applied optics

Intrinsic decomposition from a single spectral image

Xi Chen, Weixin Zhu, Yang Zhao, Yao Yu, Yu Zhou, Tao Yue, Sidan Du, and Xun Cao*

School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
**Corresponding author: caoxun@nju.edu.cn*

Received 7 March 2017; revised 8 May 2017; accepted 10 May 2017; posted 10 May 2017 (Doc. ID 290223); published 7 July 2017

In this paper, we present a *spectral intrinsic image decomposition* (SIID) model, which is dedicated to resolve a natural scene into its purely independent intrinsic components: illumination, shading, and reflectance. By introducing spectral information, our work can solve many challenging cases, such as scenes with metameric effects, which are hard to tackle for trichromatic intrinsic image decomposition (IID), and thus offers potential benefits to many higher-level vision tasks, e.g., materials classification and recognition, shape-from-shading, and spectral image relighting. A both effective and efficient algorithm is presented to decompose a spectral image into its independent intrinsic components. To facilitate future SIID research, we present a public dataset with ground-truth illumination, shading, reflectance and specularity, and a meaningful error metric, so that the quantitative comparison becomes achievable. The experiments on this dataset and other images demonstrate the accuracy and robustness of the proposed method on diverse scenes, and reveal that more spectral channels indeed facilitate the vision task (i.e., segmentation and recognition).

OCIS codes: (140.3490) Lasers, distributed-feedback; (060.2420) Fibers, polarization-maintaining; (060.3735) Fiber Bragg gratings.

https://doi.org/10.1364/AO.56.005676

1. INTRODUCTION

With the rapid development of spectrometry and its comprehensive wide utilities, field spectral imaging [1–5], aiming at daily life vision applications, has been more and more affordable, so as to be used for classification [6,7], recognition [8], and tracking [9] by taking advantage of richer color channels than traditional trichromatic imaging. Unfortunately, different from remote sensing, the complicated nature of field spectral imaging, e.g., scene geometry (shading), inter-reflections, and complicated artificial illumination, makes the further utility of field spectral imaging a very challenging problem.

To tackle this problem, in this paper, we present a *spectral intrinsic image decomposition* (*SIID*) model in which Lambertain diffuse of a captured scene is decomposed into its constituent terms: illumination, shading, and reflectance, as shown in Fig. 1.

Such a decomposition could be advantageous to many vision tasks. For instance, image segmentation and recognition should be easier in a world without illumination and shading; shape-from-shading algorithms could benefit from an image with only shading effects. Image relighting would require a thorough removal of illumination condition.

Different than *illumination and reflectance spectra separation* (*IRSS*) works [10–13], which assume a spatially uniform illumination and mix shading, material reflectance as a generalized "reflectance term," SIID is devised to resolve a natural scene into its more elaborate components, which allows much more flexibility in manipulating spectral images. Comparing to the *intrinsic image decomposition* (*IID*) [14–16], which is dedicated to separate shading and reflectance from trichromatic images, SIID could be far more challenging due to its high dimensional nature. Nevertheless, spectral images can provide many more color channels than traditional trichromatic images, leading to a more accurate analysis for potential applications.

Generally, the major contributions of our paper lie in: (1) A representative and solvable SIID model is presented, which makes a fundamentally more sophisticated physical decomposition of a natural scene. (2) A both effective and efficient algorithm is proposed to decompose a single spectral image according to the model. (3) We provide a ground-truth dataset and a meaningful error metric for SIID, which will facilitate evaluations and future research of SIID.

2. RELATED WORK

In this section, we review several works highly relevant to ours, which can be roughly divided into two categories: one is IID, the other is IRSS.

A. Intrinsic Image Decomposition

Barrow and Tenenbaum [10] first proposed to represent distinct scene properties as separate "intrinsic" images. For

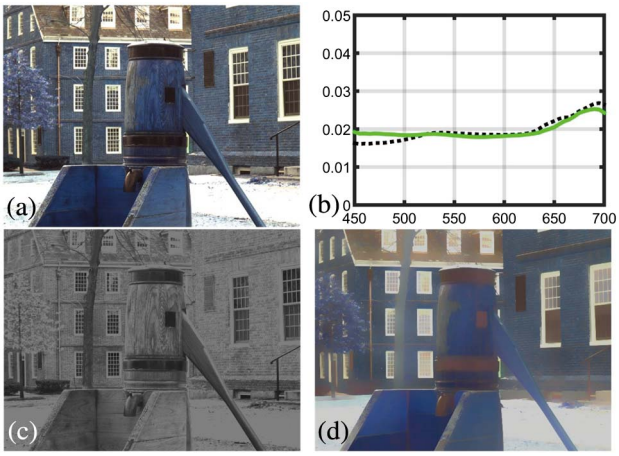

Fig. 1. Example for SIID. (a) Original spectral image is decomposed into three independent components: (b) illumination curve (solid green, predicted; dotted black, ground-truth); (c) shading; and (d) reflectance. Note that both (a) the original spectral image and (b) multispectral reflectance are integrated to three-channel images by using the response curves of RGB sensors for visualization.

decades, vision scientists have devoted themselves to this challenging problem and devised several reliable algorithms. These algorithms can be generally summarized as retinex-based approaches, learning methods [17,18], and multi-image algorithm [19,20]. Representative cutting-edge methods belong to the retinex family, which holds the assumption that artificial materials tend to have sharp reflectance edges, while natural surfaces (e.g., shadows and smooth curvature) tend to produce soft shading edges [21]. This principle was later applied to color images [22]. Kimmel *et al.* [23] steps forward to propose a variational model for the retinex problem that unifies previous methods. Shen *et al.* [24] introduce an optimization scheme in which the IID is formulated by minimizing an energy function in addition to a weighting constraint of local image properties. A major drawback of these methods is that the decomposition is implemented locally within a small window. Sinha and Adelson [25] sought to make a difference, whose work targets for global consistency of local inferences with respect to a plausible three-dimensional (3D) scene structure and illumination condition. Rother *et al.* [14] introduce a global sparsity prior on reflectance. By modeling reflectance intensity as a sparse set of basic colors, they achieve high-quality results without edge information. Followed by Zhao *et al.* [16], they propose non-local texture constraints, namely distant pixels with the same reflectance, to significantly reduce the ambiguity in decomposition.

