Color Constancy, Intrinsic Images and Shape Estimation

Madwaraj Hatwar

Computational Science and Engineering - Technische Universität München

Abstract

The first unified model to recover shape, chromatic illumination and reflectance from a single image is presented here. This approach requires a modified problem formulation, novel priors on reflectance, illumination, and shape. This algorithm outperforms all previously established algorithms for intrinsic image decomposition and shape-from-shading on the MIT intrinsic images dataset and also on the naturally illuminated version of the dataset.

1 Introduction

Color constancy problem means decomposing the image into illuminant colour and surfce colour [2]. This problem is a subset of bigger problem called intrinsic images problem - decomposing image into different intrinsic images layers such as shape, reflectance, shading, illumination etc. All the previous approaches recover some of the layers by ignoring the rest. This paper defines the first unified approach called SIRFS (shape, illumination, and reflectance from shading) to recover shape, illumination, and reflectance from a single image. The model is tested on MIT intrinsic images dataset with laboratory illumination and an own version of the dataset with natural illumination. An example is shown in Fig. 1. Given just the masked input image, the model produces the illumination model, depth-map, reflectance image, and shading image that together exactly explain the input image.

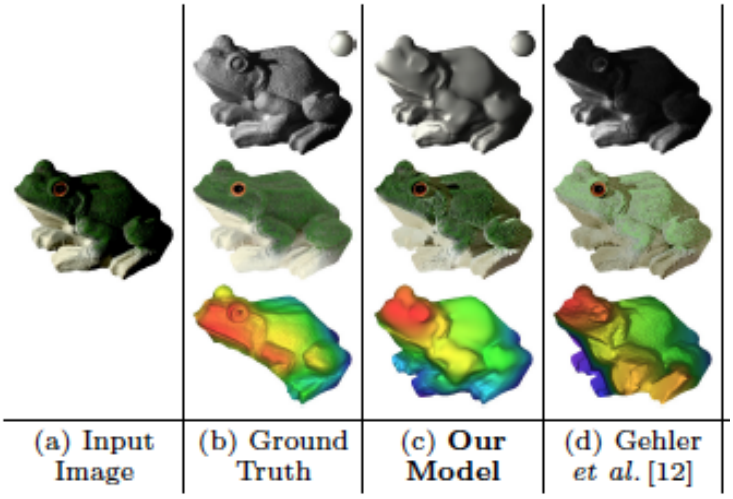


Figure 1: An object from the dataset. Illumination is rendered on a sphere, and shape is shown as a pseudocolor visualization where red is near and blue is far

2 Problem Formulation

The problem is formulated by optimizing over the depth map, reflectance image, and model of illumination such that cost functions on those quantities are minimized and the input image is exactly recreated by the output shape, albedo and illumination.
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 function of Z and L. Assuming Lambertian reflectance, the log-intensity image I is equal to R + S(Z,L). I is observed and S() is defined, but Z, R, and L are unknown. We want the most likely explanation for image I, which can be obtained by solving the optimization problem.

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

where g(R) is the cost of reflectance R, f(Z) is the cost of shape Z, and h(L) is the cost of illumination L.
The constraint is removed by writing R = I-S(Z,L) and minimize the resulting unconstrained optimization to produce depth map Z' and illumination L'. This will give us the reflectance image R' = I - S(Z',L').
To address the additional complications due to allowing non-white illuminations and to use the additional information in the colour reflectance images, we define a prior on reflectance g(R) as:

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

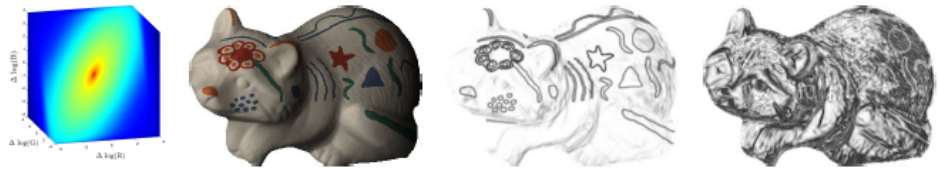
where λ are the weights learned using cross-validation on the training set. g_s(R) and g_e(R) are our priors on local smoothness and global entropy of reflectance, and g_a(R) is a new absolute prior on each pixel. Since, illumination is a free parameter, a new prior on illumination h(L) is defined.

