

Challenges and Applications of Intrinsic Image Decomposition: A Short Review

Diclehan Ulucan^{1*}, Oguzhan Ulucan¹ and Marc Ebner¹

^{1*}Computer Science, University of Greifswald, Walther-Rathenau-Str., Greifswald, 17489, Mecklenburg-Vorpommern, Germany.

*Corresponding author(s). E-mail(s): diclehan.ulucan@uni-greifswald.de;
Contributing authors: oguzhan.ulucan@uni-greifswald.de; marc.ebner@uni-greifswald.de;

Abstract

Intrinsic image decomposition has become an immensely studied problem over the last decades. It holds many challenges but also provides large benefits if it can be solved. In this study, we provide a short review of intrinsic image decomposition algorithms, datasets, and applications, while also addressing the challenges of the field. Aside from creating an algorithm for this under-constrained problem, another challenge is to evaluate the performance of the developed methods since there are certain limitations in existing evaluation strategies. Thereupon, we introduce two new error metrics, namely the ensemble of metrics and the imperceptible weighted score. The ensemble of metrics integrates different perceptual quality metrics in scale-space, while the imperceptible weighted score is the modified version of the well-known ΔE metric. We present the usability of our metrics on two datasets by utilizing various intrinsic image decomposition algorithms.

Keywords: Intrinsic image decomposition, performance evaluation, error metrics, review

1 Introduction

Our visual system is able to differentiate between colors, discount the illuminant, and estimate distances unconsciously, while these abilities are difficult to perform for computer vision systems [1, 2]. The efficiency of machine vision applications might deteriorate in the presence of reflection, occlusion, glare effect, ambiguity caused by light at edges, over-saturated regions, and dark areas due to detail loss [3]. One way to increase the performance of machine systems is to use intrinsic image decomposition in the application pipeline to overcome challenges arising due to these issues.

Images can be decomposed into a "family of intrinsic characteristics" where each component is a low-level feature of the input and it is referred to as

an *intrinsic image* [4]. The reflectance, shading, shadows, depth, surface normals, and illuminant can be given as examples for intrinsic images. Each low-level feature enables us to determine different characteristics of a scene more precisely, e.g., we can perform object segmentation by using reflectance rather than the pixel RGB values [2].

While intrinsic image decomposition provides benefits to numerous computer vision applications, developing effective intrinsic image decomposition algorithms is troublesome due to the ill-posed nature of the problem which is why intrinsic image decomposition is framed as a computational challenge. Studies in this field generally relax this problem by assuming that the scene captured by a set of sensors can be formed by the reflectance and shading elements as

follows;

$$I(x, y) = R(x, y) \cdot S(x, y) \quad (1)$$

where, I is the image at the spatial location (x, y) , R represents the reflectance which provides the ratio between the total reflected and total incident illumination, and S presents the interaction between the illumination and the surfaces [4, 5].

The challenge in this field is not only the ill-posed nature of the problem but also the lack of common evaluation strategies, i.e., datasets and error metrics. Different datasets are used to evaluate the proposed algorithms, and the employed benchmarks tend to meet the assumptions made in the proposed methods, thus objectively identifying the best-performing algorithm is quite challenging [6]. Also, most of the datasets have different characteristics that make it difficult to evaluate the algorithms in a robust manner. For instance, some datasets contain single objects placed in front of a plain black background, others contain 3D models rendered with an environmental map where the object is positioned in the foreground and can be easily segmented, some benchmarks contain subjective ground truths, and some datasets are not specifically designed for intrinsic image decomposition but are used in this field [7–11]. Based on all of these observations, we recently created a large-scale intrinsic image decomposition dataset called IID-NORD by taking the shortcomings of existing benchmarks into account [3].

Not only the absence of a common benchmark but the lack of quality metrics reflecting the actual performance of the intrinsic image decomposition methods is a critical drawback [12]. Therefore, it is troublesome to report the algorithms’ performance in a robust and fair manner. Also, the fact that intrinsic image decomposition usually requires the evaluation of different intrinsics at once makes the assessment task even more challenging, i.e., an algorithm can extract reflectance much more precisely than shading. Therefore, we need a metric that can analyze different intrinsics by considering the individual characteristics of each intrinsic and that provides us a global quality score. Moreover, we need metrics that can analyze the intrinsics individually since a single intrinsic image can be used in computer vision applications, i.e., reflectance can be utilized for image segmentation.

The development of evaluation strategies is as important as the design of algorithms in the field of intrinsic image decomposition since evaluation methods allow us to determine the shortcomings and strengths of algorithms which helps us to create more

efficient intrinsic image decomposition approaches. Therefore, in our previous work [13], we discussed the challenges of intrinsic image decomposition and made an attempt to provide new perspectives in this field by introducing two evaluation strategies. In this study, we extend our previous work. In particular, we improve the *imperceptible ΔE* metric and introduce the *imperceptible weighted score*. Furthermore, we provide a more detailed review of intrinsic image decomposition algorithms and datasets, while we also discuss the intrinsic image decomposition applications. To the best of our knowledge, only two comprehensive surveys exist in the field of intrinsic image decomposition. In the study of Garces *et al.* [12], an extensive review of deep learning-based intrinsic image decomposition methods is provided, and these algorithms are investigated in detail. In the study of Bonneel *et al.* [6], a survey focusing on evaluating the intrinsic image decomposition algorithms in the field of image editing is provided. Also, in the same study, the authors summarized the typical applications of intrinsic image decomposition in image editing. In this study, we provide a different perspective compared to similar studies. While we cover a larger range of applications of intrinsic image decomposition from image fusion to license plate recognition, we also review the traditional and learning-based algorithms.

This paper is organized as follows. In Sec. 2 we provide a review of intrinsic image decomposition algorithms. In Sec. 3 we discuss the intrinsic image decomposition applications. In Sec. 4 we review the intrinsic image decomposition benchmarks. In Sec. 5 we detail the evaluation methods and introduce our error metrics. In Sec. 6 we provide our experimental results. Lastly, in Sec. 7 we give a brief summary of our work and discuss possible future directions.

2 Algorithms

Intrinsic image decomposition algorithms provide many beneficial cues for various computer vision pipelines ranging from object segmentation to image fusion [14]. Therefore, over the last decades, numerous algorithms based on traditional and data-dependent methods have been proposed in this field. These algorithms have various input requirements, i.e., a time-varying image stack, multiple images taken under different lights, multiple images with different viewing conditions, an input sequence where the light source is placed at distinct locations in each image,

different focal distances, depth information, or a single RGB image [6]. In this section, we provide a brief review of intrinsic image decomposition algorithms by grouping them into two categories as traditional algorithms and learning-based algorithms.

2.1 Traditional Algorithms

Over the last five decades, various traditional intrinsic image decomposition algorithms have been introduced [15–20]. One of the earliest studies is the biologically inspired Retinex algorithm [21]. The method is based on the observations that planar surfaces and shadows have smooth intensity differences, while adjacent regions of various objects have sharp reflectance changes since the difference between the intensities is large. Hence, it can be concluded that large gradient changes usually occur due to reflectance changes, while small gradients are related to the shading element. Over the years, the Retinex algorithm has been modified and exploited in various studies. For instance, it is combined with a non-local reflectance constraint, where it is assumed that two pixels having the same chromaticity texture vectors have the same reflectance [22]. In another method it is assumed that small patches should have similar reflectance, and an energy function is optimized with constraints assigning larger weights to the spatially local neighboring pixels [5]. In the cluster-based algorithm, pixels with a similar reflectance are clustered, and a model is formed that describes the connections and relations between these groups [23].

