Intrinsic Image Evaluation on Synthetic Complex Scenes

S. Beigpour*, M. Serra†, J. van de Weijer†, R. Benavente†, M. Vanrell†, O. Penacchio†, D. Samaras◇

*shidab@hig.no The Norwegian Colour and Visual Computing Laboratory, Gjøvik University College, Norway

† {mserra, joost, robert, maria, penacchio}@cvc.uab.cat Computer Vision Center, Universitat Autònoma de Barcelona, Spain

◇samaras@cs.sunysb.edu Image Analysis Lab, Computer Science Department, Stony Brook University, NY, USA

Abstract

Scene decomposition into its illuminant, shading, and reflectance intrinsic images is an essential step for scene understanding. Collecting intrinsic image groundtruth data is a laborious task. The assumptions on which the ground-truth procedures are based limit their application to simple scenes with a single object taken in the absence of indirect lighting and interreflections. We investigate synthetic data for intrinsic image research since the extraction of ground truth is straightforward, and it allows for scenes in more realistic situations (e.g, multiple illuminants and interreflections). With this dataset we aim to motivate researchers to further explore intrinsic image decomposition in complex scenes.

Motivation

The reflection model [4] compiles the color observation f^c with $c \in \{R, G, B\}$ as:

$$f^c(\mathbf{x}) = m(\mathbf{x}) \int_\omega s(\lambda, \mathbf{x})\, e(\lambda, \mathbf{x}) \rho^c(\lambda)\, d\lambda. \qquad (1)$$

- The state-of-the-art, MIT dataset [5], only considers "white" illuminant i.e. $e(\lambda, \mathbf{x}) = 1$.
- The traditional way of collecting data for intrinsic image study is time-consuming, and restrictive.
- Our synthetic image dataset shows how scenes can be rendered with photo realistic quality.
- Our dataset allows for collecting precise **pixel-wise** ground truth of **Multi-illuminant** scenes.
- We model the light rays bouncing off objects (e.g., color cast, inter-reflection) as in Real world scenes.

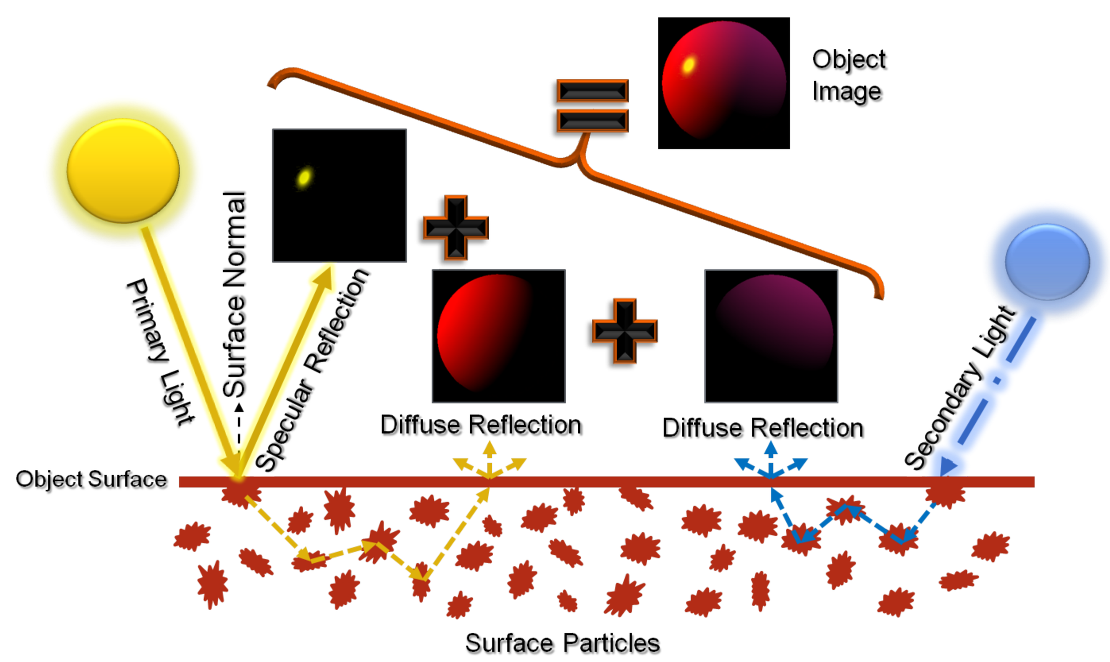
Example multi-Illuminant scenes: The 2nd and 4th are real-world photos, while the 1st and 3rd are computer generated.

