



# Single Image Highlight Removal with a Sparse and Low-Rank Reflection Model

Jie Guo<sup>1</sup>, Zuojian Zhou<sup>2</sup>, and Limin Wang<sup>1</sup>(✉)

<sup>1</sup> State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China  
guojie@nju.edu.cn, 07wanglimin@gmail.com

<sup>2</sup> School of Information Technology, Nanjing University of Chinese Medicine,  
Nanjing, China  
lygzzj@163.com

**Abstract.** We propose a sparse and low-rank reflection model for specular highlight detection and removal using a single input image. This model is motivated by the observation that the specular highlight of a natural image usually has large intensity but is rather sparsely distributed while the remaining diffuse reflection can be well approximated by a linear combination of several distinct colors with a sparse and low-rank weighting matrix. We further impose the non-negativity constraint on the weighting matrix as well as the highlight component to ensure that the model is purely additive. With this reflection model, we reformulate the task of highlight removal as a constrained nuclear norm and  $l_1$ -norm minimization problem which can be solved effectively by the augmented Lagrange multiplier method. Experimental results show that our method performs well on both synthetic images and many real-world examples and is competitive with previous methods, especially in some challenging scenarios featuring natural illumination, hue-saturation ambiguity and strong noises.

**Keywords:** Highlight removal · Low-rank · Sparse · Diffuse reflection

## 1 Introduction

A vast majority of objects in real-world scenes exhibit both diffuse and specular reflections. The existence of specular reflection is recognized as a hindrance for a variety of computer vision tasks including image segmentation, pattern recognition, and object detection, since it creates undesired discontinuities and reduces image contrast. Therefore, separating specular highlights from diffuse reflection is of crucial importance and forms the core of many high-level vision tasks. Although highlights can be well suppressed by some special facilities such as polarizing filters [1–4] or multi-spectral light stages [5], it is more appealing

**Electronic supplementary material** The online version of this chapter ([https://doi.org/10.1007/978-3-030-01225-0\\_17](https://doi.org/10.1007/978-3-030-01225-0_17)) contains supplementary material, which is available to authorized users.

to remove them using just a single color image without any hardware assistance. However, highlight removal from a single image constitutes an ill-posed problem with more unknowns than equations to solve.

Our goal in this paper is to separate reflection components of a single input image based on the dichromatic reflection model [6]. We notice that highlight regions in many real-world scenes are contiguous pieces with relatively small size while colors of diffuse reflection can be well approximated by a small number of distinct colors. The former observation implies that highlight regions in an image tend to be sparse. The latter observation, which resembles the observation in non-local image dehazing [7], reveals that diffuse colors form tight clusters in RGB space with a low-rank and sparse weighting matrix. This inspires us to separate diffuse and specular reflections with sparse and low-rank matrix decomposition [8–10].

Based on the insights gained from these observations, we propose a sparse and low-rank reflection (SLRR) model. This model assumes that specular highlights have the same spectral distribution with the incident illumination (*i.e.*, the neutral interface reflection assumption [11]) and are rather sparse in the spatial domain of a given image. We also assume that each diffuse color can be represented by a linear combination of several “basis colors” from a “color dictionary” and the weighting matrix formed by the coefficients is low-rank and sparse. The low-rankness indicates some global structures in the weighting matrix [12, 13] while the sparsity stems from the fact that the diffuse colors are clustered as blocks in the image. In this context, the highlight removal process is formulated as a constrained nuclear norm and  $l_1$ -norm minimization problem, which can be efficiently solved by the augmented Lagrange multiplier (ALM) method with alternating direction minimizing (ADM) strategy [14]. To ensure that our model is purely additive and avoids counteracting each other by subtraction, we further explicitly impose the non-negativity constraint on it. This will bring in some inequality constraints that are difficult for optimization. We address this issue by introducing slack variables and converting inequalities into equalities. These equalities are added into the augmented Lagrange function.

To demonstrate the effectiveness and robustness of our method, we conduct an extensive evaluation using various synthetic and real-world images from several public image datasets. Experimental results show that our method achieves better performance on many tasks than state-of-the-art. Particularly, it can effectively handle some challenging scenarios including natural illumination and hue-saturation ambiguity.

## 2 Related Work

Highlight removal has been paid much attention in recent years [15]. Existing work on highlight removal is generally grouped into two high-level categories based on the number of images used.

In the first category are approaches that remove highlight with multiple images. As highlight regions are direction-dependent, it is natural to use image sequences from different points of view [16, 17] or from multiple light positions [5, 18, 19] to restore the diffuse reflection. The polarization based methods

[1–4] require a set of images captured with different polarization orientations for accurate highlight removal, considering that specular and diffuse reflections hold different degrees of polarization. Other auxiliary data, such as that generated by a multi-spectral light stage [16, 17] or a flash system [20, 21], also benefits this task. Despite their effectiveness, these methods are less appealing to everyday users because such image datasets are not often available in practice.

The second line of work focuses on removing highlight from a single image. As this problem is inherently ill-posed, prior knowledge or assumptions on the characteristics of natural images should be exploited to make the problem tractable. Early work relies on color space analysis [22–24] which is only limited to dealing with uniform surface colors and probably involves image segmentation.

In order to handle textured surfaces, Tan and Ikeuchi [25] pioneered the idea of specular-free image which has been widely studied ever since. A specular-free image is a pseudo-diffuse image that has the same geometrical profile as the true diffuse component of the input image. It can be generated by setting the diffuse maximum chromaticity of each pixel to an arbitrary value [25, 26] or by subtracting the minimum value of the RGB channels for each pixel [27–30]. Kim *et al.* [31] obtained an approximated specular-free image via applying the dark channel prior. Suo *et al.* [32] defined  $l_2$  chromaticity and used it to generate the specular-free image. Yang *et al.* [26] proposed a fast bilateral filter adopting the specular-free image as the range weighting function. Several methods [28, 30, 33] use the specular-free image for pixel clustering and then recover the diffuse colors in each cluster. The main drawback of the specular-free image is it suffers from hue-saturation ambiguity which exists in many natural images. Liu *et al.* [33] suggested using an additional compensation step to raise the achromatic component of the diffuse chromaticity.

Some single image based methods do not explicitly rely on a specular-free image. For instance, Mallick *et al.* [34] proposed a PDE algorithm, which iteratively erodes the specular component in the SUV color space. But this method performs poorly on large specular regions. Ren *et al.* [35] introduced the color-lines constraint into the dichromatic reflection model and proposed a fast highlight removal method. Li *et al.* [36] made use of specialized domain knowledge to guide the removal of specular highlights in facial images. Inpainting techniques which synthetically fill in the missing regions using the neighboring patterns have also been applied to recover diffuse colors [37, 38]. Our method bears some similarity to that of Akashi and Okatani [39] which formulates the separation of reflections as a sparse non-negative matrix factorization (NMF) problem. However, current algorithms for NMF are sensitive to initial values and only guarantee finding a local minimum rather than a global minimum. Therefore this method requires running several times to get the most reasonable result. Furthermore, since NMF is highly sensitive to outliers in general cases, this method may fail in the presence of strong specularities or noises. Oppositely, our method relies on sparse and low-rank decomposition which is much more robust to outliers.

