OSA Publishing

CVC Centre de Visió per Computador

UAB Universitat Autònoma de Barcelona

Deep Intrinsic decomposition trained on surreal scenes yet with realistic light effects

HASSAN A. SIAL, RAMON BALDRICH, AND MARIA VANRELL

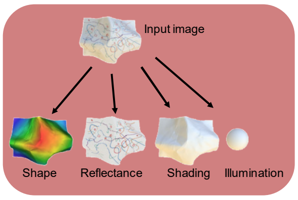
Journal of the Optical Society of America A, Vol. 37, Issue 1, 2020

Introduction

Intrinsic image decomposition is an inverse optics process to get internal characteristics such as shape, shading, reflectance, illumination and specular highlights[1]. Estimation of intrinsic images still remains a challenging task due to weaknesses of ground-truth datasets, which either are too small, or present non-realistic issues. On the other hand, end-to-end deep learning architectures start to achieve interesting results that we believe could be improved if important physical hints were not ignored. In this work we present a twofold framework:

- ➢Flexible generation of images overcoming some classical dataset problems like larger size jointly with coherent lighting appearance;
- ➢Flexible architecture tying physical properties through intrinsic losses.

Our proposal is versatile, presents low computation time and achieves state-of-art results.

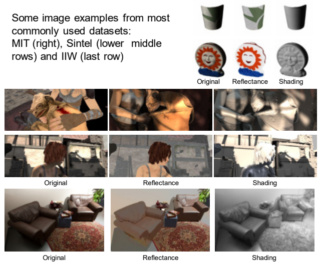


Current intrinsic image dataset and challenges

Making/Building intrinsic image datasets is a challenging task that requires accurate controlling of lights, camera and objects positions. Following table lists few intrinsic image datasets with corresponding properties.

Dataset	Size (#Images)	Model Fulfillment	Training on full Image	Diversified Background	Cast Shadows	Consistent Lighting
MIT (*Grosse et al. [2]*)	220	yes★	no	no	no	yes
IIW (*Bell et al. [3]*)	5230	no	no	yes	yes	yes
MIII (*Beigpour et al. [39]*)	75	yes	no	no	no	yes
Sintel (*Butler et al. [4]*)	890	no	yes	yes‡	yes	yes
ShapeNet (*Shi et al. [5]*)	330K	yes	no	yes	no	no
Shapenet-Intrinsic (*Baslamisli et al. [6]*)	20K	yes	no	yes	no	no
Our dataset (SID)	25K	yes	yes	yes	yes	yes
Baslamisli et al. [7]	35K	yes	no†	yes	yes	no
CGintrinsics (*Li and Snavely [8]*)	20K	yes	no†	yes	yes	no
InteriorNet (*Li et al. [41]*)	20M	no	no†	yes	yes	no

Table 1. Comparison on current available dataset according to several properties. From left to write we account for: Number of images, GT perfectly fulfills the physical model, GT is on the full image or only a part, GT is presenting the influence of a diverse background, GT is presenting cast shadows apart from shading, and global image present physically consistent lighting. Meaning of special cases: (★) MIT dataset generally fulfills product model by including a factor i.e. $I = \alpha(R \cdot S)$, but it does not completely hold for all images and have small deviation; (‡) Sintel dataset present diverse backgrounds compared to the rest, but with a strong bias towards specific colors due to high correlation of a video sequences. (†) Training area is large, but still does not cover the full image.

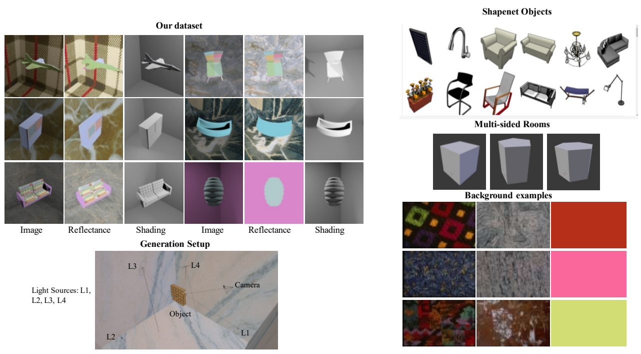


Some image examples from most commonly used datasets: MIT (right), Sintel (lower middle rows) and IIW (last row)

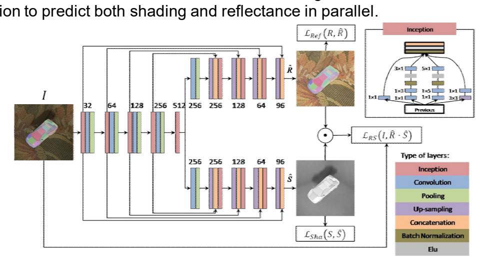
Surreal Intrinsic Dataset (SID)

Based on random Shapenet[5] objects

- ➢ Random selection of background, which corresponds to only reflectance change.
- ➢ 4 fixed light with random intensity
- ➢ 3 different indoor environment for shading variation.
- ➢ Random Camera position in semi sphere around objects. 2 images for each object.