The SIRFS algorithm is designed to decompose an image containing a single masked object into several intrinsics by making use of a multi-scale optimization method relying on prior information [24]. In the generative and probabilistic algorithm, a Dirichlet process Gaussian mixture model is utilized together with Markov chain Monte Carlo sampling methods [25]. In another intrinsic image study, an RGB-D image is used to estimate the reflectance and shading by making use of an optimization method [26].

2.2 Learning-based Algorithms

After their efficiency is exploited in a wide range of applications in computer vision, both supervised and unsupervised learning-based strategies are utilized to provide a solution to the field of intrinsic image decomposition.

Most of the learning-based strategies in the field of intrinsic image decomposition do not build their

models on the well-established fundamentals of traditional image formation [27]. Hence, even if the performance of these data-dependent algorithms surpasses the traditional intrinsic image decomposition methods quantitatively, their performance is lacking qualitative investigations. From this motivation, Baslamisli *et al.* created a learning-based strategy by combining the ideas of the best of the two worlds. They investigated the capabilities of the proposed convolutional neural network framework which relies on a physics-based reflection model and utilizes the high-frequency components of both the reflectance and shading elements of a scene [27]. After the supervised models have proved their effectiveness in the field of intrinsic image decomposition, their usage in this domain gradually increased. They are utilized either to purely perform intrinsic image decomposition or to enable the usage of image decomposition as an intermediate step to improve the performance of many different computer vision pipelines [28–34]. Nevertheless, in several studies, researchers have stressed their concerns about developing supervised frameworks by stating that a limited number of datasets contain accurate intrinsic elements which is not surprising since obtaining ground truths for both the reflectance and shading is burdensome. From this motivation, several unsupervised models are created to tackle the ill-posed nature of intrinsic image decomposition and to estimate the intrinsics of the scenes [35–37].

Although obtaining the intrinsic elements from a single image is applicable in numerous computer vision applications, several learning-based studies guide their estimations by using additional inputs, such as the distance sensor measurements, i.e., depth maps, based on the motivation that these sensors are now widely present in many capturing devices. One of the earliest attempts to find the illumination intrinsic of the scene by incorporating a depth map is performed by Ebner to improve the performance of his learning-free computational color constancy method [38]. In more recent years, there have been several works in the field of intrinsic image decomposition that utilize distance measurements to guide their learning-based models to estimate a more accurate reflectance and shading for various applications [26, 39–42].

3 Applications

Different computer vision pipelines, and applications benefiting from image processing utilize intrinsic

image decomposition algorithms. For instance, the field of robotics is one of the areas where intrinsic images are widely used. In the study of Krajník *et al.*, a road-following method is designed for mobile robots that operate in outdoor environments [43]. The pathways are detected by using intrinsic images, and the robot is steered along these pathways. In the work of Strisciuglio *et al.*, a CNNs-based intrinsic image decomposition algorithm is used in the computer vision pipeline of a gardening robotics application [44]. The robot is designed for automatic bush trimming and rose pruning. In the study of Brandao *et al.*, it is aimed to predict the friction for robot locomotion, which is the mutual name for methods robots use to move from one place to another [45]. The Retinex algorithm is used to extract the shading element of the scene that is utilized to estimate the friction coefficient between surfaces and a robot's foot.

Apart from the field of robotics, the low-level features of images are used in applications related to security. In the study of Li *et al.*, intrinsic image decomposition is used in a license plate recognition algorithm. The reflectance element is utilized for license plate localization [46]. In the study of Tong *et al.*, the shadow effect in road scenes is weakened by using intrinsic images [47]. The developed algorithm is able to weaken the shadows without causing any color deviation and to provide promising results for road region extraction. In the work of Li *et al.*, a 3D face mask presentation attack detection algorithm is developed by utilizing the reflectance image, where the intensity variation features are extracted by using a 1D CNN model [48].

Intrinsic images are also used in classical image processing tasks such as classification, segmentation, and enhancement. In the work of Kang *et al.* [49], intrinsic image decomposition is used in a hyperspectral image classification pipeline where it removes the redundant information caused by the shading element of the hyperspectral image, i.e., intrinsic image decomposition enables the algorithm to carry out the pixel-wise classification by only using the reflectance image. In the study of Baslamisli *et al.* [30], a supervised end-to-end CNN model is introduced which explores the relationship between intrinsic images and semantic segmentation, and jointly estimates the intrinsic and semantic features of a scene. Hence, intrinsic image decomposition aids the task of semantic segmentation, while semantic features also help to estimate the shading and reflectance components. In the work of Ren *et al.* [50], which is a low-light

enhancement study, a Retinex-based intrinsic image decomposition method is used to estimate a piece-wise smooth illumination and a noise-suppressed reflectance for improving the visual quality of the image. In the study of Yue *et al.* [51], a contrast enhancement method based on intrinsic images is proposed. The decomposition is carried out in the V channel of the HSV color space where it is assumed that the reflectance is piece-wise constant and illumination is locally smooth.

Other applications of intrinsic images are object recoloring, and surface re-texturing. In the study of Beigpour and van de Weijer [52], intrinsic image decomposition is used for the recoloring of objects illuminated by colored and multiple lights. After an image is decomposed, the body reflectance is changed for object recoloring and the specular reflectance is changed for illuminant recoloring. In the study of Xu *et al.* [53], intrinsic images are used in a fabric image recolorization pipeline which aims at helping designers to create new color designs for fabric. The algorithm that integrates intrinsic image decomposition into a variational framework together with an image segmentation method is able to preserve yarn boundary details as well as texture details. In the study of Bi *et al.* [54] an intrinsic image decomposition method based on the L_1 norm is proposed which is utilized in surface re-texturing and 3D object compositing applications. For the former application, the reflectance image is edited and then its product with the shading element is computed to avoid unrealistic and flat outcomes, while for the latter application the spatially varying illumination in the shading element is estimated via a simplified version of the SIRFS algorithm [24] and then a 3D object is inserted into the image, i.e. consistency of the illumination conditions in the scene is ensured [54].

Lastly, intrinsic image decomposition can also be utilized in image fusion applications. In the medical imaging study of Du *et al.* [55], magnetic resonance imaging (MRI) and positron emission tomography (PET) data are fused by utilizing intrinsic image decomposition. In the pipeline, the illumination and reflectance are extracted from the MRI image, while the PET input is separated as normal image and lesion image. These images are then used to fuse the MRI and PET data. In the study of Zhang and Ma [14], a multi-exposure image fusion model is developed that benefits from intrinsic image decomposition. The model extracts the reflectance and shading images

from the input sequence which are later fused individually. The model is also able to perform low-light image enhancement and overexposed image correction. In the work of Kang *et al.* [56], two satellite images, the high-resolution panchromatic image, and the low-resolution multi-spectral image are fused by using intrinsic image decomposition. The aim of fusion is to obtain a high-resolution multi-spectral image where the illumination is considered as the panchromatic image.

