IID-NORD: A COMPREHENSIVE INTRINSIC IMAGE DECOMPOSITION DATASET

Diclehan Ulucan, Oguzhan Ulucan, Marc Ebner

Institut für Mathematik und Informatik, Universität Greifswald, Greifswald, Germany

ABSTRACT

The goal of intrinsic image decomposition is to recover low level features of images. Most of the studies tend to consider only reflectance and shading, even though it is known that increasing the number of intrinsics is beneficial for many applications. Existing intrinsic image datasets are quite limited. In this study, a dataset is introduced to provide a comprehensive benchmark to the field of intrinsic image decomposition. IID-NORD contains a large number of scenes and for each scene ground truth reflectance, shading, surface normal vectors, light vectors and depth map is provided to allow detailed decomposition. Moreover, diverse illuminants, viewing angles, and dynamic shadows are used to prevent any bias. To the best of available knowledge, IID-NORD is the most comprehensive dataset in the field of intrinsic image decomposition. IID-NORD will be available on the first author's official webpage.

***Index Terms*—** Intrinsic image decomposition, dataset, computer graphics

1. INTRODUCTION

While the human visual system enables us to differentiate between colors, estimate distances and perceive details of a scene under different illumination without any prior information, machines have difficulties in carrying out such tasks [1, 2]. For instance, in computer vision systems efficiency decreases due to detail loss in the presence of over- and/or under-saturated regions, glare effect and reflection in applications such as object detection and shape extraction [3]. Moreover, an ambiguity caused by light may disturb the pixel continuity at edges, thus carrying out a segmentation task can be troublesome for a machine system. Solutions to these problems can be proposed through intrinsic image decomposition (IID) algorithms.

Low-level features of scenes can be represented individually and each representation is called an *intrinsic image* [2]. There are several types of intrinsic images such as the reflectance, distance, surface normal, illumination, luminosity, and specularity [2]. Different intrinsic images allow us to present details depending on distinct characteristics of a scene more effectively, i.e. the perceived color of an object can be easily extracted from the reflectance image [1]. Moreover, it is more meaningful to use the depth map to estimate distance between objects in a scene [2].

In IID studies mostly the reflectance and shading components are extracted to obtain beneficial information from an image. While the reflectance image (R) provides us the albedo, which corresponds to the ratio between total reflected and total incident illumination, the shading image (S) can be defined as the element indicating the interaction between the surfaces and illumination in a scene [2, 4]. The product of these two components form the input image (I) as follows;

$$I(x, y) = R(x, y) \cdot S(x, y) \tag{1}$$

where x and y represent the pixel location.

Eqn. 1 is known as the so-called intrinsic image decomposition problem, which is by nature under-constrained and has been studied for more than four decades [5]. In order to solve this ill-posed problem, several algorithms have been proposed and a few of them also considered extracting other intrinsic images such as depth and shape to increase the usability of their methods [6, 7]. As it can also be deduced from the milestone study of Barrow *et al.* [2], it is important to extract several intrinsic images from a scene to obtain the most possible information from an input. Therefore, rather than simplifying the IID-problem by only taking the R and S components into account, as many as possible intrinsic images should be obtained to benefit from the low-level features in diverse applications.

