Intrinsic Image Decomposition based on Retinex Theory, Superpixel Segmentation and Scale-Space Computations

Diclehan Ulucan[0000−0002−7059−302X], Oguzhan Ulucan[0000−0003−2077−9691], and Marc Ebner[0000−0003−2725−2454]

University of Greifswald, Institute of Mathematics and Computer Science
17489 Greifswald, Germany
{diclehan.ulucan, oguzhan.ulucan, marc.ebner}@uni-greifswald.de

Abstract. Intrinsic image decomposition enables us to estimate the low-level features of images. Due to the benefits it provides and the challenges it holds, intrinsic image decomposition has been extensively studied over more than five decades. It can be utilized in various computer vision and computer graphics pipelines to improve the efficiency of tasks such as object classification and recoloring, and image segmentation. In this study, we introduce an algorithm for reflectance and shading estimation, offering a simple yet effective solution to the ill-posed intrinsic image decomposition problem. Our learning-free method leverages a combination of the fundamentals of the Retinex theory, scale-space computations, and superpixel segmentation. The assumptions of the Retinex theory enable us to provide a straightforward solution to a complex problem, while scale-space computations allow us to highlight low-level features and superpixel segmentation helps us to preserve local information. We evaluated our algorithm that mainly focuses on single objects on three benchmarks, namely, MIT Intrinsic Images, Bonneel, and MPI Sintel, which either consist of single objects or complex scenes having different characteristics. According to our experiments, our algorithm provides competitive results compared to the other methods.

Keywords: Intrinsic image decomposition · Reflectance estimation · Shading estimation.

1 Introduction

The human visual system estimates the true colors of the objects unconsciously by discounting the effects of the light source present in the environment [21]. However, machine vision systems have difficulty performing this task effectively due to the presence of occlusion, reflections, glare effect, and over- and under-saturated regions [42,45]. To assist machine vision systems and increase their performance in numerous applications, we can benefit from the field of intrinsic image decomposition (IID) [3]. The main aim of intrinsic image decomposition is to recover a scene's low-level features such as reflectance, and shading, which are also referred to as "family of intrinsics" [3]. While

[1] This is the pre-print version of the manuscript accepted at the Computational Color Imaging Workshop (CCIW 2024)

reflectance carries information about the actual color of the objects irrespective of the scene's lighting conditions and the position of the capturing device, the shading component includes features related to the illumination, shadows, and geometry of the scene. Hence, each family member contains distinct characteristics that carry cues about the input scene which might not be directly observable in a single scene. Therefore, using intrinsics rather than the input image itself can benefit both high- and low-level computer vision tasks, including but not limited to object recoloring [8], surface retexturing [48], exposure correction [52], and object classification [5].

Intrinsic image decomposition is framed as a computational challenge, since the problem is severely under-constraint, i.e., we only have a single input image but multiple unknown intrinsics. Due to this reason, studies in this field generally relax the ill-posed nature of the problem by assuming that any scene can be represented using the Lambertian image formation model. The under-constraint nature of the problem is further relaxed by assuming that the sensor responses of the capturing device are narrow-band, and there is a point light source illuminating the scene uniformly. Consequently, any input scene I at the spatial location $(\mathbf{x}, \mathbf{y})$ can be formulated as follows:

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

where R is the reflectance that represents the ratio between the total reflected and total incident illumination, and S is the shading that presents the interaction between the illumination and the surface [3].

Even though these relaxations simplify the intrinsic image decomposition problem, it is still quite challenging to estimate the reflectance and shading of a scene. Therefore, this field has gained significant attention over more than five decades, and various studies have been conducted. Both traditional algorithms relying on image statistics [2,3,11,16,17,22,23,24,26,28,36,37,41,47,53] and learning-based methods relying on neural networks [4,5,6,7,19,30,32,33,35,39,49,50,51] have been introduced to provide an effective solution to the problem. These algorithms may have different necessities, i.e., they may require user scribbles, different focal distances, a depth map, an image sequence captured with different viewing conditions, an input sequence where the light source position varies in each scene, an image stack captured under distinct illumination conditions, or a time-varying image sequence [12,43]. Algorithms performing decomposition by using only a single input image are generally considered to be more beneficial since in real-world applications we may not have any other information than the input scene, requiring user interaction might be inefficient, and creating image stacks in the appropriate format is troublesome [43].

