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.

1. Introduction

The intrinsic image decomposition problem calls for factorizing an input image into component images that separate the intrinsic material properties of depicted objects from illumination effects [6]. 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.

Intrinsic image decomposition has been studied extensively, in part due to its potential utility for applications in computer vision and computer graphics. Many computer vision algorithms, such as segmentation, recognition, and motion estimation are confounded by illumination effects in the image. The performance of these algorithms may benefit substantially from reliable estimation of illumination-invariant material properties for all objects in the scene. Furthermore, advanced image manipulation applications such as editing the scene's lighting, editing the material properties of depicted objects, and integrating new objects into photographs would all benefit from the ability to decompose an image into material properties and illumination effects.

Despite the practical relevance of the problem, progress on intrinsic decomposition of single images has been limited. Until recently, the state of the art was set by algorithms based on the classical Retinex model of image formation, which was developed in the context of flat painted canvases and is known to break down in the presence of occlusions, shadows, and other phenomena commonly encountered in real-world scenes [17]. Part of the difficulty is that the problem is ill-posed: a single input image can be explained by a continuum of reflectance and illumination combinations. Researchers have thus turned to additional sources of input that can help disambiguate the problem, such as using a sequence of images taken from a fixed viewpoint [34, 24, 23], using manual annotation to guide the decomposition [10, 27], and using collections of images [22, 32, 19]. While the use of temporal sampling, human assistance, and image collections has been shown to help, the problem of automatic intrinsic decomposition of a single image remains difficult and unsolved.

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. RGB-D imaging sensors are now widespread, with tens of millions shipped since initial commercial deployment and new generations being developed for integration into mobile devices. While the availability of depth cues makes intrinsic image decomposition more tractable, the problem is by no means trivial, as demonstrated by the performance of existing approaches to intrinsic decomposition of RGB-D images (Figure 1).

Our approach is based on a simple linear least squares formulation of the problem. We decompose the shading component into a number of constituent components that account for different aspects of image formation. Specifically, the shading image is decomposed into a direct irradiance component, an indirect irradiance component, and a color component. These components are described in detail in Section 3. We take advantage of well-known smoothness properties of direct and indirect irradiance and design simple nonlocal regularizers that model these properties. These regularizers alleviate the ambiguity of the decomposition by

(a) Input (b) Lee et al. [21] (c) Barron et al. [5] (d) Our approach

Figure 1. Intrinsic decomposition of an RGB-D image from the NYU Depth dataset [29]. (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.

encoding specific assumptions about image formation and substantially improve the fidelity of estimated reflectance and shading.

We evaluate the presented model on real-world images from the NYU Depth dataset [29] and on synthetic images from the MPI-Sintel dataset [11]. The presented model outperforms prior models for intrinsic decomposition of RGB-D images both qualitatively and quantitatively.

2. Background

The problem of estimating the intrinsic reflectance of objects depicted in an image was studied by Land and McCann [20], whose Retinex model formed the basis for subsequent work on the problem. The Retinex model captures image formation for Mondrian images: images of a planar canvas that is covered by patches of constant reflectance and illuminated by multiple light sources. In such images, strong luminance gradients can be assumed to correspond to reflectance boundaries. Based on this assumption, Land and McCann described an algorithm that can compute the relative reflectance of two points in an image by integrating strong luminance gradients along a path that connects the points. The algorithm was extended to two-dimensional images by Horn [18], who observed that a complete decomposition of an image into reflectance and shading fields can be obtained by zeroing out high Laplacians in the input and solving the corresponding Poisson equation to obtain the shading field. This approach was further extended by Blake [9], who advocated for operating on gradients instead of Laplacians and by Funt et al. [14], who applied the approach to color images by analyzing chromaticity gradients. Related ideas were developed for the removal of shadows from images [13, 12].