4 Datasets

In the field of intrinsic image decomposition, there are a few publicly available datasets that have different characteristics (Fig. 1). The first dataset that contains explicit labels for the reflectance and shading images is the MIT Intrinsic Images Dataset and it is formed by Grosse *et al.* in 2009 [7]. The benchmark contains 220 images based on scenes created with 20 real objects. In each scene, a single object is positioned in front of a plain black background which is captured under different lighting conditions. The dataset contains the reflectance and shading components, the binary mask, the diffuse component, and the specular information of the image. The MPI Sintel Flow Dataset is introduced by Butler *et al.*, and it is in fact created for optical flow evaluation but since it also provides the reflectance image it is used in intrinsic image decomposition as well [57]. The dataset contains a limited number of images that are extracted from a 3D fantasy short-film called Sintel. The Intrinsic Images in the Wild Dataset is introduced by Bell *et al.*, and it provides the ground images for the reflectance image of real-world scenes [8]. This large-scale dataset is formed with the help of human operators. The Multi-illuminant Intrinsic Image Dataset is created by Beigpour *et al.*, and it consists of real photographs of 5 different scenes captured under multi-illuminant and multi-colored illumination conditions [58]. The scenes contain a few objects placed in front of a plain black background and it is a rather small-scale dataset. The Multi-view Multi-illuminant Intrinsic Dataset is formed by Beigpour *et al.*, and it consists of 600 high-resolution images [59]. The scenes contain either plain black or a partial background. The SUNCG dataset is proposed by Song *et al.* and includes a large number of manually created indoor scenes [60]. This dataset is also used in other studies as a baseline and it is modified to add new intrinsics, i.e., surface normals, and to add more material models for various usage



Fig. 1: Example scenes from datasets. (Left-to-right) The MIT Intrinsic Images Dataset [7], the MPI Sintel Flow Dataset [57], the Intrinsic Images in the Wild Dataset [8], and the IID-NORD Dataset [3].

purposes, i.e., semantic and glossiness segmentation, inverse rendering [61, 62]. The CGIntrinsics dataset is formed by Li and Snavely in 2018 to assist the researchers aiming to solve the problem of intrinsic elements in Internet images of real-world scenes [63]. The dataset is generated by taking the textures and the models of the indoor scenes of the SUNCG dataset. The CGIntrinsics benchmark contains 20k high-quality rendered scenes via the Mitsuba Renderer. The authors also provide the reflectance elements of each scene. Recently, we created a large-scale dataset called IID-NORD by using computer graphics [3]. The dataset consists of indoor scenes with various layouts. A high number of distinct illuminants, textures, and 3D objects are used to increase the variety of the dataset.

5 Evaluation Metrics

There are two evaluation methods that have been specifically designed for the task of intrinsic image decomposition, namely the local mean squared error (LMSE) [7] and the weighted human disagreement rate (WHDR) [8]. While the former is an objective error metric, the latter is based on human judgments. Since in this study, we focus on the objective evaluation techniques, we do not further discuss WHDR for which the reader may refer to the work of Bell *et al.* [8].

LMSE is widely used in the field of intrinsic image decomposition to evaluate the algorithms. LMSE is based on the classical mean squared error and it is computed by averaging MSE over overlapping patches. LMSE has difficulties in presenting the actual performance of an algorithm [6, 12], i.e., the available information in the extracted intrinsic images and the LMSE score have a tendency not to coincide. Also, when large areas with constant reflectance are decomposed correctly, often a very low LMSE score is obtained irrespective of the remaining parts of the image [6].

In several intrinsic image decomposition studies the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) are utilized alongside LMSE to benchmark the algorithms [26, 64]. PSNR computes the peak signal-to-noise ratio in decibels (dB) between the ground truth and processed images [65]. Due to pixel-wise calculations, PSNR does not consider the neighboring relationships of the image elements, hence the scores do not necessarily reflect the available information. On the other hand, SSIM is inspired by the top-to-bottom assumption of the human visual system and it is based on patch-wise computations, thus it takes the local spatial information into consideration [66]. SSIM is a perceptual quality metric that regards the structure, contrast, and luminance elements of the images to analyze the structural similarity between the ground truth and processed images. SSIM outputs a score in the range $[0, 1]$, where a score closer to 1 refers to a better result. SSIM can be computed as follows [66];

$$SSIM = \frac{(2\mu_{I_1}\mu_{I_2} + C_1)(2\sigma_{I_1I_2} + C_2)}{(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1)(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2)} \quad (2)$$

where, I_1 and I_2 represent the ground truth and output image, μ , σ , and σ^2 are the mean, covariance, and variance, respectively, while C_1 and C_2 are small constants. It is worth noting that to prevent any problems related to viewing conditions, later on, SSIM is modified and the multi-scale SSIM (MS-SSIM) is introduced [67].

Rarely, the correlation (CORR) between the estimated intrinsics and the ground truths is used in intrinsic image decomposition studies to analyze the performance of algorithms [68]. Correlation measures how similar two images are in terms of intensity. A high correlation score indicates that the estimated intrinsic image closely matches the ground truth, while a low correlation score demonstrates that there are discrepancies between the estimated intrinsic and the ground truth [69].

While these metrics exist for the analysis of intrinsic images, there is a need for developing and considering different error metrics to better evaluate the outcomes of the intrinsic image decomposition algorithms as explained in Sec. 1. Thereupon, we discuss three existing metrics that are used in various image processing tasks but to the best of our knowledge have not been utilized in the field of intrinsic image decomposition yet.

The visual information fidelity (VIF) computes a similarity score by measuring how much of the information that can be extracted from the ground truth image can also be derived from the output image [70]. VIF is calculated in the wavelet domain by utilizing Gaussian scale mixtures C , which are a random field that can be presented as the product of two independent random fields. VIF outputs scores in the range $[0, 1]$, where scores closer to 1 indicate better results. The VIF score can be calculated as follows;

$$VIF = \frac{\sum_{k \in w} \mathcal{S}_{\mathcal{RF}}(\vec{C}^{T,k}, \vec{F}^{T,k} |_{s^{T,k}})}{\sum_{k \in w} \mathcal{S}_{\mathcal{RF}}(\vec{C}^{T,k}, \vec{E}^{T,k} |_{s^{T,k}})} \quad (3)$$

where, $\mathcal{S}_{\mathcal{RF}}$ represents the set of spatial locations for the random field, w are the subbands of the image, $\vec{C}^{T,k}$ presents T elements of C_k , $\vec{F}^{T,k}$ and $\vec{E}^{T,k}$ denote the T elements of the output image and ground truth images in one subband, respectively, and s^T are the model parameters of the associated image.

The feature similarity index (FSIM) takes the local structures and contrast information of the images into account [71]. FSIM is calculated by utilizing the phase congruency (PC) which is contrast invariant, and the gradient magnitude (GM). The PC element presumes that points having a maximal phase in the frequency domain correspond to perceivable features. This assumption correlates with the human visual system's behavior while it detects significant features in images. GM is added during the computation of FSIM to consider the contrast information of a scene since the contrast information affects the human visual system during perception. FSIM can be calculated as follows:

$$FSIM = \frac{\sum_{x,y \in N} (F_{PC}(x,y) \cdot F_{GM}(x,y)) \cdot PC_{max}(x,y)}{\sum_{x,y \in N} PC_{max}(x,y)} \quad (4)$$

where, $F_{PC}(x,y)$ and $F_{GM}(x,y)$ indicate the PC and GM components of the image, respectively, $PC_{max}(x,y)$ is the maximum PC value of the input images, and N is the number of pixels in the image. Similar to SSIM and VIF, FSIM presents scores in the range $[0, 1]$ and results closer to 1 correspond to superior results. Note that FSIM is calculated for grayscale images but it has a straightforward extension to RGB images.

The ΔE (CIEDE2000) computes the color difference between two CIELAB samples, and in case we calculate the color difference between two images in CIELAB color domain, we average the results of the individual samples [72, 73]. ΔE can be calculated

by investigating the lightness, chroma, and hue components. It is known that ΔE scores less than 1 are unnoticeable to human observers, while a score in the range $[1, 4)$ might also be imperceptible [2, 74].

In the following, we utilize the SSIM, FSIM, and VIF metrics to introduce an "ensemble of metrics", and we modify the classical ΔE metric to form the "imperceptible weighted score".