Several of the intrinsic image decomposition algorithms build their strategies on the fundamentals of the assumptions of the Retinex theory [28]. The utilization of this theory is not surprising since the Retinex method is based on the analysis of the human visual system, and it is known that algorithms making use of biological findings tend to provide efficient solutions with a low computational complexity [45,53]. In this study, we propose a Retinex-based learning-free intrinsic image decomposition method operating in scale-space and relying on superpixel segmentation. We carry out our computations at multiple scales since the effectiveness of using scale-space is already proven in the field of intrinsic image decomposition [20,38]. Furthermore, we prefer this approach since in our previous studies, we explicitly demonstrated that the effectiveness of

methods relying on color information can be significantly enhanced using scale-space, which is sensitive to low-level features [46]. Moreover, we integrate superpixel segmentation, whose advantages are also proven in this field [18,27,40], into our algorithm to better preserve local information and reduce possible noise [1]. Our algorithm requires a single RGB image as input and we mainly focus on images containing a single object similar to previous works, i.e., Barron and Malik [2], and Baslamisli *et al.* [6]. Contrary to these studies, we demonstrate our method's outcomes also on more complex scenes.

The remainder of this paper is organized as follows. In Sec. 2, we introduce our method. In Sec. 3, we present the experimental results of our algorithm, and in Sec. 4, we provide a brief summary and discuss possible future directions of our study.

2 Proposed Method

We present an approach based on the fundamentals of the Retinex theory while utilizing scale-space computations, and superpixel segmentation to estimate the reflectance and shading of a single image. As aforementioned, we benefit from the Retinex theory, since it is based on the investigations of the human visual system. Numerous studies trying to solve ill-posed problems related to the image formation model utilize the Retinex theory as a building block of their strategy since algorithms based on the human visual system tend to provide simple yet effective solutions [45,53]. We prefer to use the Retinex theory in multiple scales since in our previous investigations, we explicitly demonstrated that simply carrying computations into scale-space improves the efficiency of color feature-based methods which is not surprising considering that colors are low-level features of the scene and the scale-space is sensitive to such features [46]. At each scale, we also use superpixel segmentation to better preserve local features [1]. In the following, we explain the steps of our strategy.

In our algorithm, we assume that the input image may contain noise, therefore we first apply a 3×3 median filter to the input image I. Then, we form an image pyramid containing the scale-space representation of I. We adaptively determine the number of levels of the image pyramid based on the image resolution as $N = \lfloor \frac{log(min(r,c))}{log(2)} \rfloor - 2$, where r and c are the numbers of rows and columns of the image, respectively. We do not take the coarsest two levels into account since locality degrades substantially at these layers which may negatively affect the preservation of the fine details in the scene [44,46].

At each scale, we utilize the assumptions of the Retinex theory [28] to estimate the reflectance, in particular, we are following the assumptions of the color Retinex method introduced by Grosse *et al.* [24]. The Retinex algorithm enables us to differentiate between image variations that are either caused by reflectance or shading by investigating the local derivatives of the image [53]. While large derivatives are associated with reflectance changes $\hat{r}$, small derivatives are related to changes in shading. Given a certain threshold t, the horizontal gradient ($\mathbb{G}_x$) and the vertical gradient ($\mathbb{G}_y$) can be classified as follows:

$$\hat{r}_x = \begin{cases} \mathbb{G}_x, & \text{if } |\mathbb{G}_x| \geq t_x \\ 0, & \text{otherwise} \end{cases} \qquad \hat{r}_y = \begin{cases} \mathbb{G}_y, & \text{if } |\mathbb{G}_y| \geq t_y \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where $\hat{r}_x$, and $\hat{r}_y$ represent the classified horizontal and vertical gradients for the reflectance.

We assume that there is a single light source illuminating the scene, thus all illumination changes can be represented by a single direction in RGB color space. We can denote this direction as the brightness subspace b, which can be represented by the span of the white vector $[1,1,1]^T$ [24]. This vector demonstrates the uniform changes in all three color channels equally, i.e. it controls the changes in brightness. On the other hand, the changes in chromaticity, i.e., the color quality irrespective of brightness, are available in the chromaticity subspace ch which is the null space of the white vector. We denote the projection of the input image into these subspaces as $\mathbb{G}^b$ and $\mathbb{G}^{ch}$, and utilize separate thresholds for the brightness and chromaticity changes [24]. In the color Retinex method, specific thresholds are chosen for images by utilizing cross-validation. In other words, for a given dataset, all parameter combinations are used for all images, and then for each image, the parameters that provide the lowest mean error on the other images are chosen. Although this approach provides satisfying outcomes on a specific benchmark, its computational complexity is high, and in case the scenes' statistical distributions vary significantly across the dataset the efficiency of the algorithm may decrease. Therefore, we prefer to rely on the statistical information of each image itself to determine effective thresholds adaptively, and compute the gradients for the reflectance $\hat{r}_x$, and $\hat{r}_y$ according to these thresholds as follows:

$$\hat{r}_x = \begin{cases} \mathbb{G}_x^b, & \text{if } \left|\mathbb{G}_x^b\right| \geq \alpha t_x^b \text{ or } \left|\mathbb{G}_x^{ch}\right| \geq \alpha t_x^{ch} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$$\hat{r}_y = \begin{cases} \mathbb{G}_y^b, & \text{if } \left|\mathbb{G}_y^b\right| \geq \alpha t_y^b \text{ or } \left|\mathbb{G}_y^{ch}\right| \geq \alpha t_y^{ch} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

where t_x^b and t_y^b, are associated with the thresholds for brightness with corresponding gradient directions. Similarly, t_x^{ch} and t_y^{ch} are the thresholds controlling the chromaticity information. All thresholds are computed by taking the means of $\mathbb{G}^b$ and $\mathbb{G}^{ch}$. α is a scaling factor which is determined as 0.5 based on practical experiments. Thanks to this adaptive thresholding operation, at this stage, we obtain a lower computational cost than the color Retinex method.

After classifying the gradients, we solve an optimization problem to find a reflectance image $\hat{R}$ that best matches these target gradients at each spatial location as follows:

$$\begin{aligned} \arg\min_{\hat{R}(\mathbf{x},\mathbf{y})} & [\hat{R}(\mathbf{x},\mathbf{y}) - \hat{R}(\mathbf{x}-1,\mathbf{y}) - w_y(\mathbf{x},\mathbf{y})\hat{r}_y(\mathbf{x},\mathbf{y})]^2 \\ & + [\hat{R}(\mathbf{x},\mathbf{y}) - \hat{R}(\mathbf{x},\mathbf{y}-1) - w_x(\mathbf{x},\mathbf{y})\hat{r}_x(\mathbf{x},\mathbf{y})]^2, \end{aligned} \tag{5}$$

where w_x and w_y are the weights emphasizing regions containing significant gradient changes that contribute the most to the overall gradient. We assume that in a local window, the closest pixels contribute more than the pixels that are far away from the region having a gradient change, and this can be controlled by a Gaussian function. To compute the weights, first, we find the local changes Φ by applying a Prewitt filter to the input scene and then, we fit them into a Gaussian curve as, $w = \left|\exp\left(-\Phi/2\sigma^2\right)\right|$ with a scaling factor σ of 3 which is determined based on practical experiments.

After computing an initial reflectance at each scale, we utilize superpixel segmentation to prevent color ambiguities at local regions caused by possible noise and color variations. We divide the image into its superpixels by using simple linear iterative clustering (SLIC) [1]. We prefer to adopt SLIC among other methods such as LSC [31] and SEEDS [10] since it preserves the boundaries well and it has low computational complexity. We compute local average color of each superpixel and consider it as the reflectance of that specific cluster, thus we obtain a coherent reflectance estimate. Subsequently, we further enhance this coherency by using an edge-aware smoothing filter. We apply a guided filter as a post-processing step [25], and obtain the reflectance for each scale of the image pyramid.

Afterwards, we collapse the pyramid $\mathcal{P}\{\hat{R}\}$ to obtain the final reflectance. We collapse the pyramid by first upsampling the coarsest level so that it matches the size of the reflectance image in its consecutive finer scale as follows:

$$\mathcal{R}' = upsample(\mathcal{P}\{\hat{R}\}^N, size(\mathcal{P}\{\hat{R}\}^{N-1})), \tag{6}$$

where $\mathcal{R}'$ is the upsampled image, $upsample$ is an operation that matches the size of $\mathcal{P}\{\hat{R}\}^N$ to the size of $\mathcal{P}\{\hat{R}\}^{N-1}$.

Subsequently, we linearly combine the reflectance image $\mathcal{P}\{\hat{R}\}^{N-1}$ with $\mathcal{R}'$ as follows:

$$\mathcal{R} = average(\mathcal{R}', \mathcal{P}\{\hat{R}\}^{N-1}), \tag{7}$$

where $\mathcal{R}$ is the new reflectance at the scale $N-1$, and $average$ is a function that performs linear combination.

Then, we upsample $\mathcal{R}$ to match the size of the reflectance at scale $N-2$ and linearly combine the resulting image with $\mathcal{P}\{\hat{R}\}^{N-2}$. We continue to perform Eqn. 6 and Eqn. 7 until we reach the finest scale of the pyramid. The resulting image is our final reflectance. After estimating the reflectance component of the scene, we obtain the shading component by applying Eqn. 1.