B. Illumination and Reflectance Spectral Separation

Existing works on IRSS originate from either the classical problem of separating a single spectral color signal [26–28] or the color constancy of a red–green–blue (RGB) image [11,29]. Ikari *et al.* [12] extend the assumptions and constraints of Ref. [26] and are able to separate tens of spectral channels. Researches also generalize methods of trichromatic computational color constancy for the separation problem of a spectral image [30]. To handle the large-scale spectra separation problem, a widely accepted assumption proposed by Maloney [31] that any reflectance spectrum lies in a low-dimensional subspace. Based on this assumption, Zheng *et al.* [13] regard the IRSS as a low-rank matrix factorization problem and develop a scalable algorithm for separating the illumination and reflectance. Note that the reflectance here corresponds to a multiplexed term with "material reflectance", "shading", and "inter-reflectance" among the surface structure.

The proposed SIID differs from the previous methods in that it resolves a natural scene into its more elaborate components. Therefore, the SIID allows much more flexibility in manipulating spectral images and facilitates potential vision applications, such as genuine material discrimination (excluding the affection of shading, shadows, inter-reflections, and so on) and image relighting for an arbitrary view (with the geometry information).

3. OUR METHOD

In this paper, we decompose the Lambertain diffuse of a captured scene into its constituent terms, illumination, shading, and reflectance,

$$I(x,\lambda) = \underbrace{\overbrace{L(\lambda)}^{\text{Illumination}} \cdot \overbrace{s(x)}^{\text{Geometry}} \cdot \overbrace{R(x,\lambda)}^{\text{Material}}}_{\text{Lambertian Diffuse}}, \tag{1}$$

where $I(x,\lambda)$ is the captured spectral image, $L(\lambda)$ represents the illumination, $s(x)$ denotes the grayscale shading image, and $R(x,\lambda)$ is the reflectance, which represents constituent materials of the scene.

Figure 2 shows the schematic diagram of our SIID algorithm, which consists of the following steps: first, the illumination and the diffuse spectral image, denoted as $D(x,\lambda)$, are extracted via low-rank matrix-factorization-based IRSS [13]. Then, in diffuse image segmentation, we down-sample the diffuse component by a mean-average-based super pixel algorithm. Right after that, we introduce how to optimize to get a coarse level reflectance spectral image in coarse level reflectance optimization. Finally, the reflectance and shading spectral images are extracted using bilateral filter in fine level shading and reflectance estimation.

A. Spatial and Spectrum Space Down-sample

Most state-of-the-art work in the IID field [19,24] compute directly on the original input color image with very high resolution. However, due to the complex nature of spectral image, estimating shading, and reflectance components in a lower spatial and spectral resolution can greatly reduce the computational complexity of the nonlinear optimization and make the algorithm robust to background noise.

In this paper, the super-pixel [32] -based down-sampling is used to reduce the spatial and spectral resolution. Each pixel within the same super pixel region is assumed to share the same reflectance spectra. Thus, we assign each super pixel region a mean-average diffuse spectrum as

$$d_k = \frac{\sum_k \omega(x) \cdot D(x,\lambda)}{\sum_k \omega(x)}, \qquad x \in k, \tag{2}$$

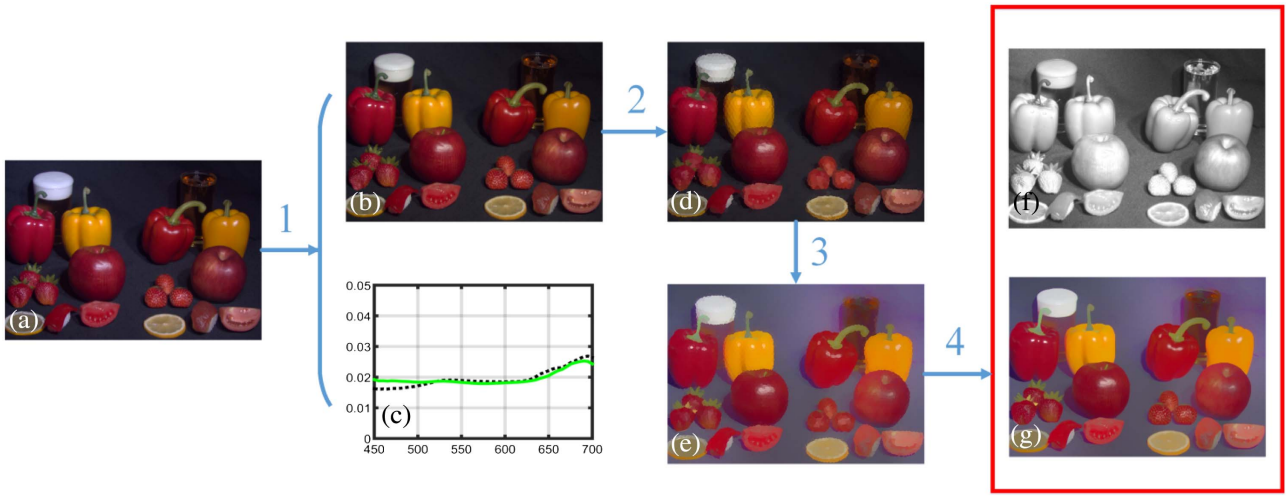

Fig. 2. Schematic of the proposed intrinsic spectral image decomposition algorithm. From a to g: (a) input spectral image; (b) diffuse spectral image; (c) illumination curve (solid green, predicted; dotted black, ground-truth); (d) down-sampled spectral image; (e) coarse level reflectance image; (f) fine level shading image; (g) fine level reflectance spectral image. Note that c, f, and g are the output of our algorithm. Digital numbers above indicate the major steps in our alorithm. From 1 to 4: 1, IRSS; 2, spatial space down-sample; 3, coarse level reflectance optimization; 4, fine level shading and reflectance estimation.

where the summation symbol $\sum$ indicates the vector addition of each spectral pixel within super pixel region p. Here, we define weight factor as

$$\omega(x) = \underbrace{e^{-(I_k(x)-\max(I_k))^2/\sigma_{ki}^2}}_{\text{intensity term}} \cdot \underbrace{e^{-(G_k(x)-\min(G_k))^2/\sigma_{kg}^2}}_{\text{gradient term}}$$
$$\text{s.t. } \sigma_{ki} = \text{var}(I_k) \quad \text{and} \quad \sigma_{kg} = \text{var}(G_k), \tag{3}$$

where I_k and G_k denote the grayscale and the gradient values of pixels belonging to k, respectively; max/min/var is the operator returning the maximum/minimum/standard deviation value of the input argument. The first term of $\omega(x)$ is called the *intensity term*, which means penalizing pixels located in the shadow area. These pixels usually suffer from poor illumination conditions, hence, the reflectance spectra computed from these areas are unreliable. The second term of Eq. (3) is called the *gradient term*, which means penalizing pixels with sharp intensity variation. Strong inter-reflection usually appears on the small concave surface, which will lead to a significant error estimation of the reflectance spectra. However, the intensity of the reflected ray from these areas, if receiving sufficient illumination, could still be very high. Thus, we cannot use the intensity term to penalize these pixels.