A critical point in developing an IID algorithm, which can extract various intrinsic images of a scene is to access ground truth information of all the low-level features aimed to be obtained by the designed model. However, in the field of IID there are only a few publicly available datasets, which contain a limited amount of intrinsic images and also have other shortcomings. The MIT Intrinsic Images Dataset [8] contains only 20 real objects, which do not have a background. For each object an image captured under different lighting conditions is created to provide different scenes. The ground truth information contains the reflectance and shading components as well as the binary mask, diffuse component and specularity information of the image. Since it is a small dataset it is inadequate for utilization in neural networks-based algorithms. Furthermore, considering it contains only images with a single object without any background, this dataset is not sufficient for studies aiming at developing a method applicable for real-world scenes. The widely known MPI Sintel Flow Dataset [9] is in fact designed for optical flow evaluation, but it is used in some IID studies as well [10]. It consists of a limited number of scenes extracted from a 3D fantasy short-film called Sintel. Hence, it is not very suitable for real-world applications. The Intrinsic Images in the Wild Dataset [11] is formed with the help of human operators. It is a large-scale dataset, however it is based on pairwise human ranking decisions and sparse annotations, which makes the dataset subjective and limits the available cues [12, 13]. Also, the rankings are only about albedo, hence this dataset is insufficient for full IID. In the Multi-illuminant Intrinsic Image Dataset [14] real photos of 5 distinct scenes are captured under complex multi-illuminant and multi-colored illumination conditions, while shadows are also included in the scenes. This dataset consists of challenging scenes with ground truths for reflectance, shading, specularity, illumination and preliminary depth information, yet it has shortcomings. The scenes either contain a very limited background or no background at all, which causes a problem for depth estimation tasks. Furthermore, the number of images is inadequate for machine learning-based IID methods. The Multi-view Multi-illuminant Intrinsic Dataset [15] contains scenes with complex multi-illuminants and multi views. Ground truths for raw depth, 3D point cloud, reflectance and shading are provided for the images. The dataset contains 600 high-resolution images, hence it is efficient for IID algorithms based on traditional methods, but not for neural networks-based techniques. Furthermore, the scenes consist only of a few objects and a partial background, which causes the evaluation results of an IID algorithm on this dataset to be questionable for real-world scenes. There are also other datasets created to train algorithms in IID studies, which render 3D models and environmental maps [12]. However, only the implementation of these datasets are publicly available and carrying out a rendering operation for thousands of images requires a high computational power, i.e. high-cost hardware. Furthermore, the images contain only a single 3D model in the foreground, which could be easily segmented, whereas the environmental map serves as background.

For both low- and high-level computer vision tasks such as segmentation and object classification, IID algorithms, which can accurately extract intrinsic images are beneficial. However, when IID methods are benchmarked or trained on datasets containing a single object, very limited number of colors and smooth shadows, and having no background information, then their efficiency in real-world applications tends to decrease. All these mentioned limitations in existing IID datasets indicate that a comprehensive dataset containing various types of intrinsic images, textures, shadows and multiple objects is required to enhance the models designed in this field. Thereupon, in this study a new intrinsic image decomposition dataset called *IID-NORD*, which contains both realistic and artificial scenes, is created by utilizing computer graphics. In IID-NORD, ground truths for reflectance, shading, depth map, surface normal vectors, and light vector map are provided for a total of 128000 distinct scenes. To the best of our knowledge, this is the most detailed publicly available IID dataset in the literature, since it contains a high number of intrinsic images. Furthermore, IID-NORD contains scenes with various shapes, textures, viewing angles, illuminants and dynamic shadows to increase the diversity of the dataset. In order to demonstrate the usability of the IID-NORD it is employed in several IID algorithms and the results are provided in this work.

This paper is organized as follows. Section 2 explains the formation of the dataset in detail. Section 3 presents the experimental results of algorithms used to prove the usability of the dataset. Section 4 gives a brief summary of the study.

2. THE DATASET

Obtaining a dataset for algorithms requiring a high amount of data can be troublesome. An extreme amount of time and labor is needed to form a large-scale dataset with real objects and generate ground truth images for various cases. To avoid this excessive effort, any possible human error and subjectivity, in this study, computer graphics is used during data formation. All images are obtained via designing an algorithm using the open source 3D graphics toolkit called OpenSceneGraph (www.openscenegraph.com). The 3D object models are gathered from the website 3D Warehouse (3dwarehouse.sketchup.com). Some textures are taken from Pixabay (pixabay.com), others are created by the authors. Four different rooms (living room, bedroom, kitchen, garage) containing distinct objects with random placement are rendered in order to avoid possible fitting problems and bias in IID neural network models. As aforementioned, five intrinsic images namely, reflectance, shading, surface normal vectors, light vector map and depth map are provided in this dataset.

In total, 128000 scenes are rendered together with their intrinsics and all images have a resolution of 965×1600 pixels. An example scene together with all of its intrinsic images is presented in Fig. 1.

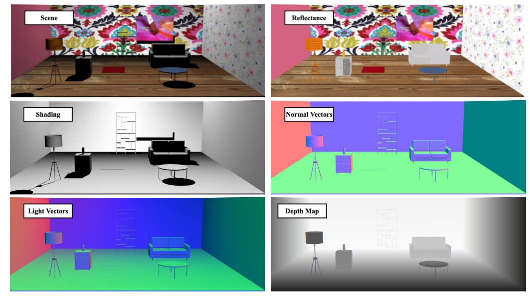

Fig. 1: Example scene and its intrinsic images.