3 Experiments and Discussions

In this section, we first briefly explain our experimental setup by introducing the datasets and the error metric we utilize. Then, we discuss our algorithm's performance while providing both statistical and visual results. It is worth mentioning that the baseline method given in our results corresponds to the baseline algorithm given in the study of Bonneel *et al.* [12] where the chromaticity image is considered as reflectance, and the square root of the direct average of the RGB channels is taken as shading.

3.1 Datasets and Evaluation Strategies

To benchmark our algorithm we utilize three datasets having different characteristics, such as containing single objects and complex scenes.

The MIT Intrinsic Images dataset [24] is created by Grosse *et al.* in 2009, and since then, it is one of the notable intrinsic image decomposition benchmarks in this field. The MIT Intrinsic Images dataset includes a total of 220 images, where the scenes are created with 20 real objects having different shapes and texture details. Each image

contains a single object placed in front of a plain black background, and each scene is illuminated via a single light source positioned at different locations. The MIT Intrinsic Images dataset considers the reflectance, shading, and diffuse components, alongside the specularity information and the binary masks highlighting the positions of the objects. It is worth to note that similar to other works [4,7,19], we statistically evaluate our method on the images utilized in the evaluation of the original study of this dataset [24].

To analyze how well our algorithm performs on complex scenes, we utilize two datasets namely, the Bonneel dataset [12], and the MPI Sintel dataset [15]. The dataset introduced by Bonneel *et al.* [12] consists of photorealistic scenes. The images contain challenging features such as transmissive surfaces, defocus blur, subsurface scattering and specular elements. The reflectances of the scenes are created using computer graphics, while the shading is obtained by applying Eqn. 1. The MPI Sintel dataset contains 23 different scenes, and 1064 rendered images obtained from an open-source 3D animated short film called Sintel [15]. The dataset is created for the evaluation of optical flow but it has also become a widely used benchmark for intrinsic image decomposition tasks. For each synthetic scene, MPI Sintel provides the ground truth reflectance component of the scene. Since the images in this dataset contain repetitive scenes, in our experiments we utilize random frames from each scene.

To provide a statistical investigation we follow the common practice of performing an objective evaluation through the widely used quality metric in the field of intrinsic image decomposition, namely, local mean squared error (LMSE) which is particularly designed for this task [24]. LMSE is based on the calculations of the classical mean squared error, and it is computed by averaging the MSE scores over overlapping windows. While reporting the results, we calculate LMSE by using a window size of 20 as in the original work of the metric. It is important to note that we report the results of the intrinsic image decomposition methods either according to recent comprehensive studies [4,19,12] or running the codes of the methods published by the authors in their default settings.

3.2 Discussion

In Table 1, we present our statistical results for the MIT Intrinsic Images dataset where we compare our algorithm with 19 different traditional and learning-based methods. Our algorithm is among the top-three best performing methods, while the two neural networks-based models achieving lower LMSE scores are fine-tuned on this dataset. Compared to the reflectance, we obtain a lower mean error for the shading component where we greatly preserve the structures in the objects (Fig. 1). It is worth mentioning that we conducted an ablation study on our algorithm where we analyzed the contributions of the individual parts of our method. First, we removed both the scale-space computations and the superpixel segmentation, then we removed either the scale-space computations or the superpixel segmentation and we noted that in case we remove one or two of our main steps, our algorithm's error increases from 0.0301 up to 0.0351 on average. Thus, each part of the algorithm has a contribution to its performance.

As given in Table 1, our algorithm focusing on single objects also performs well on the Bonneel dataset which contains significantly more complex scenes than the MIT Intrinsic Images dataset. Without any modification, we outperform several algorithms

Table 1. Statistical comparison of the performance on the MIT Intrinsic Images, Bonneel, and MPI Sintel datasets. For each dataset, LMSE scores are given. The top results are highlighted using color coding as follows, best: blue, second-best: cyan, third-best: green, forth-best: yellow, and fifth-best: orange. The methods highlighted with * are fine-tuned on the corresponding datasets. To follow the original work of Bonneel *et al.* [12], we only report the mean LMSE score on this dataset. For the MPI Sintel dataset, we do not render the shading component, and we prefer to use the actual ground truth reflectance provided by the benchmark to evaluate our results. Also, for the MPI Sintel dataset, we set a fixed threshold for the method of Grosse *et al.* [24].