### 3 Sparse and Low-Rank Reflection Model

To exploit sparse and low-rank structures in an image, we derive a new reflection model following the formulation of the dichromatic reflection model [6]. This model states that the observed image intensity  $\mathbf{I}$  of a pixel  $p$  is the sum of a diffuse component  $\mathbf{I}_d$  and a specular component  $\mathbf{I}_s$ :

$$\mathbf{I}(p) = \mathbf{I}_d(p) + \mathbf{I}_s(p) = m_d(p)\mathbf{\Lambda}(p) + m_s(p)\mathbf{\Gamma}(p) \quad (1)$$

where  $\mathbf{\Lambda}(p)$  and  $\mathbf{\Gamma}(p)$  respectively denote the chromaticities of the diffuse and specular components.  $m_d(p)$  and  $m_s(p)$  represent their corresponding coefficients which are achromatic and only depend on imaging geometry.

The specular chromaticity could be assumed to be uniform for a given image and equals to the chromaticity of the incident illumination [11]. Like many other specular removal methods, we estimate the illumination chromaticity  $\mathbf{\Gamma}$  of a real-world image via the color constancy algorithm in [40], and then normalize the original image by  $\mathbf{I}(p)/(3\mathbf{\Gamma})$  in a preprocessing step. After that, we have a pure white illumination color, *i.e.*,  $\mathbf{\Gamma}_r = \mathbf{\Gamma}_g = \mathbf{\Gamma}_b = 1/3$ . Often, we observe that highlight regions are small in size and are distributed rather sparsely. This implies that  $m_s(p)$  is non-zero only for a low density of pixels.

On the contrary, the diffuse component  $\mathbf{I}_d$  usually has a high density of valid data. However, since the diffuse reflectance of natural objects is commonly piecewise constant, the number of distinct colors of  $\mathbf{I}_d$  is orders of magnitude smaller than the number of pixels [7]. Therefore, given a proper color dictionary  $\mathbf{\Phi}_d = [\phi_1, \phi_2, \dots, \phi_K]$ ,  $\mathbf{I}_d$  can be faithfully reconstructed by  $\mathbf{I}_d(p) = \sum_{k=1}^K \phi_k w_k(p)$  in which  $w_k(p) \geq 0$  is the weighting coefficient of the pixel  $p$  w.r.t. the basis color  $\phi_k$ . The non-negativity constraint makes every pixel value stay in the convex hull of the dictionary, avoiding counteracting each other by subtraction. Ideally,  $w_k$  is non-zero for only one basis color, implying that  $w_k$  is sparse. Meanwhile, as the color dictionary  $\mathbf{\Phi}_d$  is often over-complete, the best choice of  $w_k$  should be drawn from a low-rank subspace.

We finally come up with the following sparse and low-rank reflection model:

$$\mathbf{X} = \mathbf{\Phi}_d \mathbf{W}_d + \mathbf{\Gamma} \mathbf{M}_s \quad (2)$$

where  $\mathbf{X}$  is a  $3 \times N$  matrix with each column representing a pixel color.  $N$  is the total number of pixels in an image.  $\mathbf{W}_d$  is a  $K \times N$  matrix formed by the weighting coefficients of all pixels. As the specular chromaticity  $\mathbf{\Gamma}$  is assumed to be a constant column vector for a given image, we represent the specular component using a rank-one expression in which  $\mathbf{M}_s$  is a row vector of size  $1 \times N$  encoding the position and intensity of specular highlights.

The above analysis reveals that  $\mathbf{M}_s$  tends to be sparse while  $\mathbf{W}_d$  is both sparse and low-rank. Furthermore, since pixel values are non-negative, we also impose the non-negativity constraint on  $\mathbf{M}_s$  and  $\mathbf{W}_d$ . In this way, the reflection separation

problem of an input image can be formulated as the following optimization problem with both equality and inequality constraints:

$$\begin{aligned} & \min_{\mathbf{W}_d, \mathbf{M}_s} \text{rank}(\mathbf{W}_d) + \lambda \|\mathbf{M}_s\|_0 + \tau \|\mathbf{W}_d\|_0 \\ \text{s.t. } & \mathbf{X} = \Phi_d \mathbf{W}_d + \Gamma \mathbf{M}_s, \mathbf{W}_d \geq 0, \mathbf{M}_s \geq 0 \end{aligned} \tag{3}$$

in which  $\lambda$  and  $\tau$  are parameters used to balance the effect of different components.  $\|\cdot\|_0$  denotes the  $l_0$  norm of a matrix, which counts the number of non-zero entries in the matrix.

### 4 Model Optimization

Unfortunately, Eq. 3 is highly non-convex and no efficient solution is available. To make the optimization tractable, we relax Eq. 3 via replacing  $\text{rank}(\cdot)$  with  $\|\cdot\|_*$  and  $\|\cdot\|_0$  with  $\|\cdot\|_1$  as in the sparse and low-rank matrix decomposition. Here  $\|\cdot\|_*$  is the nuclear norm of a matrix defined by the sum of its singular values and  $\|\cdot\|_1$  is the  $l_1$  norm. We also introduce two auxiliary variables  $\mathbf{J}$  and  $\mathbf{H}$  to make the objective function separable:

$$\begin{aligned} & \min_{\mathbf{J}, \mathbf{M}_s, \mathbf{H}, \mathbf{W}_d} \|\mathbf{J}\|_* + \lambda \|\mathbf{M}_s\|_1 + \tau \|\mathbf{H}\|_1 \\ \text{s.t. } & \mathbf{X} = \Phi_d \mathbf{W}_d + \Gamma \mathbf{M}_s, \mathbf{J} = \mathbf{W}_d, \mathbf{H} = \mathbf{W}_d, \mathbf{W}_d \geq 0, \mathbf{M}_s \geq 0. \end{aligned} \tag{4}$$

This optimization problem involves two non-negativity constraints. To cope with these inequalities, a straightforward strategy, as suggested by Zhuang *et al.* [41], is to clamp the negative entries in  $\mathbf{W}_d$  and  $\mathbf{M}_s$  to zero directly during each iteration. Unfortunately, we find that such a simple strategy shows poor convergence as shown in Fig. 8. Instead, we introduce two non-negative slack variables  $\mathbf{S}_1$  and  $\mathbf{S}_2$  to convert the non-negativity constraints into two equality constraints:  $\mathbf{W}_d - \mathbf{S}_1 = 0$  and  $\mathbf{M}_s - \mathbf{S}_2 = 0$ .