5.1 Proposed Metrics

We form our metrics by considering two observations: (i) several studies suggest that evaluation methods correlating with the human visual system have a tendency of computing more reliable scores [75–77], and (ii) it is widely known that computations carried out in scale-space help us to avoid problems arising due to unknown display resolution and viewing distance. Based on these observations we introduce the ensemble of metrics (EM) which aims at analyzing the reflectance and shading components together in a robust manner, and the imperceptible weighted score (IWS) which focuses on the evaluation of the reflectance component.

5.1.1 Ensemble of Metrics

The ensemble of metrics utilizes the SSIM, FSIM, and VIF metrics in scale-space to evaluate the outcomes of intrinsic image decomposition algorithms. We have chosen these metrics for the ensemble since they focus on features such as color, structure, contrast, luminance, and the amount of information coinciding between the ground truth and the estimated intrinsic image which are important for the results of intrinsic image decomposition methods.

For the ensemble of metrics, first we compute the Gaussian and Laplacian pyramids of the input image and the estimation, where we determine the number of levels adaptively based on the image resolution as follows:

$$\mathcal{L} = \lfloor \log(\min(h, w)) / \log(2) \rfloor \quad (5)$$

where, h and w are the width and height of the image.

We utilize both of the pyramids since they highlight different features of an image [78]. While the Gaussian pyramid preserves the low-frequency components of the image, i.e., color information, the Laplacian pyramid functions like a high-pass filter and contains the high-frequency elements of an image, i.e., fine details. The usage of both pyramids allows us

to take into account various details at distinct scales, thus we can investigate the outcomes of algorithms in greater detail. Moreover, by considering the high- and low-frequency separately, we can analyze the results of intrinsic image decomposition algorithms with metrics that are better suited to investigate certain features appearing more explicitly in one pyramid compared to the other.

We calculate the SSIM, VIF, and FSIM scores at every level in the Gaussian and Laplacian pyramids, however, we do not use all of them for evaluating both the reflectance and shading components. We utilize all three metrics for the evaluation of the shading element, whereas we only use the SSIM and the colored FSIM (FSIMc) for the analysis of the reflectance. We discard the VIF score for the reflectance since it is calculated in the luminance channel of images, e.g., the analysis of the reflectance and shading results in the same outcome. Furthermore, in the Gaussian scale-space, we compute the SSIM score by utilizing each feature given in its standard formulation, whereas we do not consider the luminance component in the Laplacian pyramid since it is irrelevant. On the other hand, we regard all components of the FSIM and VIF in both pyramids since they are sensitive to the information in the pyramids. After we compute the scores at each level of both pyramids, we linearly combine the corresponding scales for the three metrics separately as in the following:

$$P_{M'}^R(i) = \frac{G_{M'}^R(i) + L_{M'}^R(i)}{2} \quad (6)$$

$$P_M^S(i) = \frac{G_M^S(i) + L_M^S(i)}{2} \quad (7)$$

where, P is the mean of the Gaussian (G) and Laplacian (L) pyramids, R and S are the reflectance and shading elements, respectively, i is the scale, $M' \in \{SSIM, FSIMc\}$ and $M \in \{SSIM, VIF, FSIM\}$.

Since each P contains scores obtained at distinct scales, we have to fuse these scores into an overall score for every metric. To merge the scores, we take inspiration from the study of Wang *et al.* where the MS-SSIM metric is designed based on human observations [67]. During the experiments of Wang *et al.*, it is realized that the human visual system gives different importance to the same error at distinct scales. In other words, even when all the images at various levels in the pyramid have the same statistical error, the

perceived quality changes in each scale. The assigned importance is approximately Gaussian, i.e., we give more importance to the middle scales of the pyramid. Therefore, in the ensemble of metrics, we merge the scores at different scales by utilizing a Gaussian-based weighting strategy to assign distinct weights to each level. In other words, we weight the $P_{M'}^R$ and P_M^S scores in a Gaussian manner with a standard deviation σ of $(L - 1)/5$ as follows:

$$P_{M'}^R = \sum_i P_{M'}^R(i - \mathcal{L}/2) e^{-\frac{i^2}{2\sigma^2}} \quad (8)$$

$$P_M^S = \sum_i P_M^S(i - \mathcal{L}/2) e^{-\frac{i^2}{2\sigma^2}}. \quad (9)$$

Afterwards, we linearly combine the scores for the reflectance and shading as follows:

$$EM_R = \frac{1}{2} \sum_j P_{M'(j)}^R \quad (10)$$

$$EM_S = \frac{1}{3} \sum_j P_{M(j)}^S \quad (11)$$

where, subscript j is the j^{th} element of M' and M , and EM_R and EM_S represent the ensemble of metrics scores for the reflectance and shading elements, respectively.

Lastly, we average EM_R and EM_S to compute a global EM score. It is worth mentioning that while we compute the FSIM, we discard the coarsest scale of the pyramids since FSIM also utilizes the scale-space in its calculations which causes ambiguities in the coarsest scale of the pyramids in EM.

5.1.2 Imperceptible Weighted Score

As we discussed in Sec. 1, in cases where a particular intrinsic image needs to be analyzed in an application, i.e., usually only the reflectance is utilized for image segmentation tasks, an evaluation method for the given intrinsic is beneficial. Based on this fact, apart from EM, we introduce the imperceptible weighted score that is designed to evaluate the estimated reflectance. This score is built upon the classical ΔE (CIEDE2000) metric which focuses on the color difference of two CIELAB samples, and can be used to compute the color discrepancy between images [72, 73]. ΔE is calculated in the CIELAB color domain by considering the chroma, lightness, and hue components. It is known that ΔE scores less than 1 are imperceptible, while outcomes in the range $[1, 4)$ may also be unnoticeable to human observers [2, 79]. In our

previous study, we took these findings into consideration and modified the conventional ΔE metric [13]. We computed the ΔE score at each scale of the Gaussian pyramid since color is a low-frequency feature of images. Afterwards, we counted the number of pixels having a ΔE score in the range $[0, 4)$ at each level separately. Then, we divided the number of pixels having an unnoticeable ΔE score by the total number of pixels in the corresponding level. Subsequently, we weighted these ratios with a Gaussian function as in the EM score and summed them to find the imperceptible ΔE score.

In this study, we further modify the imperceptible ΔE score to overcome an observed shortcoming which is described in the following. Let us assume that we have two images with the same resolution, and each of them has n number of pixels with a ΔE score less than 4. In this case, even if the remaining pixels would have significantly different ΔE scores, we would obtain the same imperceptible ΔE for both images since we did not take the pixels having a ΔE greater than or equal to 4 into account. In this study, we modify our proposed metric and introduce the imperceptible weighted score to provide a simple solution to this problem, which is where the imperceptible weighted score differentiates from its previous version introduced in our recent study [13]. After we count the number of pixels having a ΔE score in the range $[0, 4)$ at a certain level (s_n), we also count the number of pixels having a ΔE score greater or equal to 4 (s'_n) at the corresponding scale. Then, we form a penalty term (ϵ) by taking the ratio of s_n and s'_n , and multiply it with the mean ΔE score of the pixels having a perceivable ΔE score (ΔE_p) as follows $\epsilon = \frac{s_n}{s'_n} \Delta E_p$.

Subsequently, we add the penalty term ϵ to the mean ΔE score of the pixels having an imperceptible ΔE score. We carry out this computation at each level of the pyramid. Then, we weight the scores with in a Gaussian manner as in the EM score as follows:

$$IWS_{s_n} = (i - \mathcal{L}/2) e^{-\frac{i^2}{2\sigma^2}} (\text{mean}(\Delta E_{s_n}) + \epsilon_{s_n}). \quad (12)$$

Finally, we take the mean of all scores to find IWS as follows:

$$IWS = \frac{1}{i} \sum_i IWS_{s_i}, \quad (13)$$

where results closer to 0 indicate that the estimated reflectance is approximating the ground truth.