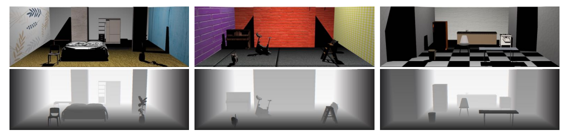

Fig. 2: Scenes (top) and corresponding depth maps (bottom).

First of all, the rooms in the scenes are designed with various shapes to avoid any bias during the training phase of learning-based IID methods. In particular, if all scenes would have the same shape as in Fig. 1, then all the depth maps would have a white spot in the back of the room, which would possibly cause a bias in learning-based IID algorithms. In Fig. 2, rooms with different shapes are presented together with their depth maps. Apart from the shape of the scenes, also different viewing angles are used during rendering to increase the diversity of the dataset (Fig. 3).

A large number of 3D objects with distinct shapes are gathered to create scenes with various items. For each type of object several versions are used to improve the variety in the dataset, i.e., for a lamp 4 different lamp designs are used and the selected lamp is changed in each consecutive render. Furthermore, to avoid any possible bias during the utilization of the dataset in neural networks based IID algorithms and to increase the diversity of the dataset, in each render every object is randomly placed into the scene. This random placement procedure allows the obtainment of various depth maps, which can also be used in other research fields such as color constancy [16].

The textures are selected for each object individually and both realistic and artificial textures are chosen for the 3D assets. Since reflectance and shading are often constant in local regions instead of the global scale, textures with sharp color changeovers are included to the dataset [17]. Furthermore, colorful textures provide beneficial features for different IID studies [18]. Especially in Retinex-based algorithms, where large gradient changes in chromaticity indicate reflectance changes, having a high number of colors in the scene is an advantage [18]. In Fig. 3, scenes with both realistic and artificial textures are demonstrated for different rooms.

For each scene a single light source is positioned at different locations, which enables the light vector maps to be more diverse. Instead of directional lights, which cause the illumination and shadows to be static, point light sources are used to obtain realistic illumination and dynamic shadows, i.e., according to the position of the light source the amount of light a particular point in a scene receives changes in each render and the orientation of the shadows varies.

The main illuminant used during rendering of scenes is pure white light, however different illuminants are also employed since they are usually ignored in other IID datasets. Lights with corresponding values at $2000K$, $3500K$, $4800K$, $5200K$ and $10000K$ on

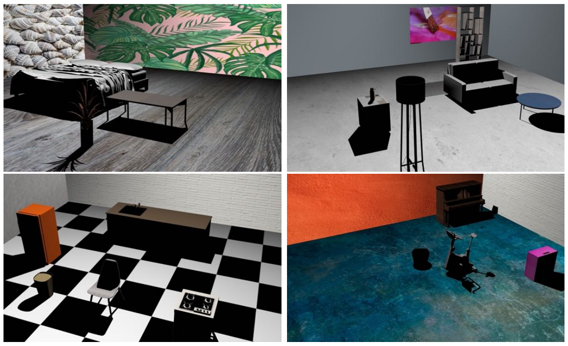

Fig. 3: Example scenes with different objects, textures and viewing angles.

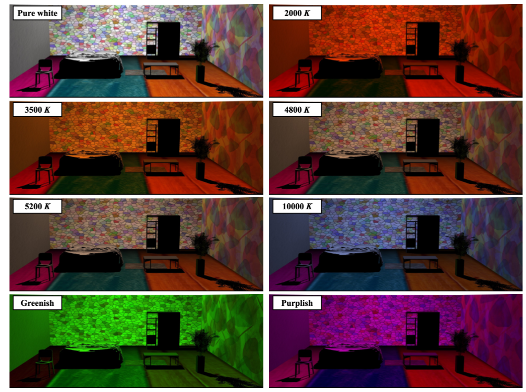

Fig. 4: The same scene under distinct illumination.

the color temperature curve (CTC) are used while rendering. These illuminants are chosen since they can be commonly observed in natural scenes, but are non-canonical illuminants. The image formation process with non-canonical light sources can be expressed as follows;

$$I(x, y) = R(x, y) \cdot S(x, y) \cdot E \tag{2}$$

where E is the color of the non-canonical global light source.