Various algorithms have been developed to solve the optimization problem, among which the ALM method [8, 14] is most widely used. This method replaces the original constrained optimization problem by a sequence of unconstrained subproblems which can be efficiently solved by soft-thresholding or singular value thresholding (SVT) [42].

The augmented Lagrange function of the above optimization problem with two slack variables is

$$\begin{aligned} \mathcal{L}(\mathbf{J}, \mathbf{M}_s, \mathbf{H}, \mathbf{W}_d, \mathbf{Y}_i, \mathbf{S}_1, \mathbf{S}_2, \mu) = & \\ & \|\mathbf{J}\|_* + \lambda \|\mathbf{M}_s\|_1 + \tau \|\mathbf{H}\|_1 + \sum_{i=1}^5 \langle \mathbf{Y}_i, \mathbf{E}_i \rangle + \frac{\mu}{2} (\|\mathbf{E}_i\|_F^2) \end{aligned} \tag{5}$$

where  $\mathbf{Y}_i$  are Lagrange multipliers and  $\mu > 0$  is a penalty parameter.  $\mathbf{E}_i$  are five equality constraints, namely,  $\mathbf{E}_1 = \mathbf{X} - \Phi_d \mathbf{W}_d - \Gamma \mathbf{M}_s$ ,  $\mathbf{E}_2 = \mathbf{J} - \mathbf{W}_d$ ,  $\mathbf{E}_3 = \mathbf{H} - \mathbf{W}_d$ ,  $\mathbf{E}_4 = \mathbf{W}_d - \mathbf{S}_1$ , and  $\mathbf{E}_5 = \mathbf{M}_s - \mathbf{S}_2$ .  $\langle \cdot, \cdot \rangle$  denotes the standard

inner product between two matrices while  $\|\cdot\|_F$  represents the Frobenius norm of a matrix. The ALM method with ADM strategy decomposes the minimization of  $\mathcal{L}$  into several subproblems in which the variables are updated alternately with other variables fixed.

#### 4.1 Update $\mathbf{J}$

With some algebra, the optimization problem over  $\mathbf{J}$ , keeping other variables fixed, is rearranged as

$$\mathbf{J}^* = \arg \min_{\mathbf{J}} \frac{1}{\mu} \|\mathbf{J}\|_* + \frac{1}{2} \left\| \mathbf{J} - \left( \mathbf{W}_d - \frac{\mathbf{Y}_2}{\mu} \right) \right\|_F^2. \quad (6)$$

For conventional nuclear norm, the solution to this subproblem is given analytically by

$$\mathbf{J}^* = \mathcal{D}_{1/\mu}(\mathbf{W}_d - \mathbf{Y}_2/\mu). \quad (7)$$

Here  $\mathcal{D}_{1/\mu}$  is an SVT operator [42] defined as  $\mathcal{D}_{1/\mu}(\mathbf{A}) = \mathbf{U}[\text{sgn}(\boldsymbol{\Sigma}) \max(|\boldsymbol{\Sigma}| - \mu^{-1}, 0)]\mathbf{V}^\top$ , in which  $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$  is a singular value decomposition of  $\mathbf{A}$ . For weighted nuclear norm, we can also obtain a similar analytical solution if the weights are in a non-ascending order [43]. In this paper, we prefer the weighted version for better performance, and the weight  $\mathbf{w}$  is set as

$$w_i = \sqrt{N}/(|\boldsymbol{\Sigma}_{i,i}| + \delta) \quad (8)$$

where a small constant  $\delta$  is used to avoid dividing by zero. With the weight  $\mathbf{w}$ , we have  $\mathcal{D}_{1/\mu}(\mathbf{A}) = \mathbf{U}[\text{sgn}(\boldsymbol{\Sigma}) \max(|\boldsymbol{\Sigma}| - \mu^{-1} \text{diag}(\mathbf{w}), 0)]\mathbf{V}^\top$ .

#### 4.2 Update $\mathbf{M}_s$

$\mathbf{M}_s$  in our context is a  $N$ -dimensional row vector. Suppose  $\boldsymbol{\Gamma}$  is known in advance, we can formulate the update of  $\mathbf{M}_s$  as

$$\mathbf{M}_s^* = \arg \min_{\mathbf{M}_s} \frac{\lambda}{\mu g} \|\mathbf{M}_s\|_1 + \frac{1}{2} \left\| \mathbf{M}_s - \frac{1}{g} \left( \boldsymbol{\Gamma}^\top \left( \mathbf{X} - \boldsymbol{\Phi}_d \mathbf{W}_d + \frac{\mathbf{Y}_1}{\mu} \right) - \frac{\mathbf{Y}_5}{\mu} + \mathbf{S}_2 \right) \right\|_2^2 \quad (9)$$

in which  $g = \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}$ . By employing the soft-thresholding operator  $\mathcal{S}_\tau(x) = \text{sgn}(x) \max(|x| - \tau, 0)$ , this subproblem also has an analytical solution:

$$\mathbf{M}_s^* = \mathcal{S}_{\frac{\lambda}{\mu g}} \left( (\boldsymbol{\Gamma}^\top (\mathbf{X} - \boldsymbol{\Phi}_d \mathbf{W}_d + \mathbf{Y}_1/\mu) - \mathbf{Y}_5/\mu + \mathbf{S}_2) / g \right). \quad (10)$$

#### 4.3 Update $\mathbf{H}$

Similarly, we can rearrange the subproblem optimizing  $\mathbf{H}$  as

$$\mathbf{H}^* = \arg \min_{\mathbf{H}} \frac{\tau}{\mu} \|\mathbf{H}\|_1 + \frac{1}{2} \left\| \mathbf{H} - \left( \mathbf{W}_d - \frac{\mathbf{Y}_3}{\mu} \right) \right\|_F^2 \quad (11)$$

and solve it efficiently with the soft-thresholding operator:

$$\mathbf{H}^* = \mathcal{S}_{\tau/\mu}(\mathbf{W}_d - \mathbf{Y}_3/\mu). \quad (12)$$

Note that this operator is performed element-wise for a matrix.