6 Experiments

In our experiments, we focus on demonstrating the usability of our proposed metrics and metrics that have not been considered in the field of intrinsic image decomposition. We would like to note that we do not benchmark intrinsic image decomposition algorithms on various datasets to present their efficiency since such experiments have already been carried out in several studies in detail. The reader may refer to the following studies Bonneel *et al.* [6], Ulucan *et al.* [3], and Garces *et al.* [12] for further information.

6.1 Experimental Setup

As mentioned in Sec. 2, intrinsic image decomposition methods have distinct input requirements. In the experiments, we utilize intrinsic image decomposition algorithms that require only a single RGB input image as input. We preferred to use these methods due to various reasons explained in the following. First of all, not all intrinsic image decomposition benchmarks contain image sequences, while real-world single images are widely available. Secondly, it is laborious to create image sequences in a proper format. Thirdly, for applications utilizing intrinsic image decomposition in their pipelines, input stacks may not be available. Fourthly, depth information is not always available for images. Lastly, requiring user interaction might be inefficient. In other words, single RGB input images are easy to access and they reflect the requisites of different computer vision applications. Consequently, while selecting the methods we use in our experiments, we choose among the algorithms that require a single RGB input image the ones that rely on different approaches such as utilizing local spatial information, being based on convolutional neural networks, relying on optimization methods, and using intrinsic image decomposition in another image processing task. By choosing algorithms with distinct approaches we increase the variety of the investigation we conduct in this study. We select the following algorithms for our experiments; Retinex [7, 21], Zhao [22], Shen [5], SIRFS [24], Lettry [35], and Ren [50], whose implementations we acquire from the official webpages of the authors and utilize without any optimization. Apart from these algorithms, we decompose all images also by a *baseline method* that is introduced in the study of Bonneel *et al.* [6]. This method is a simple technique that extracts the reflectance and shading images without considering any important aspect of

the intrinsic image decomposition problem. Therefore, any method particularly designed to compute intrinsic images is expected to outperform the baseline algorithm which assumes that the chromaticity image (I_{ch}) is the reflectance and the square root of the direct average of channels (Y), e.g., grayscale illumination, is the shading. I_{ch} and Y can be obtained as in the following;

$$I_{ch} = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B}, \frac{B}{R+G+B} \right) \quad (14)$$

$$Y = \sqrt{\frac{R+G+B}{3}} \quad (15)$$

where, R , G and B represent the color channels of the image.

In addition to the algorithms, we select 2 datasets with different characteristics, i.e., containing single-masked images and complex scenes, for our experiments. We benchmark the algorithms on a subset of the IID-NORD Dataset [3], and the well-known MIT Intrinsic Images Dataset [7] by using an Intel i7 CPU @ 3.5 GHz Quad-Core 16 GB RAM machine.

In order to statistically analyze the algorithms, we use the following 10 metrics; LMSE, PSNR, SSIM, MS-SSIM, FSIM, FSIMc, VIF, CORR, EM, and ΔE_i . By considering a wide range of metrics we aim at emphasizing the need of considering different metrics in the field of intrinsic image decomposition. The employment of various metrics may lead to the enhancement of algorithms since taking distinct evaluation strategies with different characteristics into account allows us to notice various shortcomings in the designed algorithms.

6.2 Experimental Results

The statistical results of the algorithms on both benchmarks are provided in Table 1. On the MIT Intrinsic Images dataset, all metrics except for the LMSE highlight SIRFS as the best-performing algorithm which also coincides with the visual investigation provided in Fig. 2. As mentioned in Sec. 5, LMSE discards local spatial cues, thus it may not reflect the actual performance of intrinsic image decomposition methods, while evaluation strategies such as FSIMc and FSIM take the local structures into account, hence regard the local spatial information. For instance, in Fig. 2 we provide random examples from the datasets. It is observable that the decomposition of Zhao, which is

Table 1: The statistical results of algorithms. IWS is given in 10^{-3} . CORR is normalized into the range $[0, 1]$. The best scores for each metric are highlighted. The last column gives the mean run time in seconds. The run time of Ren is not provided since its code is binary and does not only output the intrinsic images. A fixed threshold is used for Retinex. Also, SIRFS is not evaluated on IID-NORD, since it only works on inputs with single-masked objects.

		LMSE	PSNR	CORR	SSIM	MS-SSIM	FSIM	FSIMc	VIF	EM	IWS	Run time
MIT	Baseline	0.078	10.924	0.924	0.718	0.697	0.786	0.859	0.223	0.618	0.0024	0.031
	Retinex	0.091	11.128	0.937	0.726	0.720	0.795	0.882	0.179	0.631	0.0017	3.106
	Zhao	0.036	12.156	0.951	0.785	0.780	0.815	0.908	0.363	0.692	0.0019	3.671
	Shen	0.062	13.753	0.949	0.698	0.756	0.725	0.864	0.298	0.709	0.0014	38.097
	SIRFS	0.042	13.812	0.959	0.803	0.797	0.833	0.911	0.377	0.724	0.0011	171.018
	Lettry	0.056	12.275	0.944	0.527	0.722	0.706	0.873	0.122	0.639	0.0018	13.833
	Ren	0.079	9.233	0.932	0.703	0.713	0.816	0.869	0.176	0.605	0.0035	—
IID-NORD	Baseline	0.093	10.030	0.785	0.599	0.636	0.776	0.611	0.391	0.622	12.8850	0.242
	Retinex	0.101	10.581	0.805	0.658	0.688	0.786	0.677	0.386	0.652	2.2415	53.634
	Zhao	0.102	6.257	0.659	0.305	0.633	0.772	0.606	0.117	0.441	1.9746	40.219
	Shen	0.068	11.457	0.789	0.575	0.612	0.708	0.681	0.313	0.612	3.2274	390.300
	Lettry	0.097	12.093	0.788	0.621	0.710	0.766	0.702	0.235	0.643	2.6300	220.351
	Ren	0.094	9.562	0.784	0.505	0.670	0.738	0.653	0.121	0.585	1.4802	—