B. Coarse Level Reflectance Optimization

In this section, we exploit the way to generate a shading and reflectance spectral image in the super pixel scale. Right after that, a bilateral filter is used to recover fine scale shading and reflectance spectral image. *Coarse scale optimization:* We will assume the object surface is Lambertian and, hence, has diffuse reflection. Borrowing the idle fromIID, the captured luminance spectrum at every super pixel point d_p is modeled as the product of the Lambertian reflectance spectrum r_p and shading spectrum s_p. Mathematically, this model can be expressed as

$$d_p = r_p \cdot * s_p, \tag{4}$$

where d_p, r_p, and s_p are all vectors with dimensions equal to the number of spectral bands of the captured signal in the super pixel level, "$\cdot *$" denotes element wise multiplication (similar hereafter). The problem is to derive r_p and s_p by optimizing the following energy function:

$$E_C = E_{\text{LC}} + E_{\text{GC}}. \tag{5}$$

1. Local Constant Constraint

This constraint is common to the well-known assumption: Retinex [10] is where shading and reflectance differences between adjacent super pixels are assumed to be small in most cases. This assumption can be mathematically expressed as

$$E_{\text{LC}} = \sum_{p,q \in N_{p,q}} E_{\text{LCR}} + E_{\text{LCS}}, \tag{6}$$

where p, q represents the adjacent super pixels pairs in the neighborhood pair set $N_{p,q}$. E_{LCR} and E_{LCS} are reflectance and shading constancy, respectively. They are modelled as

$$E_{\text{LCR}} = \omega_{p,q}(r_q - r_p)^2, \tag{7}$$

$$E_{\text{LCS}} = (1 - \omega_{p,q})(s_p - s_q)^2, \tag{8}$$

respectively, where $\omega_{p,q}$ indicates the cosine distance between the normalized power of spectrum d_p and d_q.

2. Global Sparse Constraint

Difficulties in decomposing shading and reflectance components come mainly from the ill-posed nature of the problem. So, here, we seek to reduce the number of unknowns by adding the global sparse constraint E_{GC}, which is based on the observation that a common natural scene in the real world only consists of a few basic materials with different reflectance spectra. The global sparse constraint can be mathematically represented as

$$E_{\mathrm{GC}} = \sum_{p,q \in G_{p,q}} (r_p - r_q)^2, \tag{9}$$

which forces the reflectance spectra of super pixels p and q to be very similar. We select super pixel pairs $p, q \in G_{p,q}$ according to the following criterion:

Criterion 1 The Euclidean distance between d_p and d_q in the spectra space is smaller than a threshold D_{spec}.

Criterion 2 d_p and d_q converge to the same density peak using mean shift clustering.

Criterion 3 The Euclidean distance between d_p and d_q in image space is larger than a threshold D_{img}.

Criterion 1 and **Criterion 2** screen all the candidate super pixels potentially sharing the similar reflectance spectra. **Criterion 3** gets rid of those super pixel pairs that have appeared in local constraint term E_{LC}.

Now, we seek to minimize Eq. (5) to generate shading and reflectance components. According to Eq. (4), $s_p - s_q = 0$ could be written as

$$d_p \cdot * r_q - d_q \cdot * r_p = 0, \tag{10}$$

where $\cdot *$ means element wise multiply. With the help of Eq. (10), Eq. (5) could be represented in matrix form as

$$E_C = \|W_{\mathrm{LRC}} R\|^2 + \|V_{\mathrm{LSC}} R\|^2 + \|U_{\mathrm{GC}} R\|^2, \tag{11}$$

where R is the vector consisting of r for all super pixels in a vector, W_{LRC}, V_{LSC}, and U_{GC} depend on the coefficients in Eqs. (7)–(9). To circumvent the ambiguity about the scaling factor between shading and reflectance, we also explored the **absolute scale constraint**, where we force the coefficient sum of vector R to equal a constant value T_s, which is mathematically represented as $LR = T_s$, where L is a unit diagonal matrix. Then, we minimize E_C equally to solve the following equation:

$$\underbrace{\begin{pmatrix} W_{\mathrm{LRC}}^T W_{\mathrm{LRC}} + \\ \lambda_1 V_{\mathrm{LSC}}^T V_{\mathrm{LSC}} + \\ \lambda_2 U_{\mathrm{GC}}^T U_{\mathrm{GC}} + \\ \lambda_3 L^T L \end{pmatrix}}_{Q} R = \lambda_3 L^T T_s. \tag{12}$$

Because the matrix Q is self-adjoint and sparse, we can use the conjugate gradient method [33] to solve this equation iteratively, which typically converges very fast. Once we generate the spectra reflectance image in the coarse level, the corresponding shading spectra image can be simply computed as

$$s_p = d_p \cdot / r_p, \qquad p \in \Re, \tag{13}$$

where "$\cdot/$" mean element wise divide (similar hereafter).

C. Fine Level Shading and Reflectance Estimation

Now we discuss how to obtain shading and reflectance in the fine level. In Section 3.A, we use the mean-average method to down-sample the input diffuse spectral image, which leads to the loss of high-frequency power. This high-frequency power will in turn appear in the shading component, using Eq. (13). Instead of using the optimization method, we use a rather simple but effective way to recover the local high-frequency lost in the spectra reflectance image. According to Ref. [34], high-frequency signals can be removed easily with local filters. By using a bilateral filter in coarse shading image s_p, we generate fine level shading image $\tilde{s}_p$. Finally, the fine level reflectance can be computed simply as

$$\tilde{r}_p = d_p \cdot / \tilde{s}_p. \tag{14}$$

4. SIID DATASET

Grosse *et al.* [35] introduced a cross-polarizing approach to obtain the ground-truth dataset forIID and, thus, greatly facilitated the research for IID in recent years. We borrow the ideas from the trichromatic IID benchmark and establish the very first spectral intrinsic ground-truth dataset called the SIID dataset (Section 4.A). An error metric is also presented in this section to provide a benchmark for verifying and a comparison of the state of the arts (Section 4.B). Meanwhile, other vision research (e.g., IRSS, segmentation, recognition, relighting) can also benefit by the SIID dataset.