3 Local Reflectance Smoothness

The reflectance images of natural objects tend to be piecewise smooth, i.e. variation in reflectance images tends to be small and sparse. Variation in reflectnce affects both the colour and the brightness of the image.
The prior on reflectance smoothness is a multivariate Gaussian scale mixture (GSM) placed placed on the differences between each reflectance pixel and its neighbours. The following is the cost function to be minimized.

$$g_s(R) = \sum_i \sum_j log \left(\sum_{k=1}^{K} \alpha_k f(R_i - R_j); 0, \sigma_k \Sigma \right) \tag{3}$$

where N(i) is the 5X5 neighbourhood around pixel i, R_i -R_j is a 3-vector of the log-RGB differences from pixel i to pixel j, K= the number of discrete Gaussians in the GSM, α and σ are the coefficients, Σ is the covariance matrix of the entire GSM. This distribution prefers the nearby reflectance pixels to be similar, but allows rare discontinuities. For example, in the figure Fig. 2, it can be seen that strong colorful edges are costly and hence less likely but small edges caused by shading are more likely. But opposite is the case for influence. Since, sharp edges lie at the tail of the GSM, it has less influence and the small edges have higher influence. So during inference, the model explains the shading by varying shape while ignoring the sharp edges.



(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

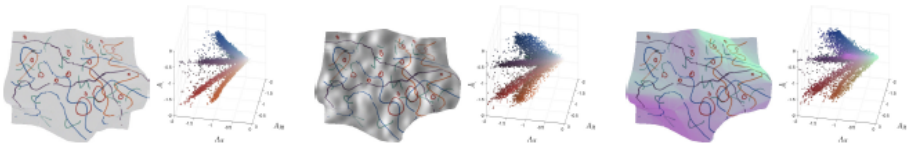
Figure 2: Visualization of the effect of the smoothness prior

4 Global Reflectnace Entropy

We assume that reflectance image of a single object is clustered in RGB space. Based on this assumption, we get the second term which is a measure of the global entropy. The entropy should be anisotropic to take care of the RGB channels. So, we first perform a whitening transformation from the training images and compute the isotropic entropy in this space giving us the anisotropy. The cost function is give below:

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

where W is the whitening transformation learned from training reflectance images, and σ determines the scale of the clusters. As seen in the Fig. 3, mistakes in estimating shape or illumination produce artifacts in the inferred reflectance, causing the the RGB distribution of the inferred reflectance to be smeared, and causing entropy to increase.



(a) Correct Everything $g_e(R) = 0.913$ (b) Wrong Shape $g_e(R) = 1.325$ (c) Wrong Light $g_e(R) = 2.366$

Figure 3: Reflectance images and their corresponding log-RGB scatterplots.

5 Absolute Colour

The assumption we make about colours is not all colours are equally likely. We must impose a prior on the absolute reflectance which is the log RGB value of each pixel. Our model is a 3D Thin-Plate spline fitted to the distribution of whitened log RGB pixels in our training set. To train the model we minimize the following:

.

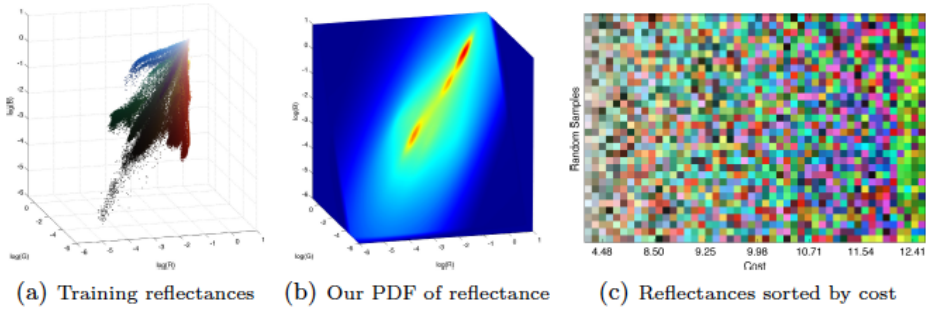
$$minimize_F \left(\sum_{i,j,k} F_{i,j,k} * N_{i,j,k} \right) + log \left(\sum_{i,j,k} \exp\left(-F_{i,j,k}\right) \right) + \lambda \sqrt{J(F) + \varepsilon^2} \tag{5}$$

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

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() is the TSP bending energy cost. Minimizing the first two terms is maximising the likelihood of the training data and the third term smoothens the TPS. During inference, we minimize the cost under the proposed model.

$$g_a(R) = \sum_i (WR_i) \tag{6}$$

Figure. 4, shows that the proposed model prefers subdued colours over bright colours.



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

Figure 4: The image on the left side is a scatterplot of the log RGB values of the training images. The image in the centre shows the proposed model that is fitted to the data. The image on the right has some randomly generated reflectances sorted by their costs.