### 4.4 Update $\mathbf{W}_d$

With other variables fixed, the subproblem w.r.t.  $\mathbf{W}_d$  is quadratic. Therefore, this is a standard least squares regression problem with a closed-form solution:

$$\mathbf{W}_d^* = (\Phi_d^\top \Phi_d + 3\mathbf{I})^{-1} (\Phi_d^\top \mathbf{X} - \Phi_d^\top \Gamma \mathbf{M}_s + \mathbf{J} + \mathbf{H} + \mathbf{S}_1 + (\Phi_d^\top \mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3 - \mathbf{Y}_4) / \mu). \tag{13}$$

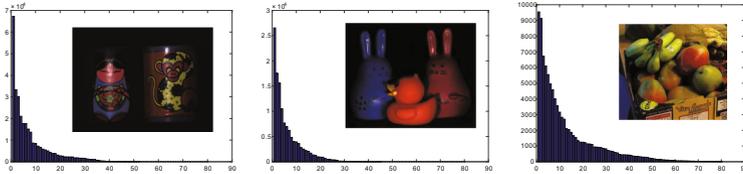


Fig. 1. Sorted color histograms of three typical natural images.

### 4.5 Update Slack Variables

The subproblems w.r.t. two slack variables ( $\mathbf{S}_1$  and  $\mathbf{S}_2$ ) are also standard least squares regression problems. Their solutions are given by

$$\mathbf{S}_1^* = \max(\mathbf{W}_d + \mathbf{Y}_4 / \mu, 0) \quad \text{and} \quad \mathbf{S}_2^* = \max(\mathbf{M}_s + \mathbf{Y}_5 / \mu, 0) \tag{14}$$

respectively. The max function is added to ensure that both  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are non-negative.

### 4.6 Construct Color Dictionary

In sparse and low-rank representations, constructing a proper dictionary is important. A simple option would take the entire input data as the dictionary [12]. However, such a large dictionary is computationally expensive and consumes too much storage space. Inspired by the non-local prior [7] of natural images, we adopt the following histogram binning method to construct a color dictionary  $\Phi_d$ .

This method requires constructing a 2-dimensional binning of longitude  $\theta$  and latitude  $\phi$  on a unit sphere. We use the same strategy as in [7] to uniformly tessellate a unit sphere. For an input image, we view each pixel  $\mathbf{I}(p)$  as a 3-dimensional vector and transform it into spherical coordinates:  $\mathbf{I}(p) = [r(p), \theta(p), \phi(p)]$ . Then, each pixel is assigned into a proper bin based on  $\theta(p)$  and  $\phi(p)$ . After that, we sort all the bins in a descending order according to their densities (see Fig. 1) and select the top  $K$  bins. The bin centers are regarded as atoms of the dictionary  $\Phi_d \in \mathbb{R}^{3 \times K}$ . Figure 1 shows sorted color histograms of three typical natural images used in this paper. Obviously, there are only a limited number of distinct colors in an image.

## 4.7 Highlight Removal Algorithm

The complete steps of our highlight removal method are outlined in Algorithm 1. The input of this algorithm is a single image that is reshaped into a  $3 \times N$  matrix in which each column stores a pixel value. The output includes a color dictionary  $\Phi_d$ , a weighting matrix  $\mathbf{W}_d$ , and a specular coefficient matrix  $\mathbf{M}_s$ . With these outputs, we can easily obtain the diffuse component and the specular component according to Eq. 2. Note that an adaptive updating strategy for the penalty parameter  $\mu$  is used as shown in line 10, which makes the convergence faster.

---

### Algorithm 1. Highlight Removal with the SLRR Model

---

**Input:** Image data  $\mathbf{X} \in \mathbb{R}^{3 \times N}$ .

- 1: **Initialize:**  $\mathbf{W}_d = \mathbf{J} = \mathbf{H} = 0$ ,  $\mathbf{M}_s = 0$ ,  $\mathbf{S}_1 = 0$ ,  $\mathbf{S}_2 = 0$ ,  $\mathbf{Y}_i = 0$ ,  $\mu = 0.1$ ,  $\mu_{max} = 10^{10}$ ,  $\rho = 1.1$ ,  $\epsilon = 10^{-6}$ ,  $K = 50$ ;
- 2: Construct a color dictionary  $\Phi_d$ ;
- 3: **while** not converged **do**
- 4:   Update  $\mathbf{J}$  according to Eq. 7;
- 5:   Update  $\mathbf{M}_s$  according to Eq. 10;
- 6:   Update  $\mathbf{H}$  according to Eq. 12;
- 7:   Update  $\mathbf{W}_d$  according to Eq. 13;
- 8:   Update slack variables  $\mathbf{S}_1$  and  $\mathbf{S}_2$  according to Eq. 14;
- 9:   Update Lagrange multipliers:  $\mathbf{Y}_i \leftarrow \mathbf{Y}_i + \mu \mathbf{E}_i$ ,  $i = 1$  to 5;
- 10:   Update  $\mu$ :  $\mu \leftarrow \min(\mu_{max}, \rho\mu)$ ;
- 11:   Check convergence:  $\max_i(\|\mathbf{E}_i\|_F / \|\mathbf{X}\|_F) < \epsilon$ ;

**Output:**  $\Phi_d \in \mathbb{R}^{3 \times K}$ ,  $\mathbf{W}_d \in \mathbb{R}^{K \times N}$ , and  $\mathbf{M}_s \in \mathbb{R}^{1 \times N}$ .

---

## 5 Experimental Results and Discussions

To verify the effectiveness and robustness of the proposed method, we evaluate it on both synthetic images and many real and practical scenarios. We also present comparisons to some recent methods proposed by Tan and Ikeuchi [25], Shen *et al.* [28], Yang *et al.* [26], Shen and Zheng [30], Akashi and Okatani [39], and Ren *et al.* [35]. Unless otherwise stated,  $\lambda$  is set as  $0.1/\sqrt{N}$  while  $\tau$  is set as  $1/\sqrt{N}$ . This is a good starting point for many cases.

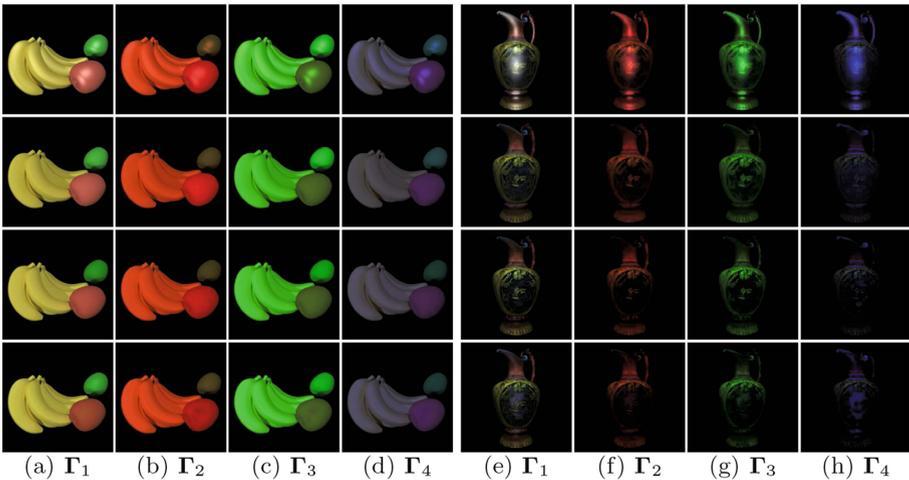
### 5.1 Study on Illumination Chromaticity