A. Ground-Truth Acquisition

We begin with extracting illumination from the scene by exposing the spectral camera directly to the light source. Different to Ref. [35], we relax the white light assumption to an arbitrary illumination condition. We manually set an exposure brightness value to assure that the illumination spectra can be observed with sufficient dynamic range.

Similar to Ref. [35], we also use the cross-polarizing approach to separate the diffuse and the specular term in Eq. (1). Specifically, polarizing filters are placed over both the light source and the spectral camera. Since the light ray reflecting off the surface of objects remains polarized, we are able to remove specularity by rotating the polarizing filter on the camera until its axis is orthogonal to the reflected light. By capturing two spectral photographs of the same scene, one with a specular variation denoted as I_{orig}, the other with only Lambertian diffuse denoted as I_{diff}, the specularity can be directly obtained by subtracting the two terms:

$$I_{\mathrm{spec}} = I_{\mathrm{orig}} - I_{\mathrm{diff}}. \tag{15}$$

The next step is the estimation of shading and reflectance. The scenes in our dataset contain nonuniform texture, for this reason, we remove the texture variation by spraying these objects with a thin layer of flat white oilpaint and re-photographing it under the same illumination as I_{diff}. Denoting the texture removed in the photograph as I_{shad}, we have the following relationship:

$$I_{\mathrm{shad}} = L \cdot s_{\mathrm{obj}} \cdot R_{\mathrm{oil}}, \tag{16}$$

$$I_{\mathrm{diff}} = L \cdot s_{\mathrm{obj}} \cdot R_{\mathrm{obj}}, \tag{17}$$

where subscript obj and oil refer to the original object and the oil-painted one, respectively. Note that the reflectance of the oil painting could not be omitted since the reflectance spectra of the oilpainting are not strictly white. Combining Eqs. (16) and (17), the reflectance and shading terms can be computed as

$$s_{\mathrm{obj}} = \frac{\|I_{\mathrm{shad}}\|_2}{\|R_{\mathrm{oil}} \cdot L\|_2} \quad \text{and} \quad R_{\mathrm{obj}} = \frac{I_{\mathrm{diff}} \cdot R_{\mathrm{oil}}}{I_{\mathrm{shad}}}. \tag{18}$$

To obtain R_{oil}, we picture the uniformly sprayed oil-painted cardboard, denoted as I_{oil}. Since the cardboard is flat, the shading component in I_{oil} could be regarded as uniform. Thus, for each pixel of the cardboard in I_{oil}, the reflectance spectra can be computed as I_{oil}/L. We average all these pixels to generate R_{oil}.

We regard those complicated effects beyond the Lambertian diffuse term like inter-reflection as outliers of our dataset. To minimize light inter-reflecting between different objects, we cover all nearby objects with black cloths. However, we still could not avoid inter-reflections happening on small concave surfaces. In these areas, the reflectance spectra computed using Eq. (18) are unreliable. To solve this problem, we manually mark these areas and replace the reflectance spectra with correct ones computed from the other area consisting of the same material without inter-reflections. Figure 3 lists some of our ground-truth segmentation results in our SIID dataset. Note that all spectral images are visualized using corresponding synthesized RGB data. For each object in our dataset, we employ two kinds of light sources in capturing the scene, one is white LED with 5500 K color temperature, the other is a warm-tone white iodine–tungsten lamp (ITL) with 3500 K color temperature. By positioning the light sources in different places, we create different illumination conditions for ground-truth capturing.

B. Error Metric

We would like to introduce a comprehensive error metric for the evaluation of different SIID algorithms over all the different intrinsic components. In the remainder of this section, we will define *percentage mean square error* (MSE) as

$$\text{pmse}_M(r_g(\lambda), r_e(\lambda)) = \frac{1}{M}\sum_{\lambda}\frac{(r_g(\lambda) - r_e(\lambda))^2}{r_g(\lambda)}, \tag{19}$$

where r_g and r_e are the ground-truth and estimated spectra with M bands, respectively; we also use L, S, and R to represent illumination, shading, and normalized reflectance, and g and e represent ground-truth and estimated results, respectively.

According to Eq. (1), the error criterion should include three terms: illumination err_L, shading err_S, and reflectance err_R.

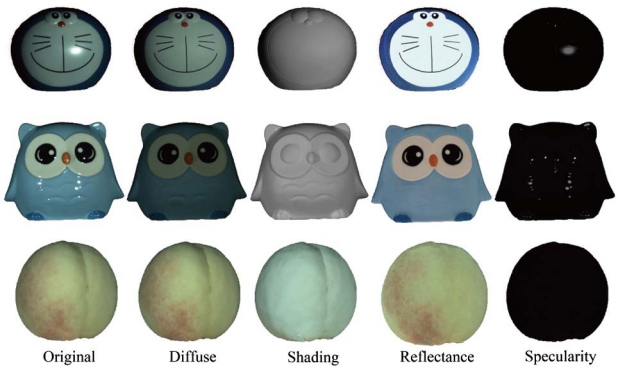

Fig. 3. Capturing ground-truth intrinsic spectral components, visualized by corresponding RGB images. From left to right: original images I_{orig} with maximum specularities; diffuse images I_{diff} with specular removed; shading images I_{shad} obtained by spraying oil paint on the surface of objects of interest; reflectance image R_{obj} and specular images I_{spec}.

Using Eq. (19), the error metric of the lighting spectra is defined as

$$\text{err}_L = \text{pmse}_M(L_g(\lambda), L_e(\lambda)). \tag{20}$$

We define the error metric of reflectance in dimension-reduced spectral space as

$$\text{err}_R = \frac{1}{X}\sum_{x}\text{pmse}_N(R_g(x,\hat{\lambda}), R_e(x,\hat{\lambda})), \tag{21}$$

where X is the total number of pixels in the computed reflectance spectral image; $\hat{\lambda}$ is the dimension-reduced spectral space. In experiments, we employ a K–L divergence-based spectral band selection algorithm [36], which is unsupervised with a very simple computational process, to select N spectral bands with the least relevance.

