# A Simple Model for Intrinsic Image Decomposition with Depth Cues

**Qifeng Chen**[1], **Vladlen Koltun**[1,2]
[1]Stanford University, [2]Adobe Research

## Abstract

We present a model for intrinsic decomposition of RGB-D images. Our approach analyzes a single RGB-D image and estimates albedo and shading fields that explain the input. To disambiguate the problem, our model estimates a number of components that jointly account for the reconstructed shading. By decomposing the shading field, we can build in assumptions about image formation that help distinguish reflectance variation from shading. These assumptions are expressed as simple nonlocal regularizers. We evaluate the model on real-world images and on a challenging synthetic dataset. The experimental results demonstrate that the presented approach outperforms prior models for intrinsic decomposition of RGB-D images.

## Introduction

The intrinsic image decomposition problemcalls for factorizing an input image into component images that separate the intrinsic material properties of depicted objects from illumination effects. The most common decomposition is into a reflectance image and a shading image. For every pixel, the reflectance image encodes the albedo of depicted surfaces, while the shading image encodes the incident illumination at corresponding points in the scene.

In this work, we consider this problem in light of the recent commoditization of cameras that acquire RGB-D images: simultaneous pairs of color and range images.

## Model

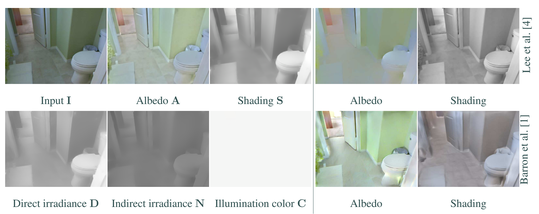


**Figure 1:** Left: the components produced by our model for an image from the NYU dataset. The top row shows the input image and the reconstructed albedo and shading images. The bottom row shows the constituent illumination components. Right: albedo and shading images produced by prior approaches.

Let $\mathbf{I}$ be the input RGB image. Our primary goal is to decompose $\mathbf{I}$ into an albedo image $\mathbf{A}$ and a shading image $\mathbf{S}$.

Our approach is based on the idea that the accuracy of this decomposition can be improved if we factorize the shading image into a number of components that can account for the different physical phenomena involved.

We factorize $\mathbf{I}$ into four component images: an albedo image $\mathbf{A}$, a direct irradiance image $\mathbf{D}$, an indirect irradiance image $\mathbf{N}$, and an illumination color image $\mathbf{C}$. These images are visualized in Figure 1. The albedo image $\mathbf{A}$ encodes the Lambertian reflectance of surfaces in the scene. The direct irradiance image $\mathbf{D}$ encodes the irradiance that each point in the scene would have received had there been no other objects that occlude or reflect the radiant flux emitted by the illuminants. The image $\mathbf{D}$ is thus intended to represent the direct irradiance that is modeled by local shading algorithms in computer graphics, which do not take shadows or inter-reflections into account. The indirect irradiance image $\mathbf{N}$ is the complement of $\mathbf{D}$, intended to absorb the contribution of shadows and indirect illumination.

For every pixel $p$, our factorization approximately satisfies

$$I_p = A_p D_p N_p C_p.$$

Taking logarithms on both sides yields

$$i_p = a_p + d_p + n_p + c_p.$$

We formulate the decomposition as an energy minimization problem:

$$\begin{aligned} &\underset{\mathbf{x}=(\mathbf{a},\mathbf{d},\mathbf{n},\mathbf{c})}{\arg\min} \quad E(\mathbf{x}) \\ &E(\mathbf{x}) = E_{\text{data}}(\mathbf{x}) + E_{\text{reg}}(\mathbf{x}). \end{aligned}$$

## Data Term

The data term is defined as a soft constraint:

$$E_{\text{data}} = \sum_p \|\text{lum}(I_p)(i_p - a_p - c_p - \mathbf{1}d_p - \mathbf{1}n_p)\|^2.$$

The objective on pixel $p$ is weighted by the luminance $\text{lum}(I_p)$ of $I_p$. Without this weight the data term would be disproportionately strong for dark pixels, since we operate in the logarithmic domain. (In the extreme, $I_p \to 0 \Rightarrow i_p \to -\infty$.) Weighting by the luminance of the input balances out the influence of the data term across the image.