Recall that we use the color constancy algorithm proposed in [40] to estimate the illumination chromaticity  $\Gamma$ . Figure 2 and Table 1 show that our method is robust to a certain amount of estimation error of  $\Gamma$ . To demonstrate this, we provide two scenes rendered with four different lighting configurations. The accurate values of the illumination chromaticity are listed in the second column of Table 1 while the estimated values for the two scenes are listed in the third and sixth columns, respectively. As expected, using accurate values of  $\Gamma$  in our

method (the third row of Fig. 2) leads to high-quality results that closely match the ground truth results (the second row of Fig. 2). Although the estimation error of  $\Gamma$  has negative influence on our method (the fourth row of Fig. 2), we still achieve satisfactory results with a small decline in PSNR (see Table 1).

## 5.2 Comparisons on Synthetic and Laboratory Images

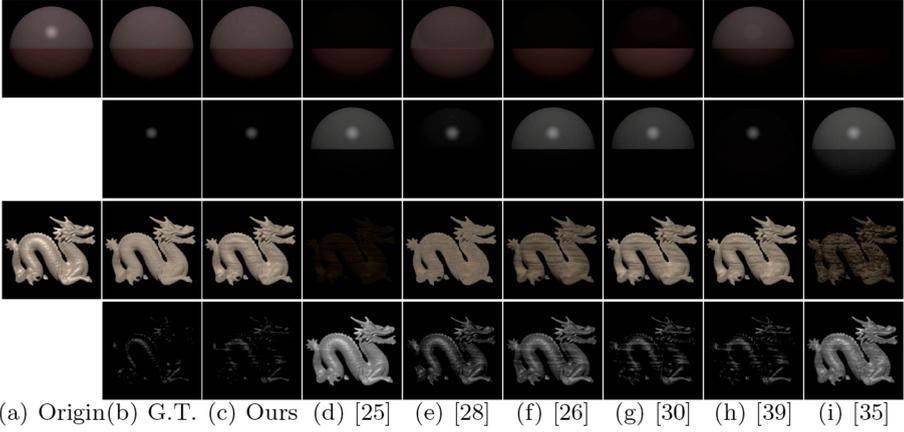
Figure 3 conducts experiments on two synthetic images. The first scene includes a single rendered sphere. Diffuse reflectances of the upper and lower hemisphere are set as  $[0.8, 0.5, 0.5]$  and  $[0.8, 0.2, 0.2]$ , respectively, such that they share the same hue but different saturation. As shown in Fig. 3, when the diffuse reflectance is close to the incident illumination color, the separation will be error-prone for many methods, especially those based on a specular-free image. These methods will mistakenly regard the entire region of the upper hemisphere as contaminated by specular reflection, making this region very dim after separation. Although the NMF method [39] tries to preserve the saturation of such regions, the separation is not sufficient in this experiment. In contrast, our method based on the SLRR model regards strong highlights as sparse outliers and encodes remaining diffuse colors using a color dictionary in which both colors:  $[0.8, 0.5, 0.5]$  and  $[0.8, 0.2, 0.2]$  after normalization are present. Consequently, we successfully remove highlights with less color distortions. Compared to other methods, our method achieves the best separation result that is very close to the ground truth (G.T.). Further discussions on the ball scene are given in the supplemental material. Another challenge



**Fig. 2.** Influence of the illumination chromaticity on highlight removal. The first row shows two rendered scenes under four different illumination chromaticities and the second row gives the corresponding diffuse component of each input image. The third and fourth rows present our method's results using the accurate and estimated illumination chromaticities, respectively.

**Table 1.** Illumination chromaticities of four different lighting configurations in Fig. 2 and the corresponding PSNR values. Here PSNR1 and PSNR2 are calculated for our method with the accurate and estimated values of  $\Gamma$ , respectively.

	Accurate	Fig. 2 left			Fig. 2 right		
		Estimated	PSNR1	PSNR2	Estimated	PSNR1	PSNR2
$\Gamma_1$	[1/3, 1/3, 1/3]	[0.36, 0.29, 0.35]	33.1	31.9	[0.34, 0.35, 0.31]	34.9	34.0
$\Gamma_2$	[0.6, 0.2, 0.2]	[0.64, 0.19, 0.17]	37.1	33.0	[0.63, 0.20, 0.17]	37.4	35.1
$\Gamma_3$	[0.2, 0.6, 0.2]	[0.21, 0.63, 0.16]	37.3	35.6	[0.20, 0.63, 0.17]	37.6	33.8
$\Gamma_4$	[0.2, 0.2, 0.6]	[0.22, 0.23, 0.55]	37.0	33.0	[0.22, 0.22, 0.56]	37.4	35.1



**Fig. 3.** Highlight removal results of two synthetic images. From (c) to (i), we compare our method to those of Tan and Ikeuchi [25], Shen *et al.* [28], Yang *et al.* [26], Shen and Zheng [30], Akashi and Okatani [39], and Ren *et al.* [35].

scene shown in the third and fourth rows of Fig. 3 also contains hue-saturation ambiguity. Again, our method outperforms previous work and retains most of the texture details.

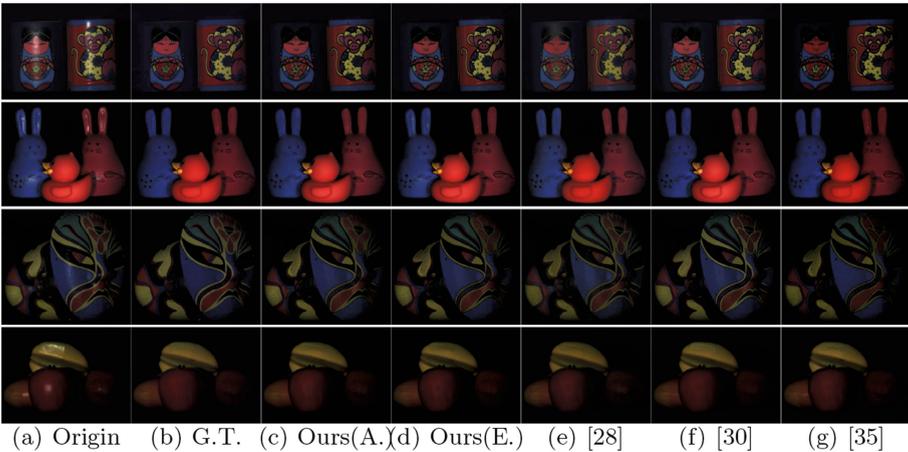
Figure 4 presents the results of different highlight removal methods on four laboratory captured images with well-controlled lighting configurations. These images accompanied with ground truth results are captured by Shen and Zheng [30] and have already been normalized w.r.t. the illumination color. Therefore, the accurate value of the illumination chromaticity is  $\Gamma = [1/3, 1/3, 1/3]$ . Figure 4 compares our method to three competing methods that perform well on this dataset. For a fair comparison, the accurate illumination chromaticity is used in each method. We also provide the results generated by our method using the estimated illumination chromaticity. Visual comparisons reveal that the separated diffuse components of our method using either accurate or estimated illumina-

tion chromaticity match the ground truth results quite well. The differences are subtle although errors exist in illumination estimation.