The local MSE (LMSE) [35] is used as the error metric of shading, which means we allow scale ambiguity between the ground-truth shadings and estimated ones. Thus, given ground-truth and estimated images, the error metric of shading is defined as

$$\text{err}_S = \frac{\text{lmse}_k(S_g(x), S_e(x))}{\text{lmse}_k(S_g(x), 0)}. \tag{22}$$

There are two parameters: window size k of the LMSE and selected spectral band size N. The typical value of k is set to 20, according to Ref. [35]. The larger value of N tends to preserve more information, while the smaller one emphasizes avoiding information redundancy. We use $N = k = 20$ to conduct experiments in Section 5.

5. EXPERIMENTS

In this section, we quantitatively validate our algorithm via the proposed dataset and make a head-to-head comparison with the naive spectral extension of the state-of-the-art IID method retinex [24] (named spectral retinex or SR for short). The detail of the extension is shown in Appendix A. We also demonstrate potential applications and justify that SIID indeed facilitates high-level vision tasks. The test cases of the dataset we used consist of various materials, from plastic (DORAEMON) and ceramic (OWL and RABBIT) to cellulose (PEACH). Some indoor and outdoor spectral images provided by Refs. [37,38] are also included to validate the robustness of our algorithm.

A. Experiments on Proposed SIID Dataset

We test our algorithm on the proposed SIID dataset and provide percentage error scores of several test cases in Table 1. Comparing it with ground-truth, our algorithm achieves desirable decomposition results in terms of both our proposed error metric score (1.59% on average for the entire database) and the visual quality of the decomposed results.

We qualitatively (Fig. 4) and quantitatively (Fig. 5) compare our method with SR, which is the spectral extension of Shen *et al.* [24]. Considering that SR assumes white light, we only compare the shading and reflectance decompositions.

The close-ups in Fig. 4 illustrate significant improvements of our algorithm. Specifically, as demonstrated in the close-up figures of OWL. The sharp shading variation caused by a small concave is recovered more accurately by our algorithm than SR,

Table 1. Percentage Error Scores[a]

Item	err_L	err_S	err_R	Average
BAYMAX LED	1.03	0.15	1.06	0.74
BAYMAX ITL	0.90	0.12	1.32	0.78
OWL LED	0.92	0.96	0.61	0.83
OWL ITL	0.97	1.25	0.83	0.92
PEACH LED	1.19	2.27	2.78	2.08
PEACH ITL	1.11	2.12	2.19	1.80
GYPSUM LED	1.15	1.17	1.05	1.19
GYPSUM ITL	1.23	1.15	1.63	1.34
RABBIT LED	1.27	1.12	2.23	1.59
RABBIT ITL	1.52	1.35	2.43	1.90

[a]Note that the first column lists the objects' code name and illumination condition, where capital ITL indicates the decompositions from iodine-tungsten lamp illumination.

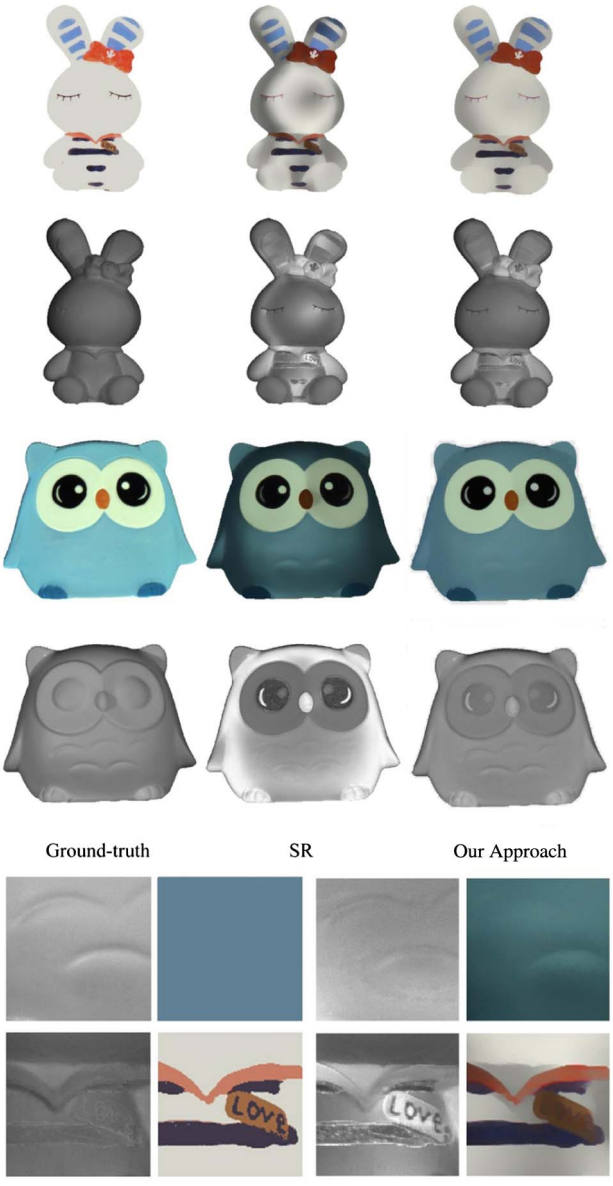

Fig. 4. Comparison between our method and the SR approach, using test case RABBIT and OWL. Odd and even rows are the reflectance and the shading results, respectively.

mainly a credit to the global sparse constraints, which forces these areas to be more uniform. Moreover, our algorithm achieves more uniform reflectance results in those concave surface, which is attributed to the use of the super pixel to remove sharp shading variations by the optimization procedure. As listed in the close-ups of RABBIT, although the richer spectral information indeed relieves the *shading–reflectance ambiguity*, it still cannot solve this problem completely: the artificial shading edge of the body area of RABBIT is still clearly visible in the estimated reflectance image. Our algorithm recovers better results in terms of the uniform reflectance in these area.

In Fig. 5, SR and our algorithm are compared: our algorithm achieves an average 38% lower MSE than SR (only spectral information is added into a retinex algorithm). This comparison shows the effectiveness of our algorithm other than a simple spectral extension of retinex.

B. Experiments on Open Source Spectral Image Datasets

We also run our method on two famous open source spectral image datasets [37,38], as shown in Fig. 6. The results indicate the effectiveness of the proposed algorithm on even very complex scenarios. For example, there is very sharp geometry variation in the second and fifth column of Fig. 6. This will directly lead to the sharp variation of shading in a small local patch, which does not comply with our local constraint term. With the help of the global constraint term, our result is still reliable in terms of visual richness. The global constraint term also helps to generate smooth reflectance results in those areas with large-scale uniform texture, e.g., the wall of buildings in the first and third column of Fig. 6.