The Retinex model is based on a heuristic classification of image derivatives into derivatives caused by changes in reflectance and derivatives caused by shading. Subsequent work proposed the use of statistical analysis to train classifiers for this purpose [8, 31]. Alternatively, a regression function can be trained for finer-grained estimation of shading and albedo derivatives [30]. Researchers have also augmented the basic Retinex model with nonlocal texture cues [36] and global sparsity priors [28, 16]. Sophisticated techniques that recover reflectance and shading along with a shape estimate have been developed [2, 4, 3]. While these developments have advanced the state of the art, the intrinsic image decomposition problem remains severely underconstrained and the performance of existing algorithms on complex real-world images remains limited.

The commoditization of RGB-D imaging sensors provides an opportunity to re-examine the intrinsic image decomposition problem and a chance to obtain highly accurate decompositions of complex scenes without human assistance. Two recent works have explored this direction. The first is due to Lee et al. [21], who developed a model for intrinsic decomposition of RGB-D video. Their model builds on Retinex with nonlocal constraints [36], augmented by constraints that regularize shading estimates based on normals obtained from the range data, as well as temporal constraints that improve the handling of view-dependent effects. This approach can also be applied to single RGB-D images: the temporal constraints simply play no role in this case. Our approach is likewise based on nonlocal constraints, but the constraints in our formulation are soft, which provides increased robustness to image noise and to violations of modeling assumptions. Our formulation is also based on a more detailed analysis of image formation,

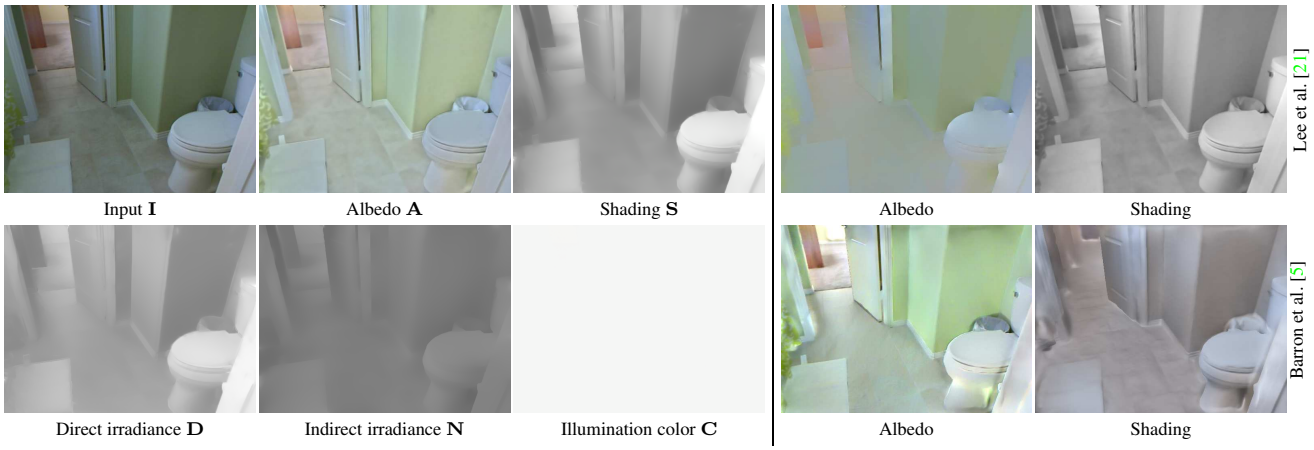

Figure 2. Left: the components produced by our model for an image from the NYU dataset. The top row shows the input image and the reconstructed albedo and shading images. The bottom row shows the constituent illumination components. Right: albedo and shading images produced by prior approaches.

which leads to improved discrimination between reflectance and illumination effects.

The second recent work on intrinsic decomposition of RGB-D images is due to Barron and Malik [5], who use non-convex optimization to obtain a smoothed depth map and a spatially varying illumination model. We observe that improved decomposition into reflectance and shading can be obtained without joint optimization of the provided depth image. While the depth images produced by existing commodity sensors are noisy, they can be smoothed by off-the-shelf algorithms. We found such a priori smoothing to be sufficient, in part because our formulation is designed to be resilient to noisy input. Since we do not attempt to solve the reflectance and shading decomposition problem while also optimizing the underlying scene geometry, we can formulate a much simpler convex objective that can be reliably optimized.