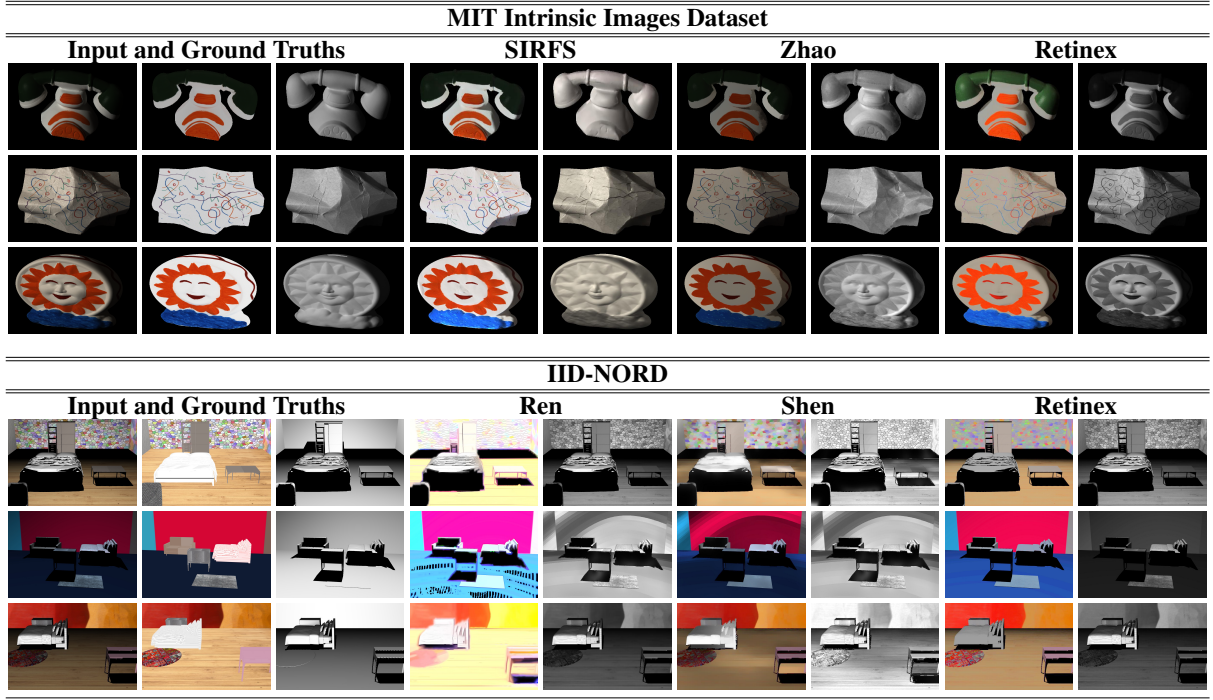


Fig. 2: Visual comparison of algorithms.

the best-performing method in terms of LMSE, has ambiguities, especially in the reflectance component,

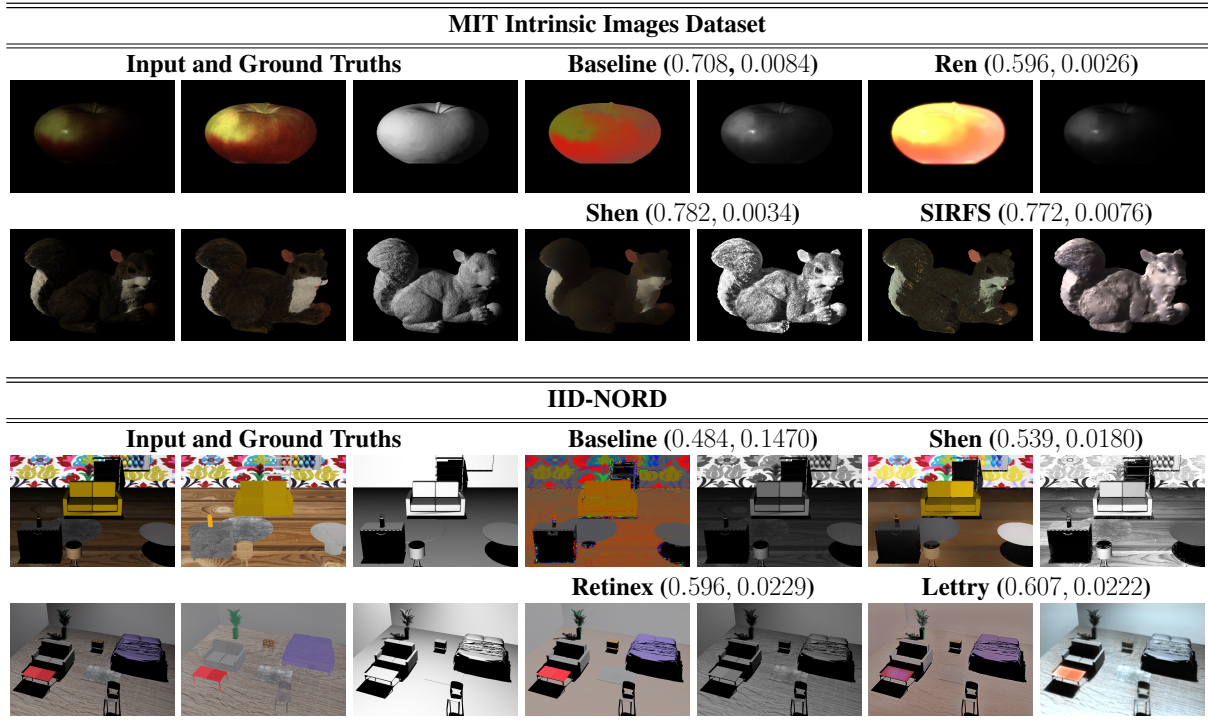


Fig. 3: Comparison of the methods on both benchmarks. The first score given in parenthesis corresponds to EM, while the second score is IWS.

while SIRFS decomposes the reflectance component more accurately which improves its IWS and EM scores on average.

On the IID-NORD dataset, the algorithms' efficiency decreases when they decompose images containing complex textures and strong shadow casts. Almost all metrics highlight a different algorithm as the best-performing method (Table 1). According to the LMSE scores Shen outperforms the other methods but as provided in Fig. 2, it may face difficulties in extracting the shading which is reflected in its EM score and it may output a reflectance which contains ambiguities that affect its average IWS. The reason behind LMSE pointing out Shen as the best-performing method can be explained by the fact that when large regions of constant reflectance are preserved during decomposition, LMSE tends to output low scores [6]. In terms of PSNR, MS-SSIM, and FSIMc, Lettry outputs more accurate intrinsics than the other algorithms. However, the visual results given in Fig. 3 present that Lettry may provide color-distorted intrinsic images which negatively affects its EM score and IWS. According to the CORR, SSIM, FSIM, and EM scores, Retinex surpasses the other

algorithms, while the best IWS is obtained by Ren. In the random examples provided in Fig. 2, we can observe that each method decomposes distinct regions of the images accurately, and handles image features such as strong shadow casts, highlights, and specularities with different accuracy. Hence, it is not surprising that the metrics highlight different algorithms as the best-performing method.

We present further results in Fig 3 where we also provide the EM scores and IWS of the images. The proposed evaluation methods produce statistical results coinciding with the perceptually available information with the intrinsics. If for a certain image, the EM score and IWS highlight different algorithms as the best-performing method, it can be deduced that while one of the intrinsic images is accurately estimated, there is an ambiguity in the estimation of the other one.

When we consider all of these investigations, we can infer that it is important to take into account the individual characteristics of the intrinsic images and weigh the outputs in a balanced manner to provide a reliable statistical score. Moreover, when we consider the fact that the results of all evaluation

strategies may not coincide with the actually available information in the output image [80], we can deduce that it is important to consider different metrics while evaluating an algorithm. For instance, benchmarking a method designed for a computer vision pipeline that utilizes the color features with LMSE or guiding a learning-based intrinsic image decomposition model by using LMSE as the loss function may not reflect the actual performance of the algorithm if large regions of constant reflectance are decomposed accurately. Therefore, different metrics such as IWS may be included to evaluate the outcomes or train the models for the task at hand.

As a final note, since benchmarking the algorithms by using metrics correlating with the human visual system mostly provides more accurate results, investigating the visual outputs of the algorithms together with their statistical outcomes may help us to understand which type of error metrics provide more reliable statistical results. Consequently, we might design more accurate benchmarking strategies in the field of intrinsic image decomposition.

7 Conclusion

Intrinsic images are the low-level features of input scenes, and they can be utilized in various applications ranging from robotics to object recoloring. The benefits it provides make intrinsic image decomposition an attractive research field, while the ill-posed nature of the problem challenges the researchers. Due to both the advantages it provides and the challenges it holds, the field of intrinsic image decomposition has been extensively studied over the last decades. In this study, we addressed the challenges in this field and provided an overview of the intrinsic image decomposition algorithms, datasets, and applications. Furthermore, we discussed the shortcomings of the existing evaluation strategies and provided new perspectives by introducing two new error metrics, namely the ensemble of metrics and the imperceptible weighted score, that are based on biological findings. We would like to mention that to further prove and analyze the proposed metrics it would be helpful to conduct experiments with human subjects. In such an experiment, the participants could rate the outcomes of algorithms and we could investigate if the ratings are correlating with the scores of the metrics.