In summary, besides the richer spectral information, our algorithm indeed plays a very important role in de-multiplexing the field spectra scene into its intrinsic components.

C. Applications

To justify whether spectral reflectance component $R(x, \lambda)$ provided by SIID helps in high-level vision tasks, we test it in several challenging material segmentation and recognition scenarios with metameric effects (the same RGB color, but different spectral curves). The RGB color images and corresponding spectral images of real skin versus printed skin, real flowers versus plastic flowers, and nylon cloth versus cotton cloth are used

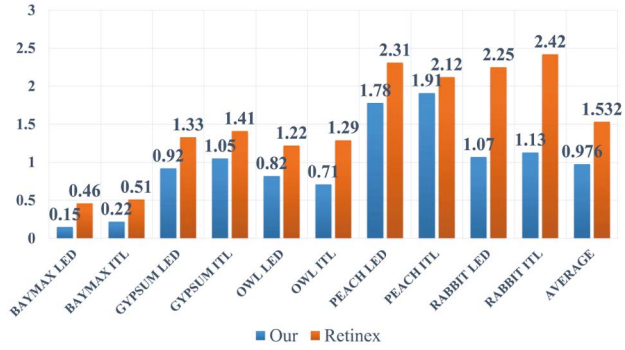

Fig. 5. Error metric scores of our method (blue bar) versus SR (orange bar). Note that the score is the average of err_S and err_R. The average scores of all test cases are listed in the last column as well.

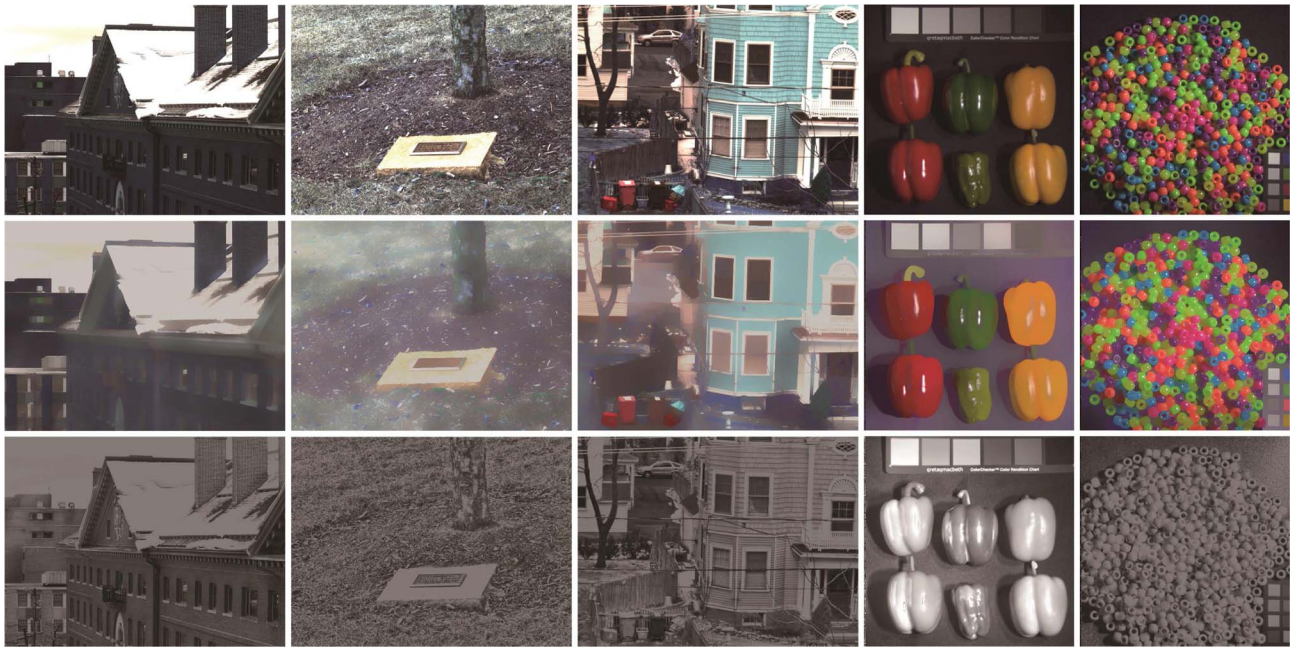

Fig. 6. Some of our decomposition results of the open source spectral database provided by Harvard University [37] (first three columns) and Columbia University [38] (last two columns). From the top row to the bottom are original, reflectance, and shading color images, respectively, synthesized from row spectral data.