MIT Intrinsic Images dataset [24]

Algorithm	Reflectance	Shading	Average	Algorithm	Reflectance	Shading	Average
Baseline	0.0704	0.0685	0.0694	Yuan *et al.* (2019) [51]	0.0462	0.0537	0.0499
Grosse *et al.* (2009) [24]	0.0529	0.0774	0.0652	Ma *et al.* (2020)* [33]	0.0212	0.0192	0.0202
Gehler *et al.* (2011) [23]	0.0393	0.0282	0.0338	Xu *et al.* (2020) [49]	0.0614	0.0672	0.0643
Shen *et al.* (2011) [37]	0.0523	0.0522	0.0523	Liu et al (2020) [32]	0.0640	0.0474	0.0557
Zhao *et al.* (2012) [53]	0.0512	0.0253	0.0383	Ren *et al.* (2020) [36]	0.0486	0.0737	0.0611
Barron and Malik (2014) [2]	0.0491	0.0281	0.0386	Baslamisli *et al.* (2021) [4]	0.0390	0.0447	0.0419
Shi *et al.* (2017) [39]	0.0606	0.0595	0.0601	Baslamisli *et al.* (2021) [7]	0.0438	0.0418	0.0428
Baslamisli *et al.* (2018) [6]	0.0854	0.2038	0.1446	Qian *et al.* (2021) [35]	0.0291	0.0319	0.0305
Lettry *et al.* (2018) [29]	0.0436	0.0472	0.0454	Das *et al.* (2022)* [19]	0.0210	0.0220	0.0215
Yu and Smith (2019) [50]	0.0573	0.0765	0.0669	**Proposed**	0.0336	0.0266	0.0301

The Bonneel dataset [12]

Algorithm	Average	Algorithm	Average
Bousseau *et al.* (2009)* [14]	0.0299	Barron and Malik (2014)* [2]	0.0352
Grosse *et al.* (2009)* [24]	0.0289	Bell *et al.* (2014)* [9]	0.0262
Shen *et al.* (2011)* [37]	0.0201	Narihira *et al.* (2015)* [34]	0.0212
Gehler *et al.* (2011)* [23]	0.0217	Zhou *et al.* (2015)* [54]	0.0190
Zhao *et al.* (2012)* [53]	0.0301	Lettry *et al.* (2018) [29]	0.0260
Garces *et al.* (2012)* [22]	0.0405	Ren *et al.* (2020) [36]	0.0287
Bonneel *et al.* (2014)* [13]	0.0275	**Proposed**	0.0244

MPI Sintel dataset [15]

Algorithm	Reflectance	Algorithm	Reflectance
Baseline	0.0415	Lettry *et al.* (2018) [29]	0.0464
Grosse *et al.* (2009) [24]	0.0450	Ren *et al.* (2020) [36]	0.0393
Shen *et al.* (2011) [37]	0.0399	**Proposed**	0.0402
Zhao *et al.* (2012) [53]	0.0676		

on this benchmark, while some of the existing methods achieve lower LMSE scores. However, it is important to note that most of the results utilized for comparison are taken from the study of Bonneel *et al.*, where the authors fine-tuned all methods on this dataset [12]. In other words, several distinct parameter combinations are used for each method, and for each image, the best-performing parameters are kept, e.g., the lowest LMSE achieved for each scene is reported. As shown in Fig. 2, overall, our algorithm recovers the colors in the reflectance well, while strong illumination conditions can leak into the reflectance component as in other algorithms.

We present our algorithm's statistical analysis on a subset of the MPI Sintel dataset in the last part of Table 1. Our method designed for single objects, can perform effi-

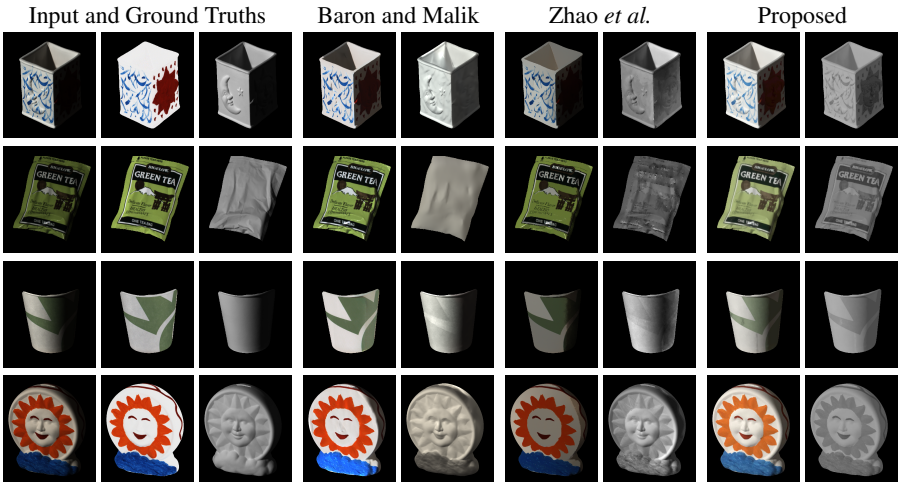

Fig. 1. Visual comparison of the algorithms on random examples on the MIT Intrinsic Images dataset.