To prevent propagating the full multispectral data, which is computationally very expensive, rendering engines approximate Eq1 with:

$$\hat{f}^c = \int_\omega s(\lambda) \rho^c(\lambda)\, d\lambda \int_\omega e(\lambda) \rho^c(\lambda)\, d\lambda = \mathbf{s} \circ \mathbf{e}. \qquad (2)$$

Multi-illuminant & Photo-realism

Schematic of a multi-Illuminant scene.



Here's a comparison between direct illumination and global illumination (left and right respectively) shows the importance of modeling the light energy bouncing back from objects in the scene (e.g., color cast and inter-reflection)

Here's the Eq2 when accounting for the bounce of light ray.

$$\hat{\mathbf{f}} = \mathbf{s}^1 \circ \mathbf{e} + \mathbf{s}^2 \circ \mathbf{s}^1 \circ \mathbf{e}, \qquad (3)$$

The reconstruction error:

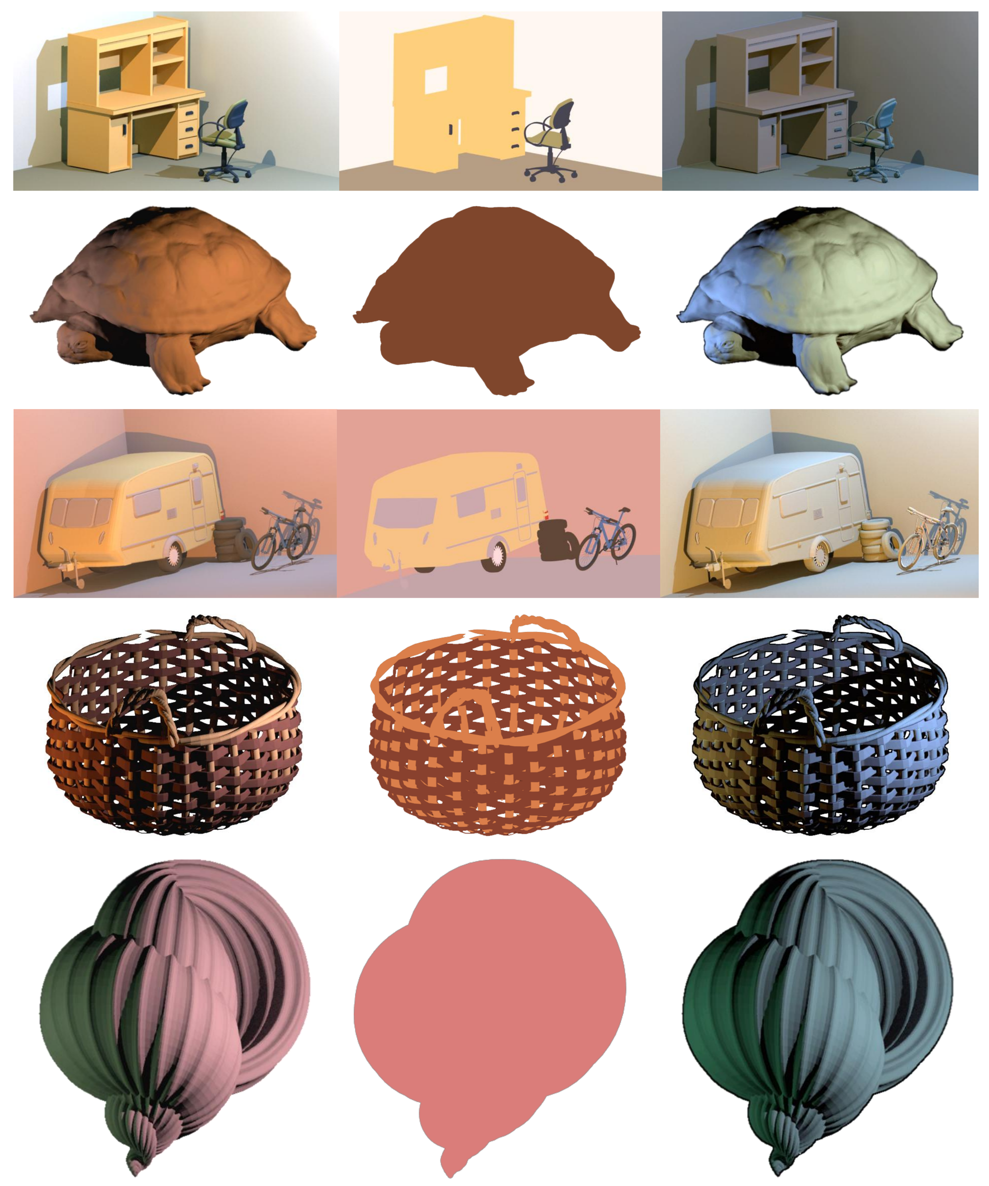
$$\varepsilon = \frac{\|\mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\mathbf{x})\|}{\|\mathbf{f}(\mathbf{x})\|} \qquad (4)$$