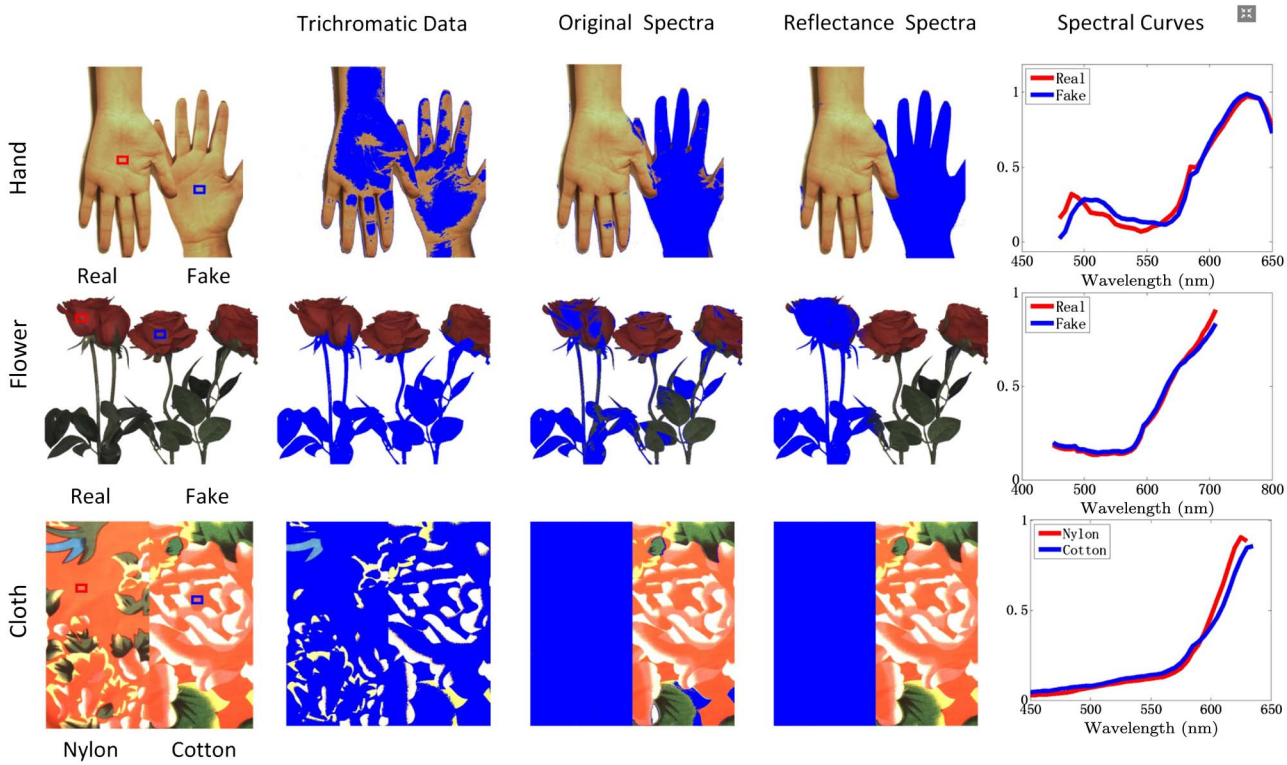

Fig. 7. Application of the proposed SIID model for high-level vision tasks. The first column is the original spectral images, and recognition and segmentation results of trichromatic images (i.e., RGB images), captured original spectral images, and reflectance spectral images generated by our SIID algorithm are shown in the second column to the forth column, respectively. The spectral curves of the two semblable materials are shown in the last column. From top to bottom: skin versus printed skin, real flowers versus plastic flowers, and nylon cloth versus cotton cloth.

Table 2. Recognition Rate

Item	Trichromatic	Original	Reflectance
Hand	47.3%	95.1%	97.6%
Flower	46.8%	56.9%	96.4%
Cloth	57.4%	95.2%	98.2%

as input. The support vector machine (SVM) is applied for recognizing and segmenting these semblable materials from the RGB images, original spectral images, and the intrinsic reflectance images extracted from original spectra to remove illumination and shading effects, respectively. As shown in Fig. 7, the SVM completely fails to recognize the real and fake materials on RGB images for the reason of the metameric effect. As for original spectral images, the performance of the SVM is improved greatly in the cases Hand and Cloth. However, it is still hard for original images to generate promising recognition results when scenarios are polluted by the shading effect caused by complex surface geometry, i.e. the petal area of the case Flower. After removing the illumination and shading effects by our SIID algorithm, the semblable materials are recognized and segmented very well in the three cases. The detail recognition rate of each case in Fig. 7 is shown in Table 2. It is obvious that the proposed SIID model and algorithm can promisingly benefit the high-level vision tasks, and thus have many potential applications.

6. CONCLUSION

We propose the SIID model to resolve a natural scene into its purely independent intrinsic components: illumination, shading, and reflectance. Experiments reveal that our SIID model could make use of the extra spectral channels to relieve certain tricky problems, e.g., shading–reflectance ambiguity, that are intractable by traditional trichromatic intrinsic decomposition. Moreover, the de-multiplexed results of natural scenes will offer practical benefits in many vision applications, such as materials classification and recognition, shape-from-shading, spectral image relighting, etc. We also propose both an effective and efficient algorithm to decouple the SIID model from a single spectral image. By making full use of rich spectral information, our algorithm achieves promising results, even in concave surfaces with strong inter-reflection. To stimulate more and more vision scientists to devise high-performance SIID algorithms and to facilitate research on diverse SIID algorithms, we provide a ground-truth dataset and a meaningful error metric for better evaluation, benchmark, and potential data for other tasks.

Our algorithm has left out quite a few important aspects that deserve to be explored in depth. For example, our algorithm appears to be less effective in those areas with strong shading or occlusion. This is the common problem for the retinex-based optimization method, since these areas do not comply with the local constraint term mentioned in Section 3.B, namely, sharp diffuse variation caused by shading.

Regarding future works, we will primarily focus on the application of SIID, such as materials recognition, geometry reconstruction, and spectral images relighting. We will also dedicate ourselves to improving the effectiveness of our algorithm and making it work for more general conditions.

APPENDIX A: SPECTRAL RETINEX

As discussed in Ref. [35], the retinex assumption, originally proposed by Ref. [21], is still the most effective and widely used assumption of IID. Thus, we would like to explore how well these approaches perform in the SIID field. Several state-of-art retinex-based approaches [14,16,24] can be revised to work for SIID. Considering that the approach proposed by Shen *et al.* [24] has been employed in a feature extraction-based spectral images segmentation and achieved desirable results [6], we would like to quantitatively compare this approach with ours. Thus, we review and revise this approach to make it work for the SIID field in this section. For an easier expression, we would like to call the original version of [24] as *trichromatic retinex* (TR), and the revised version as SR.

The TR is based on the assumption of local color characteristics in natural images: in a local window of an image, the changes of pixel values are usually caused by the changes of the reflectance, i.e., the texture variation. This assumption was inspired by the behavioral evidence that humans usually distinguish variation in shading from variation in reflectance and has been well approved and used in several IID works [14,16,24]. Specifically, this method is based on the following weight function that models the reflectance value of one pixel as a weighted average of its adjacent reflectance value:

$$R_i = \sum_{j \in N_i} \omega_{ij} \cdot R_j, \qquad \omega_{ij} = e^{-[(Y_i - Y_j)^2/\sigma_{iY}^2 + \langle I_i, I_j\rangle^2/\sigma_{iT}^2]}, \quad (23)$$

where ω_{ij} measures the similarity of the reflectance values between pixel i and pixel j, N_i is a local window of $k \times k$ pixels near pixel i. For a trichromatic image, Y denotes the intensity image, which can be simply computed as the average of three color channels. $\langle I_i, I_j\rangle = \arccos(I_{ir} \cdot I_{jr} + I_{ig} \cdot I_{jg} + I_{ib} \cdot I_{jb})$ represents the angle between pixel vector I_i and I_j. σ_{iY}^2 and σ_{iT}^2 are the variances of the angle of pixel vectors mentioned above and pixel intensity in the local window ω_i, respectively. Combined with the decomposition function $I_i = R_i S_i$, the IID problem could be solved by optimizing the following equation:

$$E(R, S) = \sum_i \left(R_i - \sum_{j \in \omega_i} \omega_{ij} R_j \right)^2 + \sum_i \left(\frac{I_i}{S_i} - R_i \right)^2. \quad (24)$$

For a detailed description of the optimization method, we refer the reader to Ref. [24].