Furthermore, to the best available knowledge, in existing IID datasets, lights outside the CTC are not considered during scene creation although these lights can be commonly observed in diverse applications such as agriculture [19]. In this dataset, strong greenish and purplish lights, which lie outside the CTC are used to render various scenes. In Fig. 4, the same scene is rendered under different illuminants on and outside the CTC.

Alongside creating natural illumination conditions in the scenes, it is also important to obtain realistic shadows which vary according to the light position. In this study, the Light Space Perspective Shadow Maps (LiSPSM) [20] technique is preferred to create dynamic shadows in the scenes, since it mostly avoids aliasing artifacts, which are common in other methods, such as uniform or perspective shadow mapping, due to quantization and perspective projection. While uniform shadow mapping presents fine results for distant objects and perspective shadow maps for close items, LiSPSM distributes the perspective error throughout the scene and provides satisfying outcomes for all objects. In Fig. 3, it can be observed that the shadows change according to the light position in the scene.

Lastly, it is worth to mention that, aside from intrinsic image decomposition studies, IID-NORD can be employed in various fields including but not limited to color constancy, image segmentation, shadow removal and depth estimation.

3. EXPERIMENTS

In this section, several existing IID studies and an image enhancement method which includes an intrinsic image decomposition part are investigated to evaluate the usability of IID-NORD and observe the algorithms' response to its challenges. The utilized algorithms are briefly explained in the following. In the study of Shen *et al.* [21], IID is formulated as an optimization problem relying on the assumption that in a local image patch the neighboring pixels with similar intensity values should have similar reflectance values. To obtain better results, user scribbles are also added to the problem formulation. Experiments are carried out on various images including the MIT Intrinsic Images Dataset. In the study of Hauagge *et al.* [22], an ambient occlusion measure is computed via a statistical approach and used to derive the reflectance and illumination components of an image. The ambient occlusion is estimated from a stack of images captured under different and unknown illumination. The experiments are carried out on various scenes and the MIT Intrinsic Images Dataset. In the unsupervised deep learning approach of Lettry *et al.* [13], an IID algorithm trained on images pairs is proposed to obtain reflectance and shading. The method takes two images of the same scene under distinct illuminants and benefits from the fact that the reflectance is constant in both images. Unlike most IID algorithms, this method outputs a colored shading component. Both synthetic and real-world data are used in this study. The work of Ren *et al.* [23] is in fact a low-light enhancement method but it proposes an IID method within the study, which relies on the illumination and reflectance estimation of the Retinex model. The algorithm estimates a piece-wise smoothed illumination and a noise-suppressed reflectance.

Table 1: Statistical analysis of the algorithms on the subset of IID-NORD.

Algorithm	Shen	Lettry	Hauagge	Ren
PSNR(dB)	10.236	8.338	**10.237**	8.634
SSIM	**0.679**	0.370	0.591	0.612

For the experiments, all the codes are taken from the official webpages of the authors. The input requirements of the algorithms are met and no optimization is carried out on the methods.

IID-NORD is a challenging dataset with complex scenes, multiple colors and shadows, thus analyzing it only statistically would not provide enough insight to its usability, since most of the state-of-the-art intrinsic image decomposition algorithms face difficulty in realistic scenes with strong shadow casts. On the other hand, providing only visual comparisons would lead to subjectivity. Therefore, both quantitative analyses and visual comparisons are carried out in this study. The statistical performance of the algorithms on IID-NORD are presented by using the well-known error metrics; structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). SSIM is inspired from the human visual system and evaluates the structural differences between the images and presents a score in range $[0, 1]$, where a score closer to 1 indicates a better result [24]. PSNR computes the peak signal-to-noise ratio between the inputs in decibels (dB), where a higher value indicates a superior outcome [25].

Intrinsic image decomposition algorithms can be employed in various computer vision pipelines as a pre-processing technique where they need to decompose complex scenes in to their intrinsic features. However, one of the main problems in many existing IID

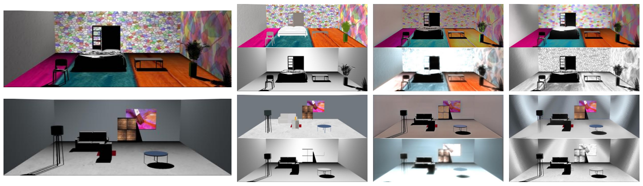

Fig. 5: Comparisons for methods of Lettry [13] and Shen [21]. (Left-to-right) Input scene, ground-truth intrinsics, results of Lettry and Shen.