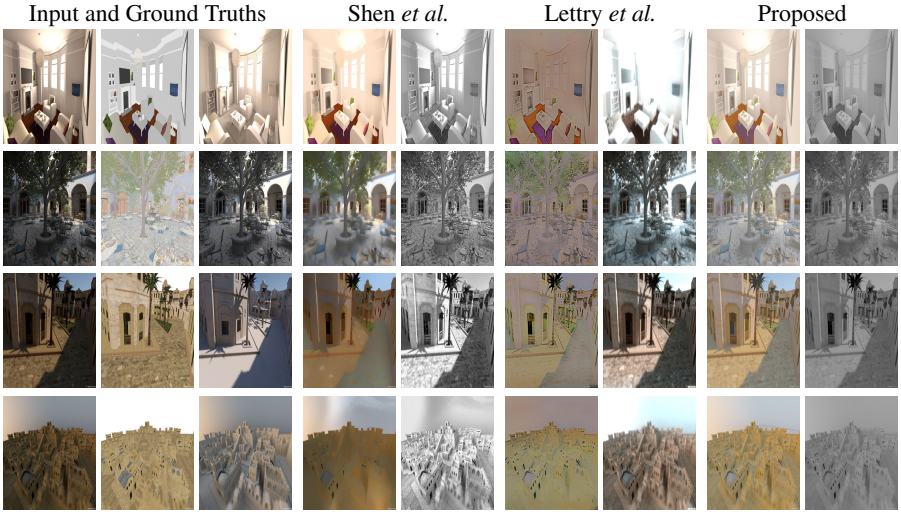

Fig. 2. Visual comparison of the algorithms on random examples on the Bonneel dataset.

ciently on this dataset and preserve the structures in the images well. The visual results for MPI Sintel are provided in Fig. 3.

As a final note, our algorithm achieves competitive results compared to other methods yet it suffers from widely recognized issues such as shadow, illumination and texture leakage, which are challenges shared by other algorithms in the field [6,17,29,37,42]. For instance, challenges due to shadows can be seen on the third row in Fig. 2 where shadows leaked into the estimated reflectances of the algorithms.

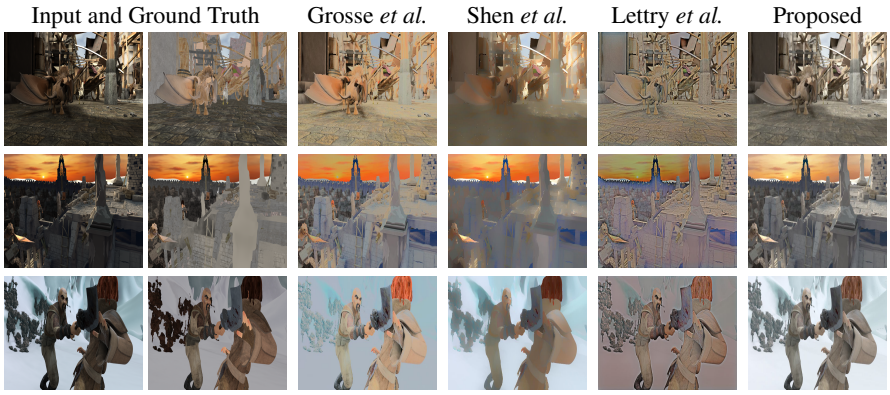

Fig. 3. Visual comparison of the algorithms on random examples on the MPI Sintel dataset.

4 Conclusion

Intrinsic images present different features of images and they can be used in various computer vision and computer graphics applications such as surface re-texturing and exposure correction to improve the outcomes. We have introduced a reflectance and shading estimation strategy that relies on the Retinex theory, scale-space computations, and superpixel segmentation. All components of our method contribute to its efficiency by giving importance to different aspects of the intrinsic image decomposition problem. We demonstrate that our simple yet effective method can achieve competitive results compared to the state-of-the-art without requiring a huge amount of data which we see as a strength of our algorithm. As future work, we will focus on the limitations of our algorithm such as shadow, illumination, and texture leakage to further improve our outcomes, particularly on complex scenes from new datasets.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
2. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. IEEE Trans. Pattern Anal. Mach. Intell. **37**, 1670–1687 (2014)
3. Barrow, H., Tenenbaum, J., Hanson, A., Riseman, E.: Recovering intrinsic scene characteristics. Comput. Vision Syst. **2**, 2 (1978)
4. Baslamisli, A.S., Das, P., Le, H.A., Karaoglu, S., Gevers, T.: Shadingnet: Image intrinsics by fine-grained shading decomposition. Int. J. Comput. Vision **129**(8), 2445–2473 (2021)
5. Baslamisli, A.S., Groenestege, T.T., Das, P., Le, H.A., Karaoglu, S., Gevers, T.: Joint learning of intrinsic images and semantic segmentation. In: Eur. Conf. Comput. Vision. pp. 286–302 (2018)
6. Baslamisli, A.S., Le, H.A., Gevers, T.: Cnn based learning using reflection and retinex models for intrinsic image decomposition. In: IEEE Conf. Comput. Vision Pattern Recognit. pp. 6674–6683 (2018)