Data availability. The datasets analyzed in this study are publicly available.

Declarations

Conflict of interest. The authors state that there are no conflicts of interest.

References

- [1] Zeki, S.: A Vision of the Brain. Blackwell Science, ISBN: 0632030545, Oxford, UK (1993)
- [2] Ebner, M.: Color Constancy, 1st Ed. Wiley Publishing, ISBN: 0470058299, Hoboken, NJ, USA (2007)
- [3] Ulucan, D., Ulucan, O., Ebner, M.: IID-NORD: A comprehensive intrinsic image decomposition dataset. In: Int. Conf. Image Process., Bordeaux, France, pp. 2831–2835 (2022). IEEE
- [4] Barrow, H., Tenenbaum, J., Hanson, A., Riesenman, E.: Recovering intrinsic scene characteristics. *Comput. Vision Syst.* **2**, 2 (1978)
- [5] Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. In: *Comput. Vision Pattern Recognit.*, Colorado Springs, CO, USA, pp. 3481–3487 (2011). IEEE
- [6] Bonneel, N., Kovacs, B., Paris, S., Bala, K.: Intrinsic decompositions for image editing. *Comput. Graph. Forum* **36**, 593–609 (2017)
- [7] Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: *Int. Conf. Comput. Vision*, Kyoto, Japan, pp. 2335–2342 (2009). IEEE
- [8] Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. *ACM Trans. Graph.* **33**(4), 1–12 (2014)
- [9] Shi, J., Dong, Y., Su, H., Yu, S.X.: Learning non-lambertian object intrinsics across shapenet categories. In: *Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, pp. 1685–1694 (2017). IEEE
- [10] Li, Z., Yu, T.-W., Sang, S., Wang, S., Song, M., Liu, Y., Yeh, Y.-Y., Zhu, R., Gundavarapu, N., Shi, J., Bi, S., Yu, H.-X., Xu, Z., Sunkavalli, K., Hasan, M., Ramamoorthi, R., Chandraker,

- M.: Openrooms: An open framework for photorealistic indoor scene datasets. In: *Comput. Vision Pattern Recognit.*, Nashville, TN, USA, pp. 7190–7199 (2021). IEEE
- [11] Roberts, M., Ramapuram, J., Ranjan, A., Kumar, A., Bautista, M.A., Paczan, N., Webb, R., Susskind, J.M.: Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In: *Int. Conf. Comput. Vision*, Montreal, QC, Canada, pp. 10912–10922 (2021). IEEE
- [12] Garces, E., Rodriguez-Pardo, C., Casas, D., Lopez-Moreno, J.: A survey on intrinsic images: Delving deep into lambert and beyond. *Int. J. Comput. Vision* **130**, 836–868 (2022)
- [13] Ulucan, D., Ulucan, O., Ebner, M.: Intrinsic image decomposition: Challenges and new perspectives. In: *Int. Conf. Image Process. Vision Eng.*, Prague, Czech Republic, pp. 57–64 (2023). INSTICC
- [14] Zhang, H., Ma, J.: IID-MEF: A multi-exposure fusion network based on intrinsic image decomposition. *Inf. Fusion* **95**, 326–340 (2023)
- [15] Weiss, Y.: Deriving intrinsic images from image sequences. In: *Int. Conf. Comput. Vision*, vol. 2. Vancouver, BC, Canada, pp. 68–75 (2001). IEEE
- [16] Tappen, M., Freeman, W., Adelson, E.: Recovering intrinsic images from a single image. *Advances Neural Inf. Process. Syst.* **15** (2002)
- [17] Finlayson, G.D., Drew, M.S., Lu, C.: Intrinsic images by entropy minimization. In: *European Conf. Comput. Vision*, Prague, Czech Republic, pp. 582–595 (2004). Springer
- [18] Bousseau, S. A. and Paris, Durand, F.: User-assisted intrinsic images. In: *SIGGRAPH Asia*, Yokohama, Japan, pp. 1–10 (2009)
- [19] Laffont, P.-Y., Bousseau, A., Drettakis, G.: Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Trans. Visualization Comput. Graph.* **19**(2), 210–224 (2012)
- [20] Pan, S., An, X., He, H.: Intrinsic image decomposition from a single image via non-linear anisotropic diffusion. In: *Int. Conf. Inf. Automat.*, Yinchuan, China, pp. 179–184 (2013). IEEE
- [21] Land, E.H.: The retinex. *Amer. Scientist* **52**, 247–264 (1964)
- [22] 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)
- [23] Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic images by clustering. *Comput. Graph. Forum* **31**, 1415–1424 (2012)
- [24] Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1670–1687 (2014)
- [25] Chang, J., Cabezas, R., Fisher III, J.W.: Bayesian nonparametric intrinsic image decomposition. In: *European Conf. Comput. Vision*, Zurich, Switzerland, pp. 704–719 (2014). Springer
- [26] Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: *Int. Conf. Comput. Vision*, Sydney, NSW, Australia, pp. 241–248 (2013). IEEE
- [27] Baslamisli, A.S., Le, H.-A., Gevers, T.: CNN based learning using reflection and Retinex models for intrinsic image decomposition. In: *Conf. Comput. Vision Pattern Recognit.*, Salt Lake City, UT, USA, pp. 6674–6683 (2018). IEEE/CVF
- [28] Janner, M., Wu, J., Kulkarni, T.D., Yildirim, I., Tenenbaum, J.: Self-supervised intrinsic image decomposition. *Advances Neural Inf. Process. Syst.* **30** (2017)
- [29] Fan, Q., Yang, J., Hua, G., Chen, B., Wipf, D.: Revisiting deep intrinsic image decompositions. In: *Conf. Comput. Vision Pattern Recognit.*, Salt Lake City, UT, USA, pp. 8944–8952 (2018). IEEE/CVF
- [30] Baslamisli, A.S., Groenesteghe, T.T., Das, P., Le, H.-A., Karaoglu, S., Gevers, T.: Joint learning of intrinsic images and semantic segmentation.