	One bounce		Two bounces	
sensors	Mean (%)	Max (%)	Mean (%)	Max (%)
3	0.58	2.88	1.38	23.84
6	0.19	1.25	0.55	9.06
9	0.12	0.86	0.34	3.77

Reconstruction error for single and two bounce reflection for 3, 6, and 9 sensors..

Ground Truth

Examples from our dataset:



From left to right: Rendered scene, Reflectance groundtruth, and Illumination groundtruth.

References

[1] J.T. Barron and J. Malik, *Color constancy, intrinsic images, and shape estimation*, in ECCV 2012

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

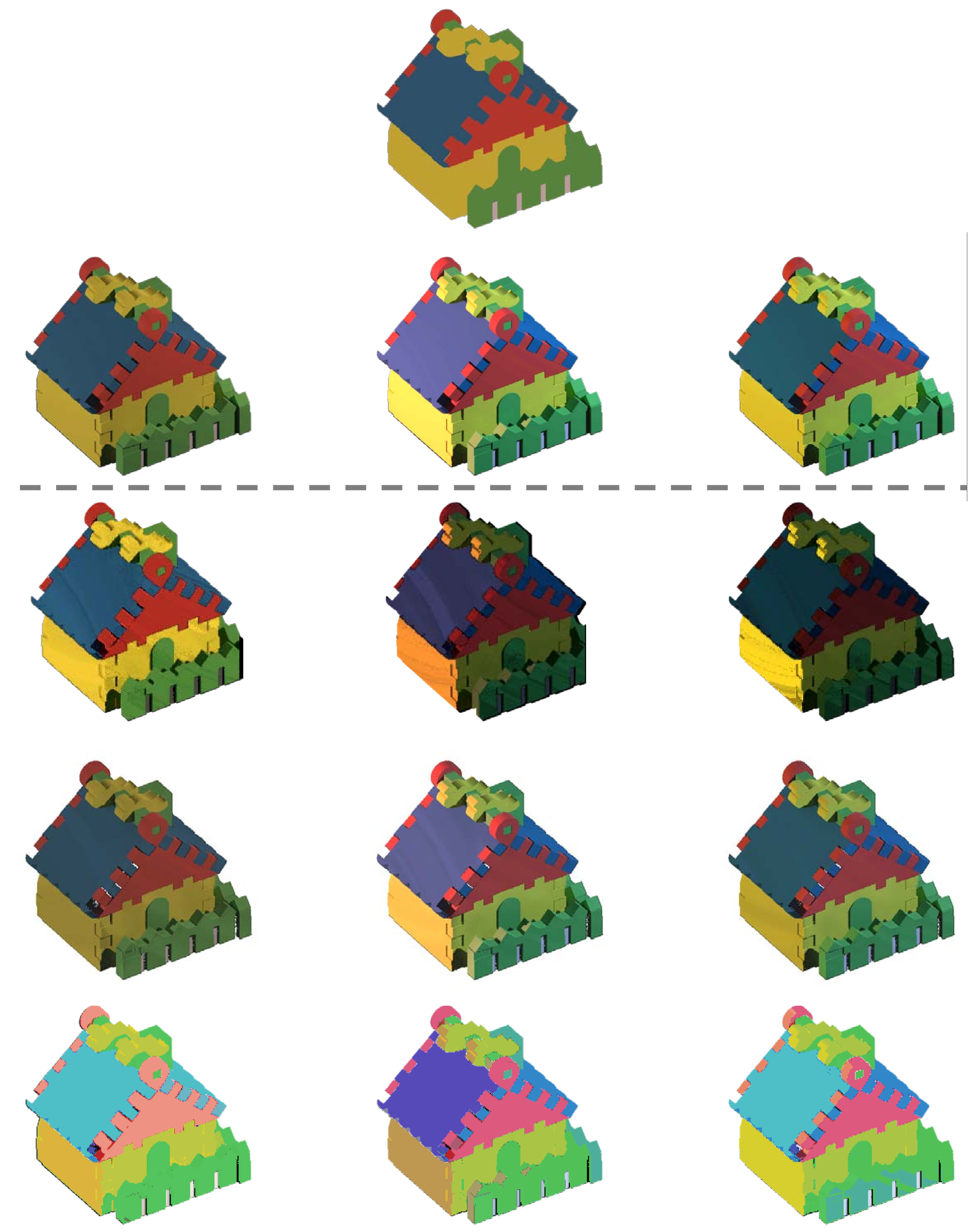
[3] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell, *Names and shades of color for intrinsic image estimation*, in CVPR 2012

[4] S.A. Shafer, *Using color to separate reflection components*, Color Research and Application, 1985

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

Qualitative Results

An example of reflectance recovery algorithm performance on an image from our dataset:



The rows from top to bottom: Groundtruth reflectance, generated image under 3 different lights, and reflectance recovery results by the 3 algorithms [1],[2],[3] (respectively). From left to right: White (WL), Single (1L) and Multiple (2L) illuminants.

Project Page

For more info and to download our dataset:
http://www.cic.uab.cat/Datasets/synthetic_intrinsic_image_dataset/

Quantitative Results

	Reflectance						Shading					
	Single Objects			Complex scenes			Single Objects			Complex scenes		
Method	WL	1L	2L	WL	1L	2L	WL	1L	2L	WL	1L	2L
Barron & Malik [1]	0.082	0.099	0.102	0.020	0.059	0.039	0.043	0.046	0.054	0.011	0.014	0.014
Gehler *et al.* [2]	0.089	0.113	0.123	0.018	0.067	0.040	0.043	0.045	0.051	0.007	0.009	0.009
Serra *et al.* [3]	0.063	0.069	0.076	0.027	0.041	0.033	0.021	0.022	0.025	0.006	0.006	0.007

LMSE results of three intrinsic image methods on our dataset. For clarity, errors for reflectance and shading are given separately. For both single objects and complex scenes, results for white illumination (WL), one illuminant (1L), and two illuminants (2L) are averaged.