6 Illumination

We optimise a sinlge model of illumination with shape. We also ave to define a prior on illumintion for this model. We define a spherical harmonic model of illumination [4]. Spherical harmonic model is an approximation to calculate the intensity of light at each pixel. This is appoximated with a series called spherical harmonics series. In this particular case, we use a 2^{nd}ordered SHA which has 9 dimensions per channel. So, in total we have 27 dimensional vector for L. To train our model, we fit a multivariate Gaussian to the spherical harmonic illuminations in our training set. The cost is given by the equation.

$$h(L) = \lambda_L (L - \mu_L)^T \Sigma_L^{-1} (L - \mu_L) \tag{7}$$

An example is shown in Fig. 5 which shows the training data and the samples from the learned models. The samples look superficially similar to the data suggesting the model is reasonable.

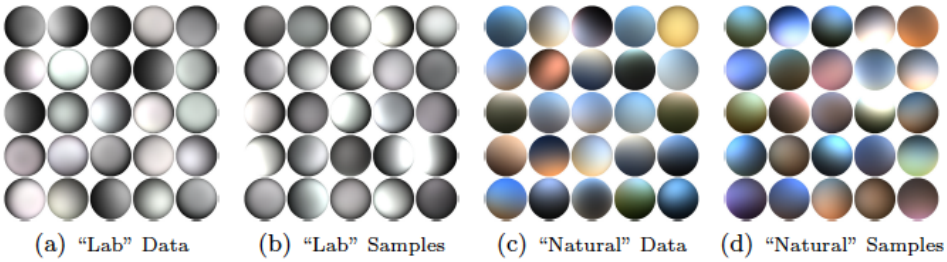


Figure 5: Training data and samples from the learned models.

7 Prior on shape

Our prior on shape is a linear combination of 3 terms [1].

$$f(Z) = \lambda_f f_f(Z) + \lambda_c f_c(Z) + \lambda_k f_k(Z) \tag{8}$$

where $f_f$is a flatness term, $f_c$encourages shapes to face outwards at the boundaries, and $f_k$is a smoothness term. λ terms are learned through cross-validation.

7.1 Flatness

We impose a prior that prefers the flattest shape in bas-relief family,[3] by minimizing the slant of Z at all points in Z. This is shown in the following equation:

$$f_f(Z) = -\sum_{x,y} log(2N^z_{x,y}(Z)) \tag{9}$$

where $N^z_{x,y}(Z)$ is the component of the surface normal of Z at position (x,y), which increases as slant decreases. This form of f_f(Z) is due to the assumption that if we observe a random surface in space, it is more likely that it is facing us ($N^z_{x,y}(Z) = 1$) than it is prependicular to us ($N^z_{x,y}(Z) = 0$).

7.2 Occluding Boundary

For every point i on the occluding boundary C, we estimate n_i, the local normal to the occluding boundary in the image plane. We try to make these normals as close to the surface normals on the depth map Z in terms of value. This is done by minimizing the loss function.

$$f_c(Z) = \sum_{i \epsilon C} \sqrt{(N^x_i(Z) - n^x_i)^2 + (N^y_i(Z) - n^y_i)^2} \tag{10}$$

7.3 Variation of mean curvature

We place a prior term to keep the shape of the surface smooth. We do this based on the variation of mean curvature of the surface. We try to minimize the variation by placing a penalty on

pairwise variation of mean curvature between all the pixels in a mall window. The prior used here is:

$$f_k(Z) = \sum_i \sum_{j \epsilon N(i)} log \left(\sum_{k=1}^{K} \alpha_k N(H(Z)_i - H(Z)_j; 0, \sigma_k) \right) \quad (11)$$

where N(i) is the 5x5 neighbourhood around the pixel i, H(Z) 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 (in this case the GSM has 40 discrete Gaussians), α and σ are the coefficients of the Gaussian. The GSM is learned using the training set.

8 Optimization

Straightforward gradient-based methods failed to work for this problem and coarse-to-fine grain techniques worked poorly. So, a new method is proposed for optimising f(X) where f is the loss function and X is the input data.

We define a function $L(X, h)$ which constructs a Laplacian pyramid from data X using a filter h, $L^{-1}(X, h)$ which reconstructs the data from the Laplacian pyramid, and $G(X, h)$ which constructs Gaussian pyramid from the data. Instead of minimizing f(X) we parametrize X as $Y = L(X, h)$ and then we minimize f'(Y). First step is to calculate the loss with respect to the Laplacian pyramid. Second step is to reconstruct the data from the pyramid. Third step is to compute the loss and the gradient with respect to the data and finally backpropagate the gardient onto the pyramid. The steps are given below mathematically.