IUI Inception based U-Net architecture for intrinsic image estimation with double loss-function to predict both shading and reflectance in parallel.

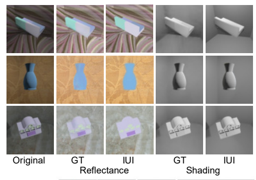


Results

Our Dataset

Method (where tested)	Reflectance MSE	Reflectance LMSE	Reflectance DSSIM	Shading MSE	Shading LMSE	Shading DSSIM
Retinex (whole image) [2]	0.0500	0.049	0.17	0.0400	0.0403	0.24
***IUI* (foreground object)**	0.0046	0.0038	0.029	0.0023	0.0020	**0.0178**
	(10.9)	*(12.9)*	*(5.9)*	*(17.4)*	*(20.2)*	*(13.5)*
***IUI* (background walls)**	**0.0016**	**0.0014**	**0.019**	**0.0010**	**0.0008**	0.023
	(31.3)	*(35.0)*	*(8.9)*	*(40.0)*	*(50.4)*	*(10.4)*
***IUI* (whole image)**	0.0020	0.0019	0.020	0.0011	0.0009	0.022
	(25.0)	*(25.8)*	*(8.5)*	*(36.4)*	*(44.8)*	*(10.9)*

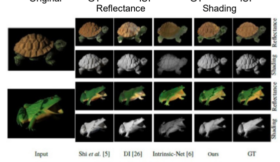
Table 2. Errors for reflectance and shading predictions on our dataset. Comparison between our IUI architecture and Retinex algorithm. IUI decreases the error of Retinex by the factor given in brackets. Errors are separately reported on object, on foreground and the on whole image.



MIT Dataset

Method	Reflectance MSE	Reflectance LMSE	Reflectance DSSIM	Shading MSE	Shading LMSE	Shading DSSIM
Retinex [2]	**0.0032**	0.0353	0.1825	0.0348	0.1027	0.3987
SIRFS [18]	0.0147	0.0416	0.1238	0.0083	0.0168	0.0985
Direct Intrinsics [26]	0.0277	0.0585	0.1526	0.0154	0.0295	0.1328
ShapeNet [5]	0.0278	0.0503	0.1465	0.0126	0.0240	0.1200
CGIntrinsics [8]	0.167	0.0319	0.1287	0.0127	0.0211	0.1376
IntrinsicNet [6]	**0.0051**	0.0295	0.0926	**0.0029**	**0.0157**	**0.0441**
RetiNet [6]	0.0128	0.0652	0.0909	0.0107	0.0746	0.1054
IUI	**0.0046**	**0.0197**	**0.054**	0.0038	0.020	0.0557

Table 3. Estimation errors on MIT dataset reported in previous works by different methods and for our IUI architecture.

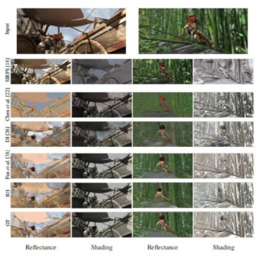


Qualitative Results on MIT intrinsic image dataset, compared to other methods, we achieved sharp and better colors and removed shading effects. Our method performed best in bringing reflectance details from dark part of image.

Sintel Dataset

Method	Reflectance MSE	Reflectance LMSE	Reflectance DSSIM	Shading MSE	Shading LMSE	Shading DSSIM
Retinex [2]	0.0606	**0.00366**	0.227	0.0727	0.0419	0.24
Lee *et al.* [21]	0.0463	0.02224	0.199	0.0507	0.0192	0.177
SIRFS [18]	0.042	0.0298	0.21	0.0436	0.0264	0.206
Chen and Koltun [22]	0.0307	0.0185	0.196	0.0277	0.019	0.165
Direct Intrinsics [26]	0.01	0.0083	**0.02014**	0.0092	0.0085	0.1505
Fan *et al.* [38]	**0.0069**	**0.0044**	**0.1194**	**0.0059**	**0.0043**	**0.0822**
IUI fine-tuned on CS	**0.0072**	**0.0054**	0.1374	0.0068	**0.0059**	0.1247
IUI without fine-tuning	0.023	0.015	0.21	0.035	0.022	0.255
IUI fine-tuned on GLS	**0.0062**	**0.0047**	0.1297	**0.0057**	**0.0048**	0.1183

Table 4. Results on Sintel Image Split dataset. Best errors are highlighted in bold.