To further validate the accuracy of our method, we provide quantitative analysis for Figs. 3 and 4 in Table 2. The error metrics used for evaluation are PSNR and SSIM w.r.t. the ground truth. As seen, our method using accurate illumination chromaticity achieves the highest scores for the two synthetic images, and also achieves two highest scores both in PSNR and SSIM for the laboratory images. Similarly, the errors of illumination estimation do not affect our results apparently. The results demonstrate the superiority of the proposed method, and are consistent with the visual impression in Figs. 3 and 4.

### 5.3 Robustness Exploration

To demonstrate that our method is robust to strong noises, we conduct an experiment in Fig. 5. We generate a noisy image by adding zero-mean Gaussian noises with variance 0.1 to each channel of a real-world image. As shown in Fig. 5, some methods, in particular those based on material clustering [28, 30], are vulnerable to noises and produce separation results with degraded accuracy. The method of Akashi and Okatani [39] also fails in the presence of strong noises because the NMF is highly sensitive to outliers. In comparison, our method can generate high quality results even in this tough case. For this image, we set  $\tau = 5/\sqrt{N}$ . More results testing the sensitivity of our method to image noises are provided in the supplemental material.

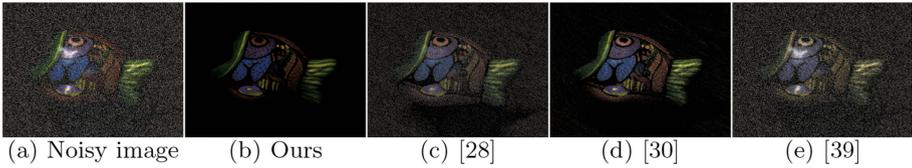


**Fig. 4.** Highlight removal results of four laboratory images with ground truth (G.T.). We compare our method using either accurate (A.) or estimated (E.) illumination chromaticity to three competing methods. See Table 2 for quantitative comparisons.

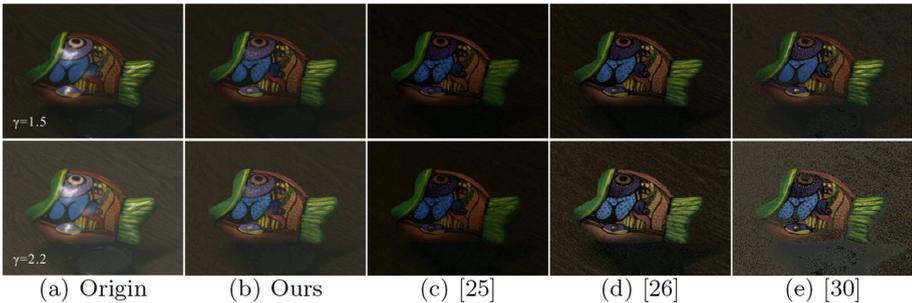
**Table 2.** Quantitative evaluation in terms of PSNR and SSIM for the images in Fig. 3 (Ball and Dragon) and Fig. 4 (Cups, Animals, Masks and Fruit). The highest scores are in red and the second highest scores are in blue.

Scenes	PSNR								SSIM							
	Ours(A.)	Ours(E.)	[25]	[28]	[26]	[30]	[39]	[35]	Ours(A.)	Ours(E.)	[25]	[28]	[26]	[30]	[39]	[35]
Ball	<b>39.6</b>	<b>38.9</b>	18.8	28.4	19.1	20.0	27.1	17.3	<b>0.998</b>	<b>0.997</b>	0.780	0.974	0.813	0.856	0.885	0.644
Dragon	<b>31.8</b>	<b>30.5</b>	11.4	21.1	17.3	25.4	27.6	13.5	<b>0.972</b>	<b>0.964</b>	0.638	0.933	0.883	0.945	0.952	0.738
Cups	<b>39.1</b>	38.6	29.3	37.5	34.1	<b>38.9</b>	35.7	38.0	<b>0.963</b>	0.959	0.767	0.962	0.941	<b>0.966</b>	0.937	0.957
Animals	<b>35.7</b>	34.4	26.1	34.2	33.0	<b>37.4</b>	26.8	30.6	<b>0.975</b>	0.938	0.929	<b>0.974</b>	0.970	0.971	0.802	0.896
Masks	<b>34.4</b>	31.5	23.9	32.1	28.4	<b>33.9</b>	32.3	30.0	<b>0.955</b>	0.911	0.789	<b>0.943</b>	0.899	0.941	0.657	0.913
Fruit	36.4	36.5	29.2	<b>38.0</b>	32.4	<b>39.2</b>	30.8	37.5	0.930	0.921	0.912	<b>0.961</b>	0.939	<b>0.960</b>	0.765	0.952

Our method is also robust to the response function of the camera. In Fig. 6, we simulate the error of camera calibration by transforming each pixel value  $\mathbf{I}$  to  $\mathbf{I}^{1/\gamma}$ .  $\gamma$  is set to 1.5 (the first row) and 2.2 (the second row), respectively. As seen, our method still works well even if there are some errors in the calibration of the cameras response function. It clearly outperforms some previous methods (*e.g.*, [25, 26] and [30]).