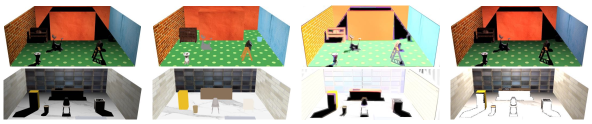

Fig. 6: Comparisons for methods of Ren [23] and Hauagge [22]. (Left-to-right) Input scene, ground-truth reflectance, results of Ren and Hauagge.

studies is their unsatisfying performance on images with multiple objects, shadow casts and complex textures. Although in real-life such features are commonly observed in captured images, they are usually not addressed in the literature. Since IID-NORD contains complex scenes it is not surprising that the intrinsic image decomposition algorithms do not present state-of-the-art performance as demonstrated in Table 1. It is important to note here that the statistical results given in Table 1 are obtained by using images rendered only under pure white light.

Figure 5 presents the outcomes for Lettry and Shen. Lettry outputs quite accurate shading components, while it appeared to have a difficulty in eliminating the shadows in the reflectance image, which is also pointed out by the authors in their study. This drawback is also reflected to the statistical scores of the algorithm and the lowest scores are obtained by Lettry. Since Lettry is a learning-based approach, it can be deduced that if its training set would contain complex samples available in IID-NORD, then the outcomes would likely be better. The outcomes of Shen present that it successfully obtains most of the reflectance information, while also handling some of the strong shadow casts. Moreover, the shading component is extracted quite accurately apart from the fact that it contains the textures of the scene.

In Fig. 6 example results are given for Hauagge and Ren. The highest PSNR score is achieved by Hauagge, whose satisfying performance can also be seen in its reflectance outputs. The complex textures are handled well and noise is relatively low. In the outcomes of Ren, the reflectance image contains blurred parts, yet the structures are mainly well recovered. It is worth to stress that this method is a part of a low-light enhancement pipeline. The reason behind using it for the comparisons of this study is to demonstrate that IID can be helpful in a variety of applications, hence the usability of IID-NORD is not limited to the field of IID. For example, with slight modifications the images can be easily used in low-light enhancement or saturation correction applications.

As aforementioned, ilumminants outside the CTC are not present in IID datasets and usually different lights are ignored during scene formation. To point out this fact, example outcomes for inputs rendered with an illuminant corresponding to $5200K$ and purplish light outside the CTC are presented in Fig. 7. As it can be observed both traditional and neural networks-based methods face a great challenge

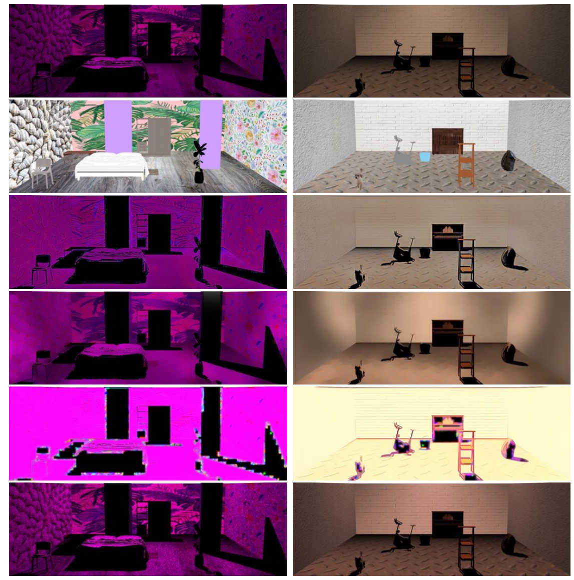

Fig. 7: Comparisons of algorithms on different illuminants. (Top-to-bottom) Input scenes, ground truth reflectances, results of Lettry [13], Shen [21], Ren [23] and Hauagge [22].

in extracting the reflectance images.

The ambiguous and less satisfying regions in the outcomes of the IID methods show that various data is required to benchmark algorithms. Furthermore, they lead to the conclusion that comprehensive and detailed datasets are essential to develop robust IID methods, which can be efficiently used in real-life applications.