Visual comparison on MPI-Sintel dataset using image split.

Visual comparison on MPI-Sintel using Scene split.

Method	Reflectance MSE	Reflectance LMSE	Reflectance DSSIM	Shading MSE	Shading LMSE	Shading DSSIM
Direct Intrinsics [26]	0.0238	0.0155	0.226	0.0205	0.0172	0.1816
Fan *et al.* [38]	**0.0189**	**0.0122**	**0.1645**	**0.0171**	**0.0117**	**0.1450**
IUI fine tuned on CS	0.0213	0.0140	0.1787	0.0253	0.01721	0.1874
IUI without fine-tuning	0.023	0.0154	0.20	0.034	0.023	0.24
IUI fine tuned on GLS	**0.01733**	**0.0110**	**0.16189**	0.0201	0.013182	0.1618

Table 5. Result on Sintel Scene Split dataset. Best errors are highlighted in bold.

IIW Dataset

Qualitative results on IIW

Method	WHDR(mean)	runtime(sec)
Shen *et al.* [48]	36.90	297
Retinex(color) [2]	26.89	198.5
Retinex(gray) [2]	26.84	225.3
Graces *et al.* [49]	25.46	5.1
Zhao *et al.* [50]	23.20	34.7
IUI *(without fine-tuning)*	22.50	**0.02**
***L1* flattening** [51]	20.94	310.94
Bell *et al.* [3]	20.64	214
Zhou *et al.* [28]	19.95	300
Nestmeyer *et al.* (CNN) [32]	19.49	**0.006**
Nestmeyer *et al.* [32]	17.69	300.086
Bi *et al.* [51]	17.67	300
Fan *et al.* [38]	**14.45**	0.1

Table 6. Result on IIW dataset

Conclusion

In this work we propose a versatile framework to define and train a convolutional network able to perform a intrinsic decomposition through training on a dataset with a large variety of light effects and color reflectances. The approach presented and evaluated here is a first version, where we have just worked with single white light sources, single background, and a limited number of room shapes, all of them based on flat surfaces. A wide range of variations can be introduced to improve the diversity of the scenes to be trained on. In parallel our proposed CNN architecture has been defined in a simplistic way to reduce its number of parameters and enough flexible to be adapted to multiple type of visual tasks related to light effect estimation. Apart from intrinsic decomposition it can be easily extended to color constancy or cast shadow removal, we already have preliminary results on these fields. The results obtained by all the experiments we report in this paper, make us to be optimistic about the capabilities of the presented approach to train networks devoted to solve taks related to the estimation of light effects. In all the reported experiments we show a performance close to the state of the art of the problem of intrinsic decomposition in shading and reflectance.

[1] Barrow, H. G., Tenenbaum, J. M., Hanson, A. R., & Riseman, E. M. (1978). *Computer vision systems* (pp. 3-26).
[2] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, Ground-truth dataset and baseline evaluations for intrinsic image algorithms," in International Conference on Computer Vision, pp. 2335{2342, 2009.
[3] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," ACM Transactions on Graphics (TOG), vol. 33, no. 4, p. 159, 2014.
[4] D. J. Butler, J. Wul, G. B. Stanley, and M. J. Black, A naturalistic open source movie for optical ow evaluation," in European Conf. on Computer Vision (ECCV) (A. Fitzgibbon et al. (Eds.), ed.), Part IV, LNCS 7577, pp. 611{625, Springer-Verlag, 2012.
[5] Shi, Jian, et al. "Learning non-lambertian object intrinsics across shapenet categories." Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017.
[6] Baslamisli, Anil S., Hoang-An Le, and Theo Gevers. "CNN based learning using reflection and retinex models for intrinsic image decomposition." CVPR 2018.
[7] Baslamisli, Anil S., et al. "Joint learning of intrinsic images and semantic segmentation." ECCV. 2018.
[8] Li, Zhengqi, and Noah Snavely. "Cgintrinsics: Better intrinsic image decomposition through physically-based rendering." *ECCV*. 2018.
[26] Narihira, Takuya, Michael Maire, and Stella X. Yu. "Direct intrinsics: Learning albedo-shading decomposition by convolutional regression." ICCV 2015.
[38] Fan, Qingnan, et al. "Revisiting deep intrinsic image decompositions." *Proceedings of the CVPR*. 2018.
[41] Li, Wenbin, et al. "InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset." *arXiv preprint arXiv:1809.00716* (2018).
For other citation see publications on https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-37-1-1