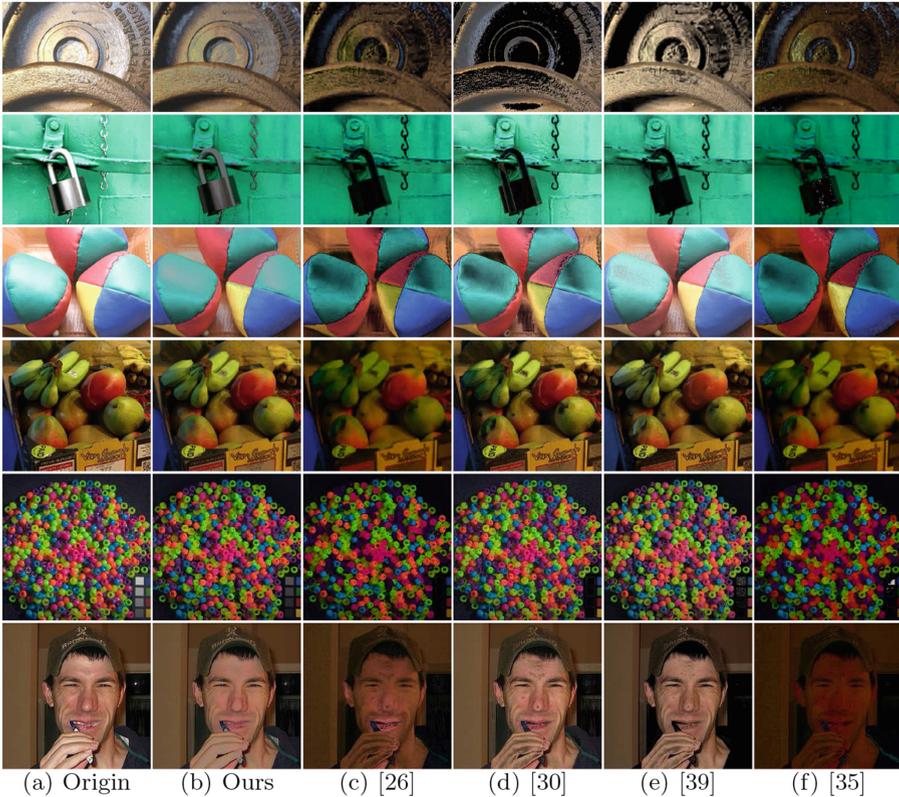
**Fig. 5.** Influence of strong noises. Zero-mean Gaussian noises with variance 0.1 are added to each channel of a real-world image. Some methods (*e.g.*, [28], [30], and [39]) are vulnerable to strong noises while our method is much more robust.



**Fig. 6.** Influence of the response function of the camera. Each pixel value  $\mathbf{I}$  of the original images is transformed to  $\mathbf{I}^{1/\gamma}$ .

### 5.4 Test on Images in the Wild

Qualitative results on images in the wild are provided in Fig. 7 where we compare our method with those of Yang *et al.* [26], Shen and Zheng [30], Akashi and Okatani [39], and Ren *et al.* [35]. We do not provide quantitative comparisons since ground truth results are unavailable for these images. Overall, our method performs well on a diversity of natural images which may contain various materials, heavy textures, occlusion, overexposure and natural illumination. Note that many methods produce obvious visual artifacts in those regions having a close-to-white diffuse reflectance (*e.g.* the labels of cardboard boxes in the fourth scene and the color palette in the fifth scene) such that the diffuse chromaticity is almost identical to the illumination chromaticity. As shown, our method can still handle this challenging situation without introducing annoying artifacts. In comparison, our method well suppresses the specular highlights and preserves the saturation and image details as much as possible. Results for additional real-world images are provided in the supplemental material.

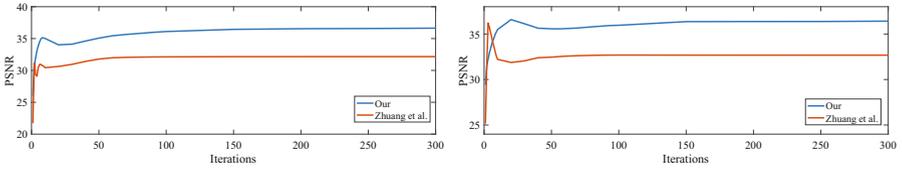


**Fig. 7.** More results of highlight removal on images in the wild which may contain natural illumination, heavy textures, hue-saturation ambiguity and overexposed pixels.

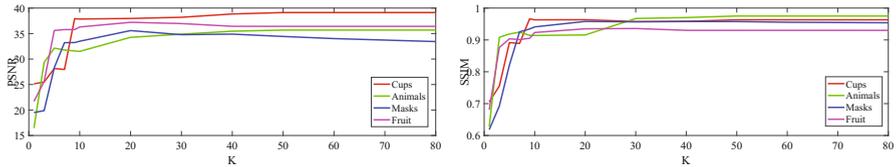
## 5.5 Study on Convergence

In this work, we use slack variables to explicitly convert non-negativity constraints into equality constraints. This leads to a fast convergence rate as evidence in Fig. 8. Here we show two examples. Compared to Zhuang *et al.* [41]’s strategy which simply clamps the negative entries to zero, our method converges faster in both cases. Empirically, 200 ~ 300 iterations are sufficient for most images. Increasing the iteration number has little effect on the final results once the algorithm converges.

We also test the influence of the color dictionary size  $K$  on the convergence rate in Fig. 9. Here, we provide the results of four real world images that have ground truths. We see that a very small  $K$  leads to low accuracy for diffuse color reconstruction. Both PSNR and SSIM obtain an obvious improvement as varying  $K$  from 1 to 10. After this, the scores of PSNR and SSIM almost saturate when  $K$  increases from 10 to 80. In practice, although high PSNR and SSIM scores could be obtained when  $K$  is relatively small (e.g.,  $K = 10$ ), the visual quality of reconstructed image is not satisfying. For a much larger  $K$  (e.g., from 60 to 80) it hardly affects the numerical results (PSNR, SSIM) as well as visual quality, but will significantly increase the computational cost. Taking these factors into account, we generally set  $K$  as 50 without any specific tuning.



**Fig. 8.** Convergence rate comparisons with Zhuang *et al.* [41]’s strategy on two images: Animals (left) and Fruit (right).



**Fig. 9.** Influence of  $K$  on the convergence rate. The line plots show the PSNR (left) and SSIM (right) w.r.t.  $K$  after 300 iterations.

## 6 Conclusion

We have presented a new method for automatically separating an original image into a diffuse component and a specular component with no user intervention. This method is built upon a sparse and low-rank reflection (SLRR) model in which we assume that highlight regions in an image are sparse while diffuse colors can be well represented by a limited number of distinct colors. We then cast the task of highlight removal into a constrained nuclear norm and  $l_1$ -norm minimization problem which can be solved effectively by the ALM method. Different from some previous work, our method do not require a specular-free image which is vulnerable to hue-saturation ambiguity. With the SLRR model and proper parameter settings, we tend to avoid this problem and preserve scene details as much as possible. Experimental results using various synthetic and real-world images validate the accuracy and robustness of our method.

**Acknowledgement.** We would like to thank the anonymous reviewers for their valuable feedback. This work was partially supported by the National Key Research and Development Program of China (No. 2018YFB1004901) and NSFC (No. 61502223).