$[l, \nabla_Y L] = f'(Y)$
$X \leftarrow L^{-1}(Y, h)$
$[l, \nabla_X L] \leftarrow f(X)$
$\nabla_Y l \leftarrow G(\nabla_X l, h)$

Finally, we solve for $X' = L^{-1}(argmin_Y f'(Y), h)$ using L-BFGS method. A binomial filter of length 5 was chosen and the filter which worked best was $[1, 4, 6, 4, 1]/(4\sqrt{2})$

9 Results

The algortihm is evaulated on 2 datasets, a MIT dataset with Laboratory like illumination and an own version of MIT dataset with natural illumination. 4 sets of illumination were conducted - and known and unknown illuminations for both laboratory illuminated naturally illuminated dataset. We calculate the errors with respect shape, illumination, shading, reflectance and the average error. The proposed model performs better for laboratory illumination and significantly better for naturally illuminated datasets. Some examples from the algorithm are shown in Fig. 7 and 8.

10 Conclusion

This paper presents 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. This is done 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, the proposed model outperforms all previously published algorithms

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	-	0.0852
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	-	0.0699
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

Figure 6: A comparison of the proposed model against previously proposed models

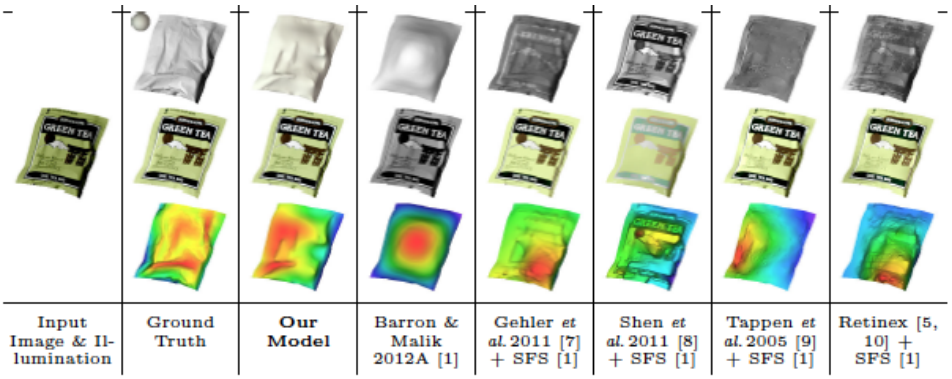


Figure 7: The output of the proposed model, and previous models for the Laboratory illuminated dataset

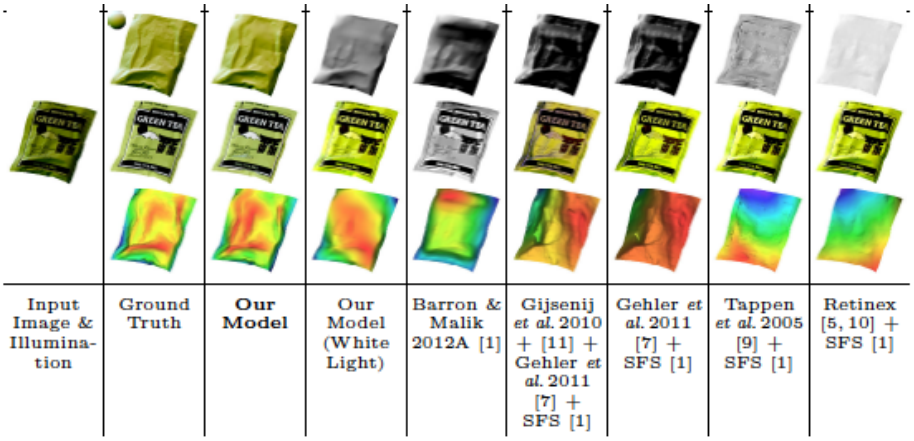


Figure 8: The output of the proposed model, and previous models for the naturally illuminated dataset

for intrinsic images and shape-from-shading, on both the MIT dataset and an own naturally illuminated variant of that dataset.

References

[1] Jonathan T. Barron and Jitendra Malik. Color constancy , intrinsic images , and shape estimation (supplementary material). 2012. 7

[2] Jonathan T Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *European Conference on Computer Vision*, pages 57–70. Springer, 2012. 1

[3] Jonathan T Barron and Jitendra Malik. Shape, albedo, and illumination from a single image of an unknown object. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 334–341. IEEE, 2012. 7.1

[4] B. Haefner, Z. Ye, M. Gao, T. Wu, Y. Quéau, and D. Cremers. Variational uncalibrated photometric stereo under general lighting. In *International Conference on Computer Vision (ICCV)*, Seoul, South Korea, October 2019. ¡a href="/$_m$*edia/spezial/bib/haefner*2019*supp.pdf*"*target* = "$_b$*lank*" > [*supp*] < /*a* > .6