We also refer the reader to the recent work of Yu et al. [35] that uses RGB-D data to disambiguate the related problem of shape-from-shading.

3. Model

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 $\mathbf{S}$. For every pixel p, the decomposition should approximately satisfy the equivalence $I_p = A_p S_p$, where the product $A_p S_p$ is performed separately in each color channel. 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. The advantage of this approach is that each component can be regularized differently. By considering the smoothness properties of each factor in the scene's illumination, we can design simple regularizers based on our understanding of image formation.

Specifically, 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 2. 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.

The factorization of irradiance into a direct component and an indirect component is one of the features that distinguish our model from prior work on intrinsic image decomposition. One of the pitfalls in intrinsic image decomposition is the absorption of genuine albedo variation in the shading images. A common approach to dealing with this is to restrict the problem by reducing its dimensionality. We take a different approach and deliberately increase the dimensionality of the problem by further decomposing the shading image to distinguish between direct and indirect irradiance. Our guiding observation is that these components have different smoothness characteristics. Direct irradiance varies slowly as a function of position and surface orientation [25, 7]. Indirect irradiance can have higher frequencies, but is spatially smooth almost everywhere [1, 26]. We employ dedicated regularizers that model these characteristics. The finer-grained decomposition of the shading image allows us to regularize it more carefully and thus reduce the leakage of albedo variation into the shading image and vice versa.

For every pixel p, our factorization approximately satisfies

$$I_p = A_p D_p N_p C_p. \quad (1)$$

As is common in intrinsic image decomposition, we operate in the logarithmic domain. 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, with a data term and a regularization term:

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

These terms are described in detail in Sections 3.1 and 3.2.

3.1. Data Term

The data term is defined as

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

The objective on pixel p is weighted by the luminance $\text{lum}(I_p)$ of I_p. (In practice, we use $\text{lum}(I_p)+\varepsilon$ to avoid zeroing out the data term.) 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.

The traditional approach in intrinsic image decomposition is to reduce the dimensionality of the problem by representing one of the components strictly in terms of the others. For example, it is common to solve for the shading $\mathbf{S}$ and then to simply obtain the albedo by taking $A_p = I_p/S_p$ for every pixel (or vice versa) [17, 16, 36, 21, 5]. In our formulation, this would mean omitting the variable a_p from the optimization and substituting $i_p - c_p - \mathbf{1}d_p - \mathbf{1}n_p$ in its place. In other words, the decomposition assumption expressed by the data term (2) is traditionally a hard constraint. In practice, however, this assumption clearly does not always hold. Participating media, blur, chromatic distortion, and sensor noise all invalidate the assumption that $I_p = A_p S_p$. For this reason, our model expresses this assumption as a soft constraint: the data term. Experimentally, the benefits of this formulation seem to clearly outweigh the costs of somewhat increased dimensionality. In particular, this model is considerably more stable in dealing with very dark input pixels, whose chromaticity can be drastically perturbed by sensor noise.

3.2. Regularization

The regularization objective comprises separate terms for regularizing the albedo, the direct irradiance, the indirect irradiance, and the illumination color:

$$E_{\text{reg}} = \sum_{i \in \{\mathbf{A},\mathbf{D},\mathbf{N},\mathbf{N}',\mathbf{C}\}} \lambda_i E_i. \quad (3)$$

We now describe each of these terms.

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

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

The weight $\alpha_{p,q}$ adjusts the strength of the regularizer based on the chromaticity difference between p and q, and the luminance of p and q:

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

where $\text{ch}(I_p)$ denotes the chromaticity of p. The left term expresses the well-established assumption that pixels that have similar chromaticity are likely to have similar albedo [14, 12, 36, 15, 21]. 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, for which the chromaticity is ill-conditioned.