For the reason that no classification and complex mathematical model is required for the TR approach, and, thus, the algorithm can be easily extended for processing spectral images that usually have more than three data channels. For a spectral image with multiple channels, TR can be revised to work for SIID by changing the definition of $\langle I_i, I_j\rangle$ in Eq. (23) as

$$\langle I_i, I_j\rangle = \arccos \sum_{l=1}^{M} I_{li} I_{lj}, \quad (25)$$

where I is a normalized spectral image with M spectral bands.

Funding. Natural Science Foundation of Jiangsu Province (BE2015152); National Natural Science Foundation of China (NSFC) (61100111, 61201425, 61271231, 61300157, 61371166, 61422107).

REFERENCES

1. G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: an introduction," IEEE Signal Process. Mag. **31**(1), 105–115 (2014).
2. X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 297–304.
3. C. Ma, X. Cao, R. Wu, and Q. Dai, "Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system," Opt. Lett. **39**, 937–940 (2014).
4. L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 4942–4950.
5. L. Gu, A. A. Robles-Kelly, and J. Zhou, "Efficient estimation of reflectance parameters from imaging spectroscopy," IEEE Trans. Image Process. **22**, 3648–3663 (2013).
6. X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," IEEE Trans. Geosci. Remote Sens. **53**, 2241–2253 (2015).
7. M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady, "3d imaging spectroscopy for measuring hyperspectral patterns on solid objects," ACM Trans. Graph. **31**, 38 (2012).
8. A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *IEEE Conference on Computer Vision and Pattern Recognition* (2011), pp. 193–200.
9. H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, 2010), pp. 44–51.
10. H. Barrow and J. Tenenbaum, "Computer vision systems," in *Computer Vision Systems* (Academic, 1978), p. 2.
11. A. Gijsenij, T. Gevers, and J. Van De Weijer, "Computational color constancy: survey and experiments," IEEE Trans. Image Process. **20**, 2475–2489 (2011).
12. A. Ikari, R. Kawakami, R. T. Tan, and K. Ikeuchi, "Separating illumination and surface spectral from multiple color signals," in *Digitally Archiving Cultural Objects* (Springer, 2008), pp. 297–321.
13. Y. Zheng, I. Sato, and Y. Sato, "Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization," in *IEEE Conference on Computer Vision and Pattern Recognition* (2015).
14. C. Rother, M. Kiefel, L. Zhang, B. Schölkopf, and P. V. Gehler, "Recovering intrinsic images with a global sparsity prior on reflectance," in *Advances in Neural Information Processing Systems* (2011), pp. 765–773.
15. L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2008), pp. 1–7.
16. Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin, "A closed-form solution to retinex with nonlocal texture constraints," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1437–1444 (2012).
17. J. Chang, R. Cabezas, and J. W. Fisher III, "Bayesian nonparametric intrinsic image decomposition," in *Computer Vision–ECCV* (Springer, 2014), pp. 704–719.
18. M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1459–1472 (2005).
19. P.-Y. Laffont, A. Bousseau, and G. Drettakis, "Rich intrinsic image decomposition of outdoor scenes from multiple views," IEEE Trans. Vis. Comput. Graphics **19**, 210–224 (2013).
20. Y. Weiss, "Deriving intrinsic images from image sequences," in *IEEE International Conference on Computer Vision* (IEEE, 2001), Vol. **2**, pp. 68–75.
21. E. H. Land and J. McCann, "Lightness and retinex theory," J. Opt. Soc. Am. **61**, 1–11 (1971).
22. B. V. Funt, M. S. Drew, and M. Brockington, "Recovering shading from color images," in *European Conference on Computer Vision* (Springer, 1992), pp. 124–132.
23. R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," Int. J. Comput. Vis. **52**, 7–23 (2003).
24. J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 3481–3487.
25. P. Sinha and E. Adelson, "Recovering reflectance and illumination in a world of painted polyhedra," in *4th International Conference on Computer Vision* (IEEE, 1993), pp. 156–163.
26. P.-R. Chang and T.-H. Hsieh, "Constrained nonlinear optimization approaches to color-signal separation," IEEE Trans. Image Process. **4**, 81–94 (1995).
27. M. S. Drew and G. D. Finlayson, "Analytic solution for separating spectra into illumination and surface reflectance components," J. Opt. Soc. Am. A **24**, 294–303 (2007).
28. J. Ho, V. Funt, and M. S. Drew, "Separating a color signal into illumination and surface reflectance components: theory and applications," IEEE Trans. Pattern Anal. Mach. Intell. **12**, 966–977 (1990).
29. K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data," IEEE Trans. Image Process. **11**, 972–984 (2002).
30. C. P. Huynh and A. Robles-Kelly, "A solution of the dichromatic model for multispectral photometric invariance," Int. J. Comput. Vis. **90**, 1–27 (2010).
31. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," J. Opt. Soc. Am. A **3**, 1673–1683 (1986).
32. K. S. Kim, D. Zhang, M. C. Kang, and S. J. Ko, "Improved simple linear iterative clustering superpixels," in *IEEE 17th International Symposium on Consumer Electronics* (2013), pp. 259–260.
33. D. S. Kershaw, "The incomplete cholesky—conjugate gradient method for the iterative solution of systems of linear equations," J. Comput. Phys. **26**, 43–65 (1978).
34. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *International Conference on Computer Vision* (1998), pp. 839–846.
35. R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *IEEE 12th International Conference on Computer Vision* (IEEE, 2009), pp. 2335–2342.
36. X.-S. Liu, L. Ge, B. Wang, and L.-M. Zhang, "An unsupervised band selection algorithm for hyperspectral imagery based on maximal information," Int. J. Infrared Millim. Waves **2**, 015 (2012).
37. D. H. Foster, K. Amano, S. M. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
38. F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," IEEE Trans. Image Process. **19**, 2241–2253 (2010).