## Regularization

**Albedo.** Our regularizer for the albedo component comprises weighted nonlocal pairwise terms that penalize albedo differences between pixels in the image:

$$E_{\mathbf{A}} = \sum_{\{p,q\} \in \mathcal{N}_{\mathbf{A}}} \alpha_{p,q} \|a_p - a_q\|^2.$$

$$\alpha_{p,q} = \left(1 - \frac{\|\text{ch}(I_p) - \text{ch}(I_q)\|}{\max\limits_{\{p,q\} \in \mathcal{N}_{\mathbf{A}}} \|\text{ch}(I_p) - \text{ch}(I_q)\|}\right) \sqrt{\text{lum}(I_p)\text{lum}(I_q)},$$

where $\text{ch}(I_p)$ denotes the chromaticity of $p$. We simply connect each pixel to $k$ random pixels in the image to form $\mathcal{N}_{\mathbf{A}}$. The left term expresses the well-established assumption that pixels that have similar chromaticity are likely to have similar albedo. The right term is the geometric mean of the luminance values of p and q and attenuates the strength of the regularizer for darker pixels. Note that there are no parameters in $\alpha_{p,q}$.

**Direct irradiance.** The direct irradiance regularizer models the spatial and angular coherence of direct illumination. Specifically, if two points in the scene have similar positions and similar normals, we expect them to have similar irradiance:

$$E_{\mathbf{D}} = \sum_{\{p,q\} \in \mathcal{N}_{\mathbf{D}}} (d_p - d_q)^2.$$

$\mathcal{N}_{\mathbf{D}}$ is constructed by connecting each pixel to its k nearest neighbors in $(x, y, z, n_x, n_y, n_z)$ feature space where $(x, y, z)$ is the 3D position of $p$ and $(n_x, n_y, n_z)$ is the surface normal at $p$. We apply a whitening transform to $(x, y, z)$ to normalize the feature values. (The other three dimensions are normalized by construction.)

**Indirect irradiance.** We assume that the indirect irradiance component is smooth in three-dimensional space:

$$E_{\mathbf{N}} = \sum_{\{p,q\} \in \mathcal{N}_{\mathbf{N}}} (n_p - n_q)^2.$$

$\mathcal{N}_{\mathbf{N}}$ is constructed by connecting each pixel to its $k$ nearest neighbors in $(x, y, z)$ feature space. We also include a simple $L^2$ regularizer on the indirect irradiance magnitude:

$$E_{\mathbf{N}'} = \sum_p n_p^2.$$

**Illumination color.** The irradiance can change quite significantly at relatively short distances but it is rare for the color of the incident illumination to vary as rapidly. We thus represent the illumination color separately, as a trichromatic field $\mathbf{C}$, so that a distinct regularizer can be applied:

$$E_{\mathbf{C}} = \sum_{\{p,q\} \in \mathcal{N}_{\mathbf{C}}} \gamma_{p,q} \|c_p - c_q\|^2.$$

The weight $\gamma_{p,q}$ adjusts the strength of the regularizer based on the Euclidean distance between the positions of $p$ and $q$ in $\mathbb{R}^3$, which we denote by $\bar{p}$ and $\bar{q}$:

$$\gamma_{p,q} = 1 - \frac{\|\bar{p} - \bar{q}\|}{\max\limits_{\{p,q\} \in \mathcal{N}_{\mathbf{C}}} \|\bar{p} - \bar{q}\|},$$

The set $\mathcal{N}_{\mathbf{C}}$ is constructed by connecting each pixel $p$ to $k$ other pixels $q$ in the image at random. Note that there are no parameters in $\gamma_{p,q}$.