7. Baslamisli, A.S., Liu, Y., Karaoglu, S., Gevers, T.: Physics-based shading reconstruction for intrinsic image decomposition. Comput. Vision Image Understanding **205**, 103183 (2021)
8. Beigpour, S., Van De Weijer, J.: Object recoloring based on intrinsic image estimation. In: IEEE Int. Conf. Comput. Vision. pp. 327–334 (2011)
9. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Trans. Graph. **33**(4), 1–12 (2014)
10. Van den Bergh, M., Boix, X., Roig, G., De Capitani, B., Van Gool, L.: Seeds: Superpixels extracted via energy-driven sampling. In: IEEE Conf. Comput. Vision Pattern Recognit. pp. 13–26 (2012)
11. Blake, A.: Boundary conditions for lightness computation in mondrian world. Comput. Vision Graph. Image Process. **32**(3), 314–327 (1985)
12. Bonneel, N., Kovacs, B., Paris, S., Bala, K.: Intrinsic decompositions for image editing. Comput. Graph. Forum **36**, 593–609 (2017)
13. Bonneel, N., Sunkavalli, K., Tompkin, J., Sun, D., Paris, S., Pfister, H.: Interactive intrinsic video editing. ACM Trans. Graph. **33**(6), 1–10 (2014)
14. Bousseau, A., Paris, S., Durand, F.: User-assisted intrinsic images. In: ACM SIGGRAPH Asia 2009 papers. pp. 1–10 (2009)
15. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: Eur. Conf. Comput. Vision. pp. 611–625 (2012)
16. Chang, J., Cabezas, R., Fisher III, J.W.: Bayesian nonparametric intrinsic image decomposition. In: Eur. Conf. Comput. Vision. pp. 704–719 (2014)
17. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: IEEE Int. Conf. Comput. Vision. pp. 241–248 (2013)
18. Das, P., Gevers, M., Karaoglu, S., Gevers, T.: Idtransformer: Transformer for intrinsic image decomposition. In: IEEE/CVF Int. Conf. Comput. Vision. pp. 816–825 (2023)
19. Das, P., Karaoglu, S., Gevers, T.: Intrinsic image decomposition using physics-based cues and cnns. Comput. Vision and Image Understanding **223**, 103538 (2022)
20. Ding, S., Sheng, B., Hou, X., Xie, Z., Ma, L.: Intrinsic image decomposition using multi-scale measurements and sparsity. Comput. Graph. Forum **36**, 251–261 (2017)
21. Ebner, M.: Color Constancy, 1st ed. Wiley Publishing, ISBN: 0470058299, Hoboken, NJ, USA (2007)
22. Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic images by clustering. Comput. Graph. Forum **31**, 1415–1424 (2012)
23. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: Advances in neural information processing systems. pp. 765–773 (2011)
24. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: IEEE Int. Conf. Comput. Vision. pp. 2335–2342 (2009)
25. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2012)
26. Horn, B.: Robot vision. MIT press (1986)
27. Jin, X., Gu, Y.: Superpixel-based intrinsic image decomposition of hyperspectral images. IEEE Trans. Geosci. Remote Sens. **55**(8), 4285–4295 (2017)
28. Land, E.H., McCann, J.J.: Lightness and retinex theory. J. Opt. Soc. Amer. **61**(1), 1–11 (1971)
29. Lettry, L., Vanhoey, K., Van Gool, L.: Unsupervised deep single-image intrinsic decomposition using illumination-varying image sequences. Comput. Graph. Forum **37**, 409–419 (2018)
30. Li, Z., Snavely, N.: Cgintrinsics: Better intrinsic image decomposition through physically-based rendering. In: Eur. Conf. Comput. Vision. pp. 371–387 (2018)