## References

1. Nayar, S.K., Fang, X.S., Boult, T.: Separation of reflection components using color and polarization. *Int. J. Comput. Vision* **21**(3), 163–186 (1997)
2. Kim, D.W., Lin, S., Hong, K.S., Shum, H.Y.: Variational specular separation using color and polarization. In: *Proceedings of the IAPR Conference on Machine Vision Applications*, pp. 176–179 (2002)
3. Umeyama, S., Godin, G.: Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(5), 639–647 (2004)
4. Wang, F., Ainouz, S., Petitjean, C., Bensrhair, A.: Specularity removal: a global energy minimization approach based on polarization imaging. *Comput. Vis. Image Underst.* **158**(Suppl. C), 31–39 (2017)
5. Kobayashi, N., Okabe, T.: Separating reflection components in images under multispectral and multidirectional light sources. In: *ICPR*, pp. 3210–3215, December 2016
6. Shafer, S.A.: Using color to separate reflection components. *Color Res. Appl.* **10**(4), 210–218 (1985)
7. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: *CVPR* (2016)
8. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *J. ACM* **58**(3), 11:1–11:37 (2011)
9. Chandrasekaran, V., Sanghavi, S., Parrilo, P.A., Willsky, A.S.: Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21**(2), 572–596 (2011)
10. Zhou, X., Yang, C., Zhao, H., Yu, W.: Low-rank modeling and its applications in image analysis. *ACM Comput. Surv.* **47**(2) (2015)
11. Lee, H.C., Breneman, E.J., Schulte, C.P.: Modeling light reflection for computer color vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(4), 402–409 (1990)

12. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: ICML, pp. 663–670 (2010)
13. Zhang, Y., Jiang, Z., Davis, L.S.: Learning structured low-rank representations for image classification. In: CVPR, June 2013
14. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, UIUC (2009)
15. Artusi, A., Banterle, F., Chetverikov, D.: A survey of specular removal methods. *Comput. Graph. Forum* (2011)
16. Lin, S., Li, Y., Kang, S.B., Tong, X., Shum, H.-Y.: Diffuse-specular separation and depth recovery from image sequences. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 210–224. Springer, Heidelberg (2002). [https://doi.org/10.1007/3-540-47977-5\\_14](https://doi.org/10.1007/3-540-47977-5_14)
17. Weiss, Y.: Deriving intrinsic images from image sequences. In: ICCV, pp. 68–75 (2001)
18. Lin, S., Shum, H.Y.: Separation of diffuse and specular reflection in color images. In: CVPR, pp. 341–346 (2001)
19. Sato, Y., Ikeuchi, K.: Temporal-color space analysis of reflection. *J. Opt. Soc. Am. A* **11**(11), 2990–3002 (1994)
20. Feris, R., Raskar, R., Tan, K.H., Turk, M.: Specular reflection reduction with multi-flash imaging. *SIBGRAP I*, 316–321 (2004)
21. Agrawal, A., Raskar, R., Nayar, S.K., Li, Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans. Graph.* **24**(3), 828–835 (2005)
22. Klinker, G.J., Shafer, S.A., Kanade, T.: The measurement of highlights in color images. *Int. J. Comput. Vision* **2**(1), 7–32 (1988)
23. Schlüns, K., Teschner, M.: Analysis of 2D color spaces for highlight elimination in 3D shape reconstruction. In: ACCV, 801–805 (1995)
24. Bajcsy, R., Lee, S.W., Leonardis, A.: Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *Int. J. Comput. Vision* **17**(3), 241–272 (1996)
25. Tan, R.T., Ikeuchi, K.: Separating reflection components of textured surfaces using a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(2), 178–193 (2005)
26. Yang, Q., Wang, S., Ahuja, N.: Real-time specular highlight removal using bilateral filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 87–100. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15561-1\\_7](https://doi.org/10.1007/978-3-642-15561-1_7)
27. j. Yoon, K., Choi, Y., Kweon, I.S.: Fast separation of reflection components using a specularity-invariant image representation. In: ICIP, pp. 973–976 (2006)
28. Shen, H.L., Zhang, H.G., Shao, S.J., Xin, J.H.: Chromaticity-based separation of reflection components in a single image. *Pattern Recogn.* **41**(8), 2461–2469 (2008)
29. Shen, H.L., Cai, Q.Y.: Simple and efficient method for specular removal in an image. *Appl. Opt.* **48**(14), 2711–2719 (2009)
30. Shen, H.L., Zheng, Z.H.: Real-time highlight removal using intensity ratio. *Appl. Opt.* **52**(19), 4483–4493 (2013)
31. Kim, H., Jin, H., Hadap, S., Kweon, I.: Specular reflection separation using dark channel prior. In: CVPR, pp. 1460–1467, June 2013
32. Suo, J., An, D., Ji, X., Wang, H., Dai, Q.: Fast and high quality highlight removal from a single image. *IEEE Trans. Image Process.* **25**(11), 5441–5454 (2016)
33. Liu, Y., Yuan, Z., Zheng, N., Wu, Y.: Saturation-preserving specular reflection separation. In: CVPR, pp. 3725–3733 (2015)

34. Mallick, S.P., Zickler, T., Belhumeur, P.N., Kriegman, D.J.: Specularity removal in images and videos: a pde approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 550–563. Springer, Heidelberg (2006). [https://doi.org/10.1007/11744023\\_43](https://doi.org/10.1007/11744023_43)
35. Ren, W., Tian, J., Tang, Y.: Specular reflection separation with color-lines constraint. *IEEE Trans. Image Process.* **26**(5), 2327–2337 (2017)
36. Li, C., Lin, S., Zhou, K., Ikeuchi, K.: Specular highlight removal in facial images. In: CVPR, July 2017
37. Tan, P., Lin, S., Quan, L., Shum, H.Y.: Highlight removal by illumination-constrained inpainting. In: ICCV, pp. 164–169, October 2003
38. Tan, P., Quan, L., Lin, S.: Separation of highlight reflections on textured surfaces. In: CVPR, pp. 1855–1860 (2006)
39. Akashi, Y., Okatani, T.: Separation of reflection components by sparse non-negative matrix factorization. *Comput. Vis. Image Underst.* **146**(C), 77–85 (2016)
40. Tan, T.T., Nishino, K., Ikeuchi, K.: Illumination chromaticity estimation using inverse-intensity chromaticity space. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I-673–I-680 (2003)
41. Zhuang, L., Gao, H., Lin, Z., Ma, Y., Zhang, X., Yu, N.: Non-negative low rank and sparse graph for semi-supervised learning. In: CVPR, pp. 2328–2335 (2012)
42. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20**(4), 1956–1982 (2010)
43. Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., Zhang, L.: Weighted nuclear norm minimization and its applications to low level vision. *Int. J. Comput. Vision* **121**(2), 183–208 (2017)