The somewhat unorthodox aspect of the regularizer is the construction of the set of pairs $\mathcal{N}_{\mathbf{A}}$ on which the regularizer operates. Given our prior belief that pixels with similar chromaticity are likely to have similar albedo, it would make sense to identify such pairs and preferentially connect them. In practice, such preferential connectivity strategies are highly liable to create largely disconnected clusters in the image with very poor communication between them. When this happens, the association of a pixel with a cluster is largely exclusive and is determined by its chromaticity. This again places too much confidence in chromaticity, which can be poorly conditioned. Instead, we simply connect each pixel to k random pixels in the image. The random connectivity strategy leads to reasonably short graph distances between pixels, while not treating input chromaticity as a hard constraint. Here too the intuition is that our assumptions on image formation have limited validity in practice. In particular, while input chromaticity is correlated with the intrinsic reflectance of the imaged surface, it is also affected by camera optics and other aspects of image formation that we do not model. Thus instead of committing to a connectivity strategy that would act as a hard constraint, we express our modeling assumptions through the weight $\alpha_{p,q}$. Note that this weight has no free parameters that need to be tuned.

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 if the contribution of other objects in the scene (in the form of shadows and inter-reflections) is not taken into account [25, 7]. (Note again that the direct irradiance component is meant to represent the "virtual" irradiance that every point in the scene would have received had the scene contained only the light sources and no other objects that cast shadows or reflect light.) The regularizer has the following form:

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

The set $\mathcal{N}_{\mathbf{D}}$ of pairwise connections is constructed as follows. For each pixel p we compute a feature vector (x, y, z, n_x, n_y, n_z). The vector (x, y, z) is the position of p in three-dimensional space, which can be easily computed from the image coordinates of p and the corresponding depth value. The vector (n_x, n_y, n_z) is the surface normal at p, computed from the depth values at p and nearby points. We thus embed all input pixels in a six-dimensional feature space. To normalize the feature values, we apply a whitening transform to the (x,y,z) dimensions. (The other three dimensions are normalized by construction.) Then, for each pixel p, we find its k nearest neighbors in this feature space. (We use $k = 5$, 10, or 20 for all regularizers in this paper.) For each such neighbor q, we add the pair $\{p, q\}$ to the set $\mathcal{N}_{\mathbf{D}}$.

This strategy connects each pixel p to k other pixels in the image that have similar spatial location and surface normal. This connectivity strategy is more confident than the one we used for albedo regularization. This is a key advantage of separating the direct and indirect irradiance components. When occlusion effects are separated out, irradiance becomes a simpler function that varies smoothly with position and surface normal. The simple approach of connecting nearest neighbors in the relevant feature space is thus sufficient.

Indirect irradiance. We assume that the indirect irradiance component is smooth in three-dimensional space. While irradiance is clearly not smooth in image space due to occlusion, it is smooth almost everywhere in object space [1, 26]. Our regularizer is a direct expression of this assumption:

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

To construct the set $\mathcal{N}_{\mathbf{N}}$ of pairwise connections, we simply connect each pixel p to its k nearest neighbors in $\mathbb{R}^3$, based on its location in three-dimensional 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 direct and indirect irradiance components $\mathbf{D}$ and $\mathbf{N}$ are modeled as scalar fields. In actuality, illumination can have nontrivial chromaticity. It would have thus been natural to model the direct irradiance, for example, as trichromatic. In our experiments, this choice led to diminished decomposition performance. The reason is that the irradiance can change quite significantly at relatively short distances when surface curvature is high. On the other hand, it is less common for the color of the incident illumination to vary as rapidly. Representing the total irradiance and its spectral power distribution jointly as a single trichromatic field would mean that the regularizer cannot easily distinguish these terms. In practice, this leads to unnatural swings in illumination color.

