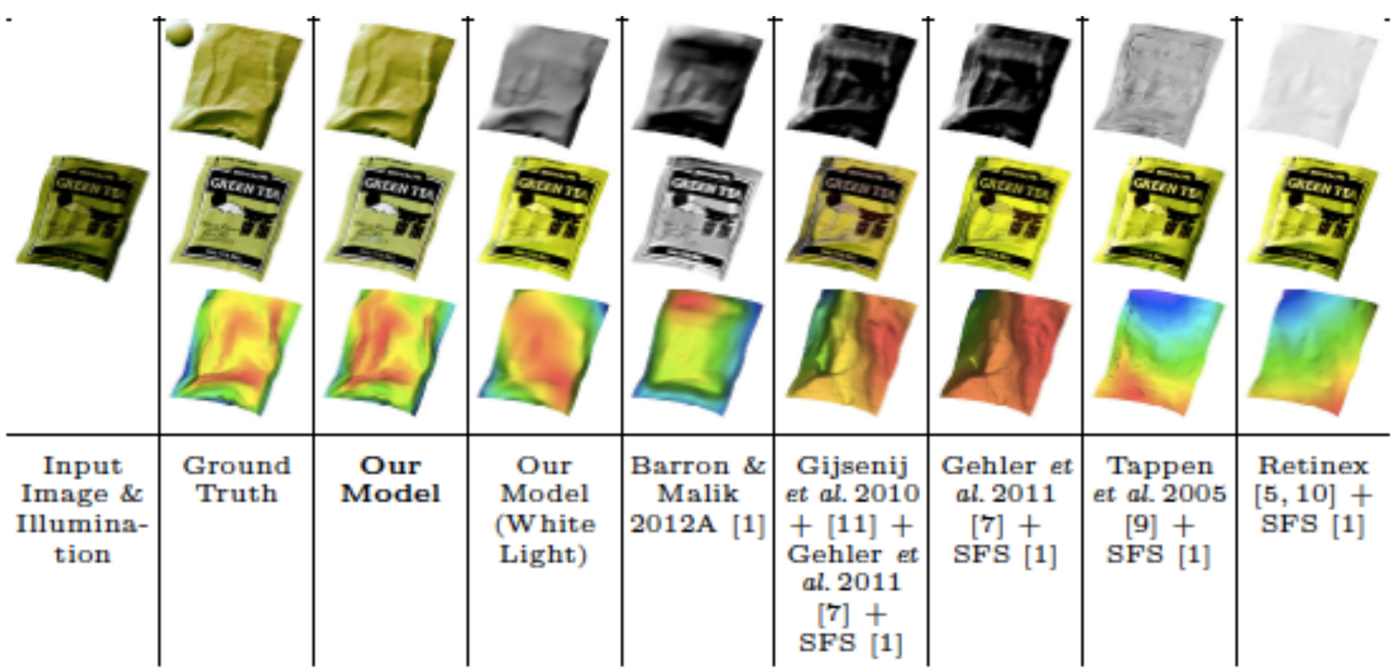




- Laboratory illumination dataset Natural illumination dataset



Comparing the proposed model with different models for Laboratory illumination



Comparing the proposed model with different models for Natural illumination

Evaluation of the algorithm on Laboratory illumination and natural illumination datasets with known and unknown illumination.

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

THANK YOU