31. Li, Z., Chen, J.: Superpixel segmentation using linear spectral clustering. In: IEEE Conf. Comput. Vision Pattern Recognit. pp. 1356–1363 (2015)
32. Liu, Y., Li, Y., You, S., Lu, F.: Unsupervised learning for intrinsic image decomposition from a single image. In: IEEE/CVF Conf. Comput. Vision Pattern Recognit. pp. 3248–3257 (2020)
33. Ma, Y., Jiang, X., Xia, Z., Gabbouj, M., Feng, X.: Casqnet: Intrinsic image decomposition based on cascaded quotient network. IEEE Trans. Circuits Syst. Video Technol. **31**(7), 2661–2674 (2020)
34. Narihira, T., Maire, M., Yu, S.X.: Learning lightness from human judgement on relative reflectance. In: IEEE Conf. Comput. Vision Pattern Recognit. pp. 2965–2973 (2015)
35. Qian, Y., Shi, M., Kamarainen, J.K., Matas, J.: Fast fourier intrinsic network. In: IEEE/CVF Winter Conf. Appl. Comput. Vision. pp. 3169–3178 (2021)
36. Ren, X., Yang, W., Cheng, W.H., Liu, J.: LR3M: Robust low-light enhancement via low-rank regularized retinex model. IEEE Trans. Image Process. **29**, 5862–5876 (2020)
37. Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. In: IEEE Comput. Vision Pattern Recognit. pp. 3481–3487 (2011)
38. Shen, L., Yeo, C., Hua, B.S.: Intrinsic image decomposition using a sparse representation of reflectance. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 2904–2915 (2013)
39. Shi, J., Dong, Y., Su, H., Yu, S.X.: Learning non-lambertian object intrinsics across shapenet categories. In: IEEE Conf. Comput. Vision Pattern Recognit. pp. 1685–1694 (2017)
40. Shi, J., Dong, Y., Tong, X., Chen, Y.: Efficient intrinsic image decomposition for rgbd images. In: ACM Symp. Virtual Reality Softw. Technol. pp. 17–25 (2015)
41. Tappen, M., Freeman, W., Adelson, E.: Recovering intrinsic images from a single image. Advances Neural Inf. Process. Syst. **15** (2002)
42. Ulucan, D., Ulucan, O., Ebner, M.: IID-NORD: A comprehensive intrinsic image decomposition dataset. In: IEEE Int. Conf. Image Process. pp. 2831–2835 (2022)
43. Ulucan, D., Ulucan, O., Ebner, M.: Intrinsic image decomposition: Challenges and new perspectives. In: Int. Conf. Image Process. Vision Eng. pp. 57–64. INSTICC (2023)
44. Ulucan, D., Ulucan, O., Ebner, M.: Multi-scale surface normal estimation from depth maps. In: Int. Conf. Image Process. Vision Eng. pp. 47–56 (2023)
45. Ulucan, O., Ulucan, D., Ebner, M.: Block-based color constancy: The deviation of salient pixels. In: IEEE Int. Conf. Acoust. Speech Signal Process. pp. 1–5 (2023)
46. Ulucan, O., Ulucan, D., Ebner, M.: Multi-scale color constancy based on salient varying local spatial statistics. The Vis. Comput. pp. 1–17 (2023)
47. Weiss, Y.: Deriving intrinsic images from image sequences. In: IEEE Int. Conf. Comput Vision. vol. 2, pp. 68–75 (2001)
48. Xu, C., Han, Y., Baciu, G., Li, M.: Fabric image recolorization based on intrinsic image decomposition. Textile Research J. **89**(17), 3617–3631 (2019)
49. Xu, J., Hou, Y., Ren, D., Liu, L., Zhu, F., Yu, M., Wang, H., Shao, L.: Star: A structure and texture aware retinex model. IEEE Trans. Image Process. **29**, 5022–5037 (2020)
50. Yu, Y., Smith, W.A.: Inverserendernet: Learning single image inverse rendering. In: IEEE/CVF Conf. Comput. Vision Pattern Recognit. pp. 3155–3164 (2019)
51. Yuan, Y., Sheng, B., Li, P., Bi, L., Kim, J., Wu, E.: Deep intrinsic image decomposition using joint parallel learning. In: Comput. Graph. Int. Conf. pp. 336–341 (2019)
52. Zhang, H., Ma, J.: IID-MEF: A multi-exposure fusion network based on intrinsic image decomposition. Inf. Fusion **95**, 326–340 (2023)
53. Zhao, Q., Tan, P., Dai, Q., Shen, L., Wu, E., Lin, S.: A closed-form solution to retinex with nonlocal texture constraints. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1437–1444 (2012)
54. Zhou, T., Krahenbuhl, P., Efros, A.A.: Learning data-driven reflectance priors for intrinsic image decomposition. In: IEEE Int. Conf. Comput. Vision. pp. 3469–3477 (2015)