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. That is, we use random connectivity for regularizing the illumination color component, akin to the albedo regularization. The reason is that using nearest neighbor connectivity in 3-space can split the pixels into multiple disconnected clusters, since occlusion boundaries in image space correspond to jump discontinuities in 3-space. Since the factorization into surface albedo and illumination color is ill-defined, the inferred illumination color can vary sharply from cluster to cluster. Using random connectivity instead leads to a globally connected graph in which all pixels communicate and the computed illumination color varies smoothly across the scene.

4. Experiments

Our approach is implemented in Matlab and uses the lsqlin function to optimize the linear least-squares objective. (The log-albedo and log-color components are constrained to be ≤ 0 for each pixel and each color channel. This encodes the constraint that the value of each color

Figure 3. Results on two images from the NYU dataset. For each image (top and bottom), the results are organized as in Figure 2.

channel in $\mathbf{A}$ and $\mathbf{C}$ has to be between 0 and 1.) Running times were measured on a laptop with an Intel Core i7-3610QM 2.3 GHz CPU and 16GB of RAM.

NYU dataset. We evaluated the presented model on the 16 images used by Barron and Malik in the main body and the supplementary material of their paper [5]. Results on one of these images are shown in Figure 1. Results on the other fifteen images are provided in supplementary material. For these images, the albedo and shading images shown for the approach of Barron and Malik are taken directly from their paper. The results for the approach of Lee et al. [21] are computed using the implementation provided by the authors.

We also tested the different approaches on additional randomly sampled images from the dataset. Three such images are shown in Figures 2 and 3. Results for the prior approaches were computed using code provided by the authors. The average running times were 12 minutes for our approach, 3 seconds for the approach of Lee et al., and 2 hours for the approach of Barron and Malik. (The maximal number of L-BFGS iterations in the implementation of Barron and Malik was set to 500.)

We used $\lambda_{\mathbf{A}} = \lambda_{\mathbf{N}'} = 0.1$ and $\lambda_{\mathbf{D}} = \lambda_{\mathbf{N}} = \lambda_{\mathbf{C}} = 1$.

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 [11]. We used the "clean pass" images as input. (Infinite depth of field, no motion blur, and no atmospheric effects like fog.) This dataset was not intended for evaluation of intrinsic image algorithms, but we use it for lack of a readily apparent alternative that would reproduce many of the challenges of real-world scenes, such as complex object shapes, occlusion, and complex lighting, and would be accompanied by the requisite ground truth data. We are grateful to the creators of the dataset for providing us with the depth maps and for improving the accuracy of the ground-truth albedo maps. They also created the ground-truth shading images by rendering all the scenes with uniform grey albedo on all objects.

A number of scenes from the dataset could not be used due to software issues that resulted in defects in the provided ground-truth albedo maps. In total, we used 15 scenes. A list is provided in supplementary material. We pruned the set of images automatically by taking every fifth image from each scene. This yielded a set of 141 images. For each of these input images, we obtained albedo and shading images using our approach, the approach of Lee et al. [21], the approach of Barron and Malik [5], the color

	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 [17]	0.0606	0.0727	0.0667	0.0366	0.0419	0.0392	0.227	0.240	0.234
Lee et al. [21]	0.0463	0.0507	0.0485	0.0224	**0.0192**	0.0208	**0.199**	0.177	0.188
Barron et al. [5]	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.

Retinex algorithm (using the implementation of [17]), and two baselines. The first baseline used the input image as the albedo image and a uniform grey image as the shading image. The second baseline did the opposite, using the input as shading and a constant image as albedo.

Table 1 provides a quantitative evaluation of the results obtained by the different approaches. We used three error measures for evaluation. Following Grosse et al. [17], we use scale-invariant measures, such that the absolute brightness of each image is adjusted to minimize the error. The first error measure is the standard mean-squared error (MSE). The second is the local mean-squared error (LMSE), introduced by Grosse et al. [17]. Specifically, we cover the image by overlapping windows of size 10% of the image in every dimension, adjust the brightness separately for each window to fit the corresponding part of the ground truth image, compute the MSE for each window, and average the results. This is a finer-grained measure, but it still suffers from many of the defects of MSE. For this reason, we also use the structural similarity index (SSIM), developed specifically to provide a better image similarity measure [33]. Since the SSIM is a similarity measure (i.e., higher is better), while MSE and LMSE are dissimilarity measures (i.e., lower is better), we report the DSSIM for consistency, defined as (1-SSIM)/2.