- In: European Conf. Comput. Vision, Munich, Germany, pp. 286–302 (2018). Springer
- [31] 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)
 - [32] Le, H., Samaras, D.: Physics-based shadow image decomposition for shadow removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(12), 9088–9101 (2021)
 - [33] Das, P., Karaoglu, S., Gevers, T.: Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In: *Conf. Comput. Vision Pattern Recognit.*, New Orleans, LA, USA, pp. 19790–19799 (2022). IEEE/CVF
 - [34] Yi, R., Zhu, C., Xu, K.: Weakly-supervised single-view image relighting. In: *Conf. Comput. Vision Pattern Recognit.*, Vancouver, Canada, pp. 8402–8411 (2023). IEEE/CVF
 - [35] 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)
 - [36] Liu, Y., Li, Y., You, S., Lu, F.: Unsupervised learning for intrinsic image decomposition from a single image. In: *Conf. Comput. Vision Pattern Recognit.*, Virtual, pp. 3248–3257 (2020). IEEE/CVF
 - [37] Forsyth, D., Rock, J.J.: Intrinsic image decomposition using paradigms. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(11), 7624–7637 (2021)
 - [38] Ebner, M., Hansen, J.: Depth map color constancy. *Bio-Algorithms Med-Syst.* **9**(4), 167–177 (2013)
 - [39] Xiao, Y., Tsougenis, E., Tang, C.-K.: Shadow removal from single rgb-d images. In: *Conf. Comput. Vision Pattern Recognit.*, Columbus, OH, USA, pp. 3011–3018 (2014). IEEE/CVF
 - [40] Han, G., Xie, X., Lai, J., Zheng, W.-S.: Learning an intrinsic image decomposer using synthesized rgb-d dataset. *IEEE Signal Process. Letters* **25**(6), 753–757 (2018)
 - [41] Liu, S., Jalab, H.A., Dai, Z.: Intrinsic face image decomposition from rgb images with depth cues. In: *Int. Visual Inf. Conf.*, Bangi, Malaysia, November 19–21, 2019, Bangi, Malaysia, pp. 149–156 (2019). Springer
 - [42] Xing, X., Groh, K., Karaoglu, S., Gevers, T.: Intrinsic appearance decomposition using point cloud representation. *arXiv preprint arXiv:2307.10924* (2023)
 - [43] Krajník, T., Blažíček, J., Santos, J.M.: Visual road following using intrinsic images. In: *European Conf. Mobile Robots*, Lincoln, UK, pp. 1–6 (2015). IEEE
 - [44] Strisciuglio, N., Tylecek, R., Blaich, M., Petkov, N., Biber, P., Hemming, J., Henten, E., Sattler, T., Pollefeys, M., Gevers, T., *et al.*: Trimbot2020: an outdoor robot for automatic gardening. In: *Int. Symposium on Robotics*, Munich, Germany, pp. 1–6 (2018). VDE
 - [45] Brandao, M., Hashimoto, K., Takanishi, A.: Friction from vision: A study of algorithmic and human performance with consequences for robot perception and teleoperation. In: *Int. Conf. Humanoid Robots*, Cancun, Mexico, pp. 428–435 (2016). IEEE
 - [46] Li, H., Zhou, C., Xue, W., Guo, Y.: License plate recognition based on intrinsic image decomposition algorithm. In: *Int. Conf. Comput. Science & Education*, Vancouver, BC, Canada, pp. 512–515 (2014). IEEE
 - [47] Tong, G., Li, Y., Sun, A., Wang, Y.: Shadow effect weakening based on intrinsic image extraction with effective projection of logarithmic domain for road scene. *Signal Image Video Process.* **14**, 683–691 (2020)
 - [48] Li, L., Xia, Z., Jiang, X., Ma, Y., Roli, F., Feng, X.: 3D face mask presentation attack detection based on intrinsic image analysis. *IET Biometrics* **9**(3), 100–108 (2020)
 - [49] Kang, X., Li, S., Fang, L., Benediktsson, J.A.: Intrinsic image decomposition for feature extraction of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **53**(4), 2241–2253 (2014)

- [50] 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)
- [51] Yue, H., Yang, J., Sun, X., Wu, F., Hou, C.: Contrast enhancement based on intrinsic image decomposition. *IEEE Trans. Image Process.* **26**(8), 3981–3994 (2017)
- [52] Beigpour, S., Van De Weijer, J.: Object recoloring based on intrinsic image estimation. In: *Int. Conf. Comput. Vision*, Barcelona, Spain, pp. 327–334 (2011). IEEE
- [53] Xu, C., Han, Y., Baci, G., Li, M.: Fabric image recolorization based on intrinsic image decomposition. *Textile Research J.* **89**(17), 3617–3631 (2019)
- [54] Bi, S., Han, X., Yu, Y.: An L_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Trans. Graphics* **34**(4), 1–12 (2015)
- [55] Du, J., Li, W., Tan, H.: Intrinsic image decomposition-based grey and pseudo-color medical image fusion. *IEEE Access* **7**, 56443–56456 (2019)
- [56] Kang, X., Li, S., Fang, L., Benediktsson, J.A.: Pansharpening based on intrinsic image decomposition. *Sens. Imag.* **15**, 1–17 (2014)
- [57] Butler, D.J., W., J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: *European Conf. Comput. Vision*, Florence, Italy, pp. 611–625 (2012). Springer
- [58] Beigpour, S., Kolb, A., Kunz, S.: A comprehensive multi-illuminant dataset for benchmarking of the intrinsic image algorithms. In: *Int. Conf. Comput. Vision*, Santiago, Chile, pp. 172–180 (2015). IEEE
- [59] Beigpour, S., Ha, M.L., Kunz, S., Kolb, A., Blanz, V.: Multi-view multi-illuminant intrinsic dataset. In: *British Mach. Vision Conf.*, York, UK (2016). BMVA Press
- [60] Song, S., Yu, F., Zeng, A., Chang, A.X., Savva, M., Funkhouser, T.: Semantic scene completion from a single depth image. In: *Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, pp. 1746–1754 (2017). IEEE/CVF
- [61] Zhou, H., Yu, X., Jacobs, D.W.: GloSh: Global-local spherical harmonics for intrinsic image decomposition. In: *Int. Conf. Comput. Vision*, Seoul, Korea, pp. 7820–7829 (2019). IEEE/CVF
- [62] Sengupta, S., Gu, J., Kim, K., Liu, G., Jacobs, D.W., Kautz, J.: Neural inverse rendering of an indoor scene from a single image. In: *Int. Conf. Comput. Vision*, Seoul, Korea, pp. 8598–8607 (2019). IEEE/CVF
- [63] Li, Z., Snavely, N.: Cgintrinsics: Better intrinsic image decomposition through physically-based rendering. In: *Eur. Conf. Comput. Vision*, Munich, Germany, pp. 371–387 (2018). CVF
- [64] Narihira, T., Maire, M., Yu, S.X.: Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In: *Int. Conf. Comput. Vision*, Santiago, Chile, pp. 2992–2992 (2015). IEEE
- [65] Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd Ed. Pearson Prentice Hall, – (2018)
- [66] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004)
- [67] Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, pp. 1398–1402 (2003). IEEE
- [68] Jiang, X., Schofield, A.J., Wyatt, J.L.: Correlation-based intrinsic image extraction from a single image. In: *European Conf. Comput. Vision*, Crete, Greece, pp. 58–71 (2010). Springer
- [69] Ahmed, S.N.: *Physics and Engineering of Radiation Detection*. Academic Press, UK (2007)
- [70] Sheikh, H.R., Bovik, A.C.: Image information and visual quality. *IEEE Trans. Image Process.* **15**, 430–444 (2006)

- [71] Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **20**, 2378–2386 (2011)
- [72] Luo, M.R., Cui, G., Rigg, B.: The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Res. Appl.* **26**, 340–350 (2001)
- [73] Sharma, G., Wu, W., Dalal, E.N.: The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **30**, 21–30 (2005)
- [74] Poynton, C.: *Digital Video and HD: Algorithms and Interfaces*. Elsevier, – (2012)
- [75] Gao, X., Lu, W., Tao, D., Li, X.: Image quality assessment and human visual system. In: *Proc. Vis. Commun. Image Process.*, Huangshan, China, pp. 316–325 (2010). SPIE
- [76] Ding, Y.: Image quality assessment based on human visual system properties. *Vis. Qual. Assessment Natural Med. Image* **37**, 63–106 (2018)
- [77] Zhu, W.-H., Sun, W., Min, X.-K., Zhai, G.-T., Yang, X.-K.: Structured computational modeling of human visual system for no-reference image quality assessment. *Int. J. Automat. Comput.* **18**, 204–218 (2021)
- [78] Ebner, M., Tischler, G., Albert, J.: Integrating color constancy into JPEG2000. *IEEE Trans. Image Process.* **16**, 2697–2706 (2007)
- [79] Ulucan, O., Ulucan, D., Ebner, M.: Color constancy beyond standard illuminants. In: *Int. Conf. Image Process.*, Bordeaux, France, pp. 2826–2830 (2022). IEEE
- [80] Karakaya, D., Ulucan, O., Turkan, M.: Image declipping: Saturation correction in single images. *Digital Signal Processing* **127**, 103537 (2022)