4. CONCLUSION

In this study, a large-scale intrinsic image decomposition dataset IID-NORD is created to provide a comprehensive benchmark. While forming IID-NORD, an open source 3D graphics toolkit is used to avoid any possible human error and subjectivity. A total of 128000 images each with five intrinsic features: reflectance, shading, surface normal vector, light vector map and depth map are rendered. Different room shapes and viewing angles, objects with distinct textures, and diverse illumination types are used to prevent possible bias and fitting problems. Several existing intrinsic image decomposition studies are evaluated on IID-NORD to analyze the usability of the dataset. It is observed that IID-NORD challenges the state-of-the-art intrinsic image decomposition methods, but also allows proper decomposition.

To the best of our knowledge, IID-NORD is the most detailed publicly available intrinsic image decomposition dataset, which is considered to contribute not only to the field of intrinsic image decomposition, but also to the studies in color constancy, shadow removal, depth estimation and segmentation. As future work, the number of objects in the scenes will be increased and more complex scenes will be created to improve the usability of the dataset. Moreover, multiple light sources will be added to the scenes.

5. ACKNOWLEDGEMENT

The authors are thankful to Trajce Nikolov and Daniel Lehmann for sharing their experience about OpenSceneGraph.

6. REFERENCES

[1] M. Ebner, *Color Constancy, 1st ed.*, Wiley Publishing, 2007.

[2] H. Barrow, J. Tenenbaum, A. Hanson, and E. Riseman, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst*, vol. 2, no. 3-26, pp. 2, 1978.

[3] O. Ulucan, D. Karakaya, and M. Turkan, "Saturated region recovery in tone-mapped HDR images," in *IEEE Int. Conf. Image Process.*, 2021, pp. 1774–1778.

[4] L. Shen and C. Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2011, pp. 697–704.

[5] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[6] J. T. Barron and J. Malik, "Intrinsic scene properties from a single RGB-D image," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 17–24.

[7] J. T. Barron and J. Malik, "Shape, albedo, and illumination from a single image of an unknown object," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 334–341.

[8] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *IEEE Int. Conf. Comput. Vision*, 2009, pp. 2335–2342.

[9] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Eur. Conf. Comput. Vision*, 2012, pp. 611–625.

[10] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *IEEE Int. Conf. Comput. Vision*, 2013, pp. 241–248.

[11] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–12, 2014.

[12] J. Shi, Y. Dong, H. Su, and S. X. Yu, "Learning non-lambertian object intrinsics across shapenet categories," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1685–1694.

[13] L. Lettry, K. Vanhoey, and L. Van Gool, "Unsupervised deep single-image intrinsic decomposition using illumination-varying image sequences," in *Comput. Graph. Forum*, 2018, pp. 409–419.

[14] S. Beigpour, A. Kolb, and S. Kunz, "A comprehensive multi-illuminant dataset for benchmarking of the intrinsic image algorithms," in *IEEE Int. Conf. Comput. Vision*, 2015, pp. 172–180.

[15] S. Beigpour, M. L. Ha, S. Kunz, A. Kolb, and V. Blanz, "Multi-view multi-illuminant intrinsic dataset.," in *Brit. Mach. Vision Conf.*, 2016.

[16] M. Ebner and J. Hansen, "Depth map color constancy," *Bio-Algorithms Med-Syst.*, vol. 9, no. 4, pp. 167–177, 2013.

[17] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 425–436, 2013.

[18] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2008, pp. 1–7.

[19] H. L. Smith, L. McAusland, and E. H. Murchie, "Don't ignore the green light: exploring diverse roles in plant processes," *J. Exp. Botany*, vol. 68, no. 9, pp. 2099–2110, 2017.

[20] M. Wimmer, D. Scherzer, and W. Purgathofer, "Light space perspective shadow maps," in *Eurographics Conf. Rendering Techn.*, 2004, pp. 143–151.

[21] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 2011, pp. 3481–3487.

[22] D. Hauagge, S. Wehrwein, K. Bala, and N. Snavely, "Photometric ambient occlusion for intrinsic image decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, pp. 639–651, 2015.

[23] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862–5876, 2020.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, 2004.

[25] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, 3rd ed.*, Prentice Hall, 2018.