The average running times were 15 minutes for our approach, 8 seconds for the approach of Lee et al., and 3 hours for the approach of Barron and Malik. Weights for our approach and for the approach of Lee et al. were set using randomized two-fold cross-validation. The approaches were trained on the SSIM measure. We did not train separately for MSE and LMSE. Due to the running time and the number of parameters for the approach of Barron and Malik, we did not perform cross-validation for this approach. We tried to adjust key parameters for this approach manually to maximize performance.

5. 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 problem is solvable (in a practically interesting sense), but is far from solved. Our results are still far from the ground truth and our model does not attempt to explicitly account for specular reflectance, translucency, participating media, camera optics, and other factors. 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. All code will be made freely available.

References

[1] J. Arvo. The irradiance Jacobian for partially occluded polyhedral sources. In *SIGGRAPH*, 1994. 3, 5

[2] J. T. Barron and J. Malik. High-frequency shape and albedo from shading using natural image statistics. In *CVPR*, 2011. 2

[3] J. T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *ECCV*, 2012. 2

[4] J. T. Barron and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*, 2012. 2

[5] J. T. Barron and J. Malik. Intrinsic scene properties from a single RGB-D image. In *CVPR*, 2013. 2, 3, 4, 6, 7, 8

[6] H. G. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. In *Computer Vision Systems*. 1978. 1

[7] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2), 2003. 3, 5

[8] M. Bell and W. T. Freeman. Learning local evidence for shading and reflectance. In *ICCV*, 2001. 2

[9] A. Blake. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics, and Image Processing*, 32(3), 1985. 2

[10] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. *ACM Trans. Graph.*, 28(5), 2009. 1

[11] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 2012. 2, 6

[12] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *IJCV*, 85(1), 2009. 2, 4

[13] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew. On the removal of shadows from images. *PAMI*, 28(1), 2006. 2

[14] B. V. Funt, M. S. Drew, and M. Brockington. Recovering shading from color images. In *ECCV*, 1992. 2, 4

[15] E. Garces, A. Muñoz, J. Lopez-Moreno, and D. Gutierrez. Intrinsic images by clustering. *Comput. Graph. Forum*, 31(4), 2012. 4