## Experiments

(a) Input (b) Lee et al. [4] (c) Barron et al. [1] (d) Our approach

**Figure 2:** Intrinsic decomposition of an RGB-D image from the NYU Depth dataset [5]. (a) Input color and depth image. (b-d) Albedo and shading images estimated by two recent approaches for intrinsic decomposition of RGB-D images and by our approach.

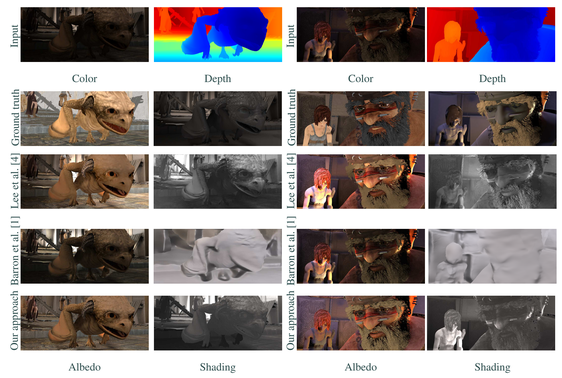


**Figure 3:** Results on two images from the MPI-Sintel dataset [2].

| | MSE albedo | MSE shading | MSE average | LMSE albedo | LMSE shading | LMSE average | DSSIM albedo | DSSIM shading | DSSIM average |
|---|---|---|---|---|---|---|---|---|---|
| Baseline 1 | 0.0531 | 0.0488 | 0.0510 | 0.0326 | 0.0284 | 0.0305 | 0.214 | 0.206 | 0.210 |
| Baseline 2 | 0.0369 | 0.0378 | 0.0373 | 0.0240 | 0.0303 | 0.0272 | 0.228 | 0.187 | 0.207 |
| Retinex [3] | 0.0606 | 0.0727 | 0.0667 | 0.0366 | 0.0419 | 0.0392 | 0.227 | 0.240 | 0.234 |
| Lee et al. [4] | 0.0463 | 0.0507 | 0.0485 | 0.0224 | **0.0192** | 0.0208 | **0.199** | 0.177 | 0.188 |
| Barron et al. [1] | 0.0452 | 0.0420 | 0.0436 | 0.0298 | 0.0264 | 0.0281 | 0.210 | 0.206 | 0.208 |
| Our approach | **0.0307** | **0.0277** | **0.0292** | **0.0185** | **0.0190** | **0.0188** | **0.196** | **0.165** | **0.181** |

**Table 1:** Quantitative evaluation of the albedo and shading images produced by different approaches on the MPI-Sintel dataset.

For quantitative evaluation, we used the MPI-Sintel dataset. This is a set of complex computer-generated images that were found to have similar statistics to natural images [2]. We used three error measures for evaluation: the standard mean-squared error (MSE), the local mean-squared error (LMSE) and the structural similarity index (SSIM). Since MSE and LMSE are dissimilarity measures, we report the DSSIM for consistency, defined as (1-SSIM)/2.

## Discussion

We view the presented work as a step towards high-fidelity estimation of reflectance properties and scene illumination from single RGB-D images. We believe that the key to progress lies in increasingly careful and detailed modeling and simulation of image formation. We hope that the simplicity of our model will encourage subsequent work on this problem.

## References

[1] Jonathan T. Barron and Jitendra Malik. Intrinsic scene properties from a single RGB-D image. In *CVPR*, 2013.

[2] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 2012.

[3] Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009.

[4] Kyong Joon Lee, Qi Zhao, Xin Tong, Minmin Gong, Shahram Izadi, Sang Uk Lee, Ping Tan, and Stephen Lin. Estimation of intrinsic image sequences from image+depth video. In *ECCV*, 2012.

[5] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In *ECCV*, 2012.