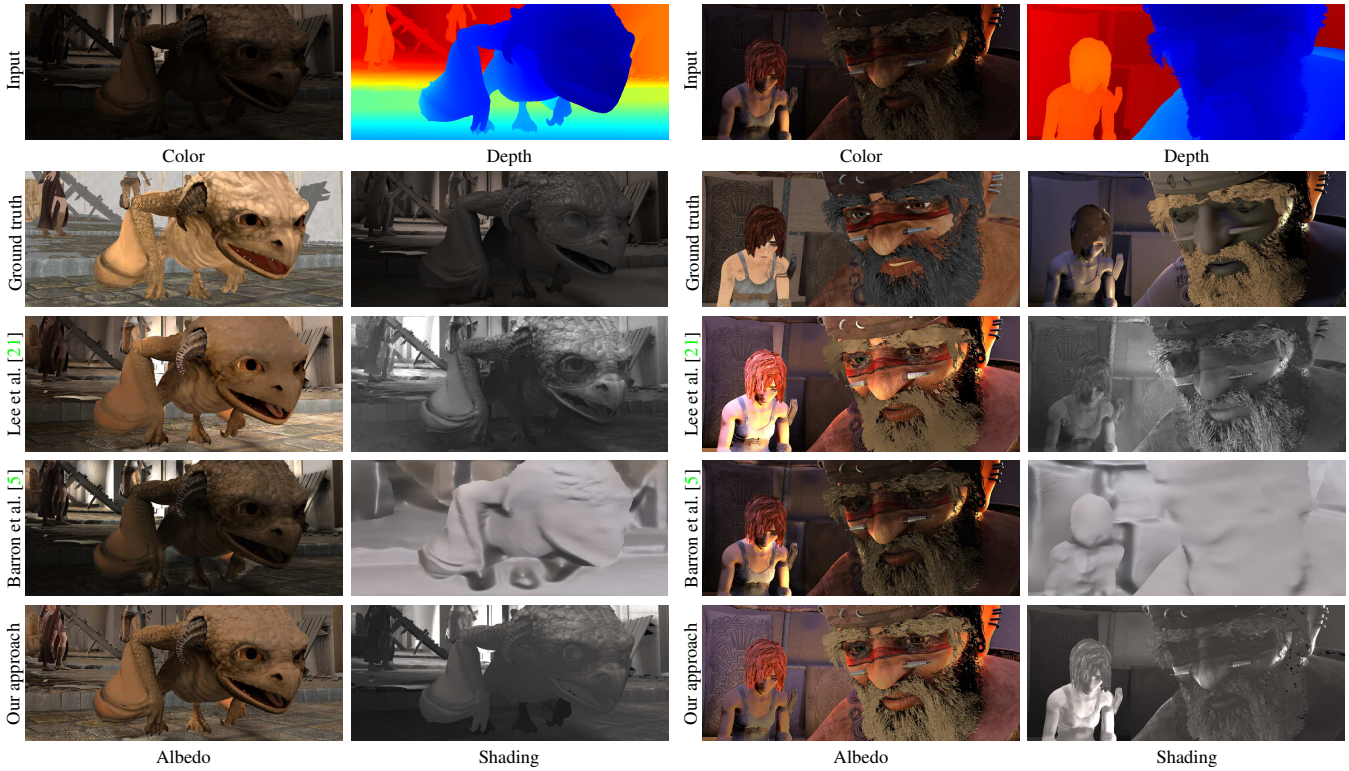

Figure 4. Results on two images from the MPI-Sintel dataset.

[16] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *NIPS*, 2011. 2, 4

[17] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009. 1, 4, 6, 7

[18] B. K. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4), 1974. 2

[19] P.-Y. Laffont, A. Bousseau, and G. Drettakis. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Trans. Vis. Comput. Graph.*, 19(2), 2013. 1

[20] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1), 1971. 2

[21] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *ECCV*, 2012. 2, 3, 4, 6, 7, 8

[22] X. Liu, L. Wan, Y. Qu, T.-T. Wong, S. Lin, C.-S. Leung, and P.-A. Heng. Intrinsic colorization. *ACM Trans. Graph.*, 27(5), 2008. 1

[23] Y. Matsushita, S. Lin, S. B. Kang, and H.-Y. Shum. Estimating intrinsic images from image sequences with biased illumination. In *ECCV*, 2004. 1

[24] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *PAMI*, 26(10), 2004. 1

[25] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH*, 2001. 3, 5

[26] R. Ramamoorthi, D. Mahajan, and P. N. Belhumeur. A first-order analysis of lighting, shading, and shadows. *ACM Trans. Graph.*, 26(1), 2007. 3, 5

[27] J. Shen, X. Yang, Y. Jia, and X. Li. Intrinsic images using optimization. In *CVPR*, 2011. 1

[28] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *CVPR*, 2011. 2

[29] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In *ECCV*, 2012. 2

[30] M. F. Tappen, E. H. Adelson, and W. T. Freeman. Estimating intrinsic component images using non-linear regression. In *CVPR*, 2006. 2

[31] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *PAMI*, 27(9), 2005. 2

[32] A. Troccoli and P. K. Allen. Building illumination coherent 3D models of large-scale outdoor scenes. *IJCV*, 78(2-3), 2008. 1

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 2004. 7

[34] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, 2001. 1

[35] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In *CVPR*, 2013. 3

[36] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin. A closed-form solution to retinex with nonlocal texture constraints. *PAMI*, 34(7), 2012. 2, 4