PAPER

White Balancing by Using Multiple Images via Intrinsic Image Decomposition*

Ryo MATSUOKA†, Tatsuya BABA†, *Student Members*, Mia RIZKINIA†, *Nonmember*, *and* Masahiro OKUDA†, *Member*

SUMMARY Using a flash/no-flash image pair, we propose a novel white-balancing technique that can effectively correct the color balance of a complex scene under multiple light sources. In the proposed method, by using multiple images of the same scene taken under different lighting conditions, we estimate the reflectance component of the scene and the multiple shading components of each image. The reflectance component is a specific object color which does not depend on scene illumination and the shading component is a shading effect caused by the illumination lights. Then, we achieve white balancing by appropriately correcting the estimated shading components. The proposed method achieves better performance than conventional methods, especially under colored illumination and mixed lighting conditions.

key words: *white balance, intrinsic image, decomposition, sparsity, flash image, ℓ_0 norm*

1. Introduction

White balancing is an important tool to correct the chrominance of images in order to replicate the color consistency of the human visual system. Several white balancing techniques have been proposed [?], [?], [?], [?,]. Many studies, however, did not focus on mixed lighting conditions. Namely, they assume that the photograph was taken under a single illuminant source. Recently, most commercial cameras and photo editing tools involve some practical functionality to recover natural white balance. Most of them, however, cannot realize adequate white balancing, especially under the colored illumination and mixed lighting conditions. Indoor scenes often have colored illumination of artificial light, or mixed illumination of artificial and natural lights. Several reported techniques for the white balancing under the mixed lighting conditions have been proposed [?], [?], [?], [?,]. However, the methods require user interaction or are based on restrictive assumptions. For example, in [?,], to correct localized color casts, they use a scribble interface and achieve an effective correction. In contrast, Hsu et al. [?,] proposed a semi-automatic white balancing technique for real scenes with two light types, but they assume that the illuminant colors are known. Although a recently proposed method [?,] achieves a better performance than the other methods, it still requires user interaction.

Manuscript received January 1, 2011.
Manuscript revised January 1, 2011.

†The authors are with the Faculty of Environmental Engineering, the University of Kitakyushu, Fukuoka, 808–0135, Japan.

*This paper is partially based on the conference paper, which is submitted to the International Conference on Image Processing at year 2015.

DOI: 10.1587/transinf.E0.D.1

Barrow and Tenenbaum [?,] introduced the technique of the intrinsic image decomposition to separate the reflectance component and the shading component, two components that represent the observed image with Lambertian surfaces. The reflectance component is a specific object color that does not depend on scene illumination unlike the shading component, which is a shading effect caused by the illumination lights. Based on their theory, many intrinsic image decomposition techniques have been proposed by a number of authors [?], [?], [?], [?], [?], [?], [?], [?], [?,], and it is useful for several applications in computer vision, such as scene illumination transfer [?,] and colorization [?,]. However, intrinsic image decomposition is a major challenge due to the ill-posedness that there are two unknown components for one observed image. In addition, many conventional methods assume a single white illuminant condition. Under the mixed lighting conditions, the color of the illuminant sources are intricately mixed on each point of a scene object surface, which makes it a more challenging problem than the case of a single illuminant.

Meanwhile, image processing with the flash/no-flash image pair [?], [?], [?,] have been actively studied and attracted attention as an effective method to overcome the performance limitation of classical single image-based methods. In these methods, the noise-free flash image taken by an electronic flash is utilized as a reference image to restore the noisy no-flash image. In addition, by estimating the scene illumination color from a flash/no-flash image pair, the method in [?,] achieves white balancing. However, this method fails and produces an unexpected color artifact under the complex lighting conditions.

In this paper, we present a novel approach for the reflectance estimation. In general, the contribution estimation of each light from a single image is a severely ill-posed problem. We overcome this difficulty by utilizing a flash image as a reference. Our technique estimates the reflectance component of the specific object color, and the shading components from a flash/no-flash image pair. Then we achieve the white balance correction by appropriately correcting the estimated shading components. The proposed method achieves good performance, especially under colored illumination and mixed lighting conditions.

In Section 2, we discuss the intrinsic image decomposition that becomes a key technique of the proposed method. Our decomposition problem is formulated by the optimization problem with $\ell_{0,1}$ norm, furthermore, a decomposition algorithm that estimates the reflectance and shading components is proposed. In Section 3, several examples are shown to verify the validity of the proposed algorithm, and to compare our work to several conventional methods. In the last Section 4, we briefly conclude this paper.

Notation: Our method mainly consists of two steps. We treat images in the log domain in the first step, and linear domain in the second step. In the manuscript, we summarize the notation for images $\mathbf{x} = \{\mathbf{q}, \mathbf{r}, \mathbf{s}_1 \mathbf{s}_2\}$ as follows

- $\mathbf{x}$: images in the log domain
- $\hat{\mathbf{x}}$: the linear version of $\mathbf{x}$, that is, $\mathbf{x} = \log \hat{\mathbf{x}}$
- $\bar{\mathbf{x}}$, $\tilde{\mathbf{x}}$: images in the linear domain, used in the second step

2. Intrinsic Image Decomposition

The intrinsic image model [?,] assumes that an image scene is the product of a scene's reflectance (also called albedo) and shading (or illuminant) at each pixel, expressed as $\hat{\mathbf{q}} = \hat{\mathbf{r}} \otimes \hat{\mathbf{s}}$ where $\hat{\mathbf{q}} = [\hat{\mathbf{q}}_R^\top\ \hat{\mathbf{q}}_G^\top\ \hat{\mathbf{q}}_B^\top]^\top \in \mathbb{R}^{3N}$ is a vectorized observed color image, where N is the number of the pixels, and $(\cdot)^\top$ stands for the transposition of $(\cdot)$. Also, $\hat{\mathbf{r}} \in \mathbb{R}^{3N}$ is the reflectance, and $\hat{\mathbf{s}} \in \mathbb{R}^{3N}$ is the shading. The operator $\otimes$ is a pixel-wise multiplication. The intrinsic image decomposition aims to estimate $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, given an input image $\hat{\mathbf{q}}$. This can be reformulated by taking the log of the images:

$$\mathbf{q} = \mathbf{r} + \mathbf{s}, \tag{1}$$

where $\mathbf{q} = \log \hat{\mathbf{q}}$ and so on. Our goal is to estimate the reflectance and shading components and to correct its white balance using them. This decomposition is inherently a challenging problem since the equation (1) is severely underdetermined. One solution is to apply tractable prior knowledge to solve the problem [?], [?], [?,]. Their methods are based on a simple assumption that the reflectance has piecewise constant region with sharp edges, whereas the shading component smoothly varies between pixels.

The proposed white balance correction mainly consists of two steps:

1. Estimate a single reflectance $\mathbf{r}$ and two shading components $\mathbf{s}_1, \mathbf{s}_2$, which correspond to a flash/no-flash image pair $\mathbf{q}_1, \mathbf{q}_2$, respectively.
2. Appropriately correct the estimated shading component $\mathbf{s}_2$ of the no-flash image $\mathbf{q}_2$ to eliminate illuminant colors, and then simply conduct the white balancing by adding the corrected shading component to the reflectance $\mathbf{r}$.

Figure 1 shows the flow of our method.

2.1 Proposed intrinsic image decomposition problem

The first step decomposes two input images into the reflectance and shading components. We assume that the inputs are well aligned and no further registration or motion compensation is needed, and the two inputs have a common reflectance. We find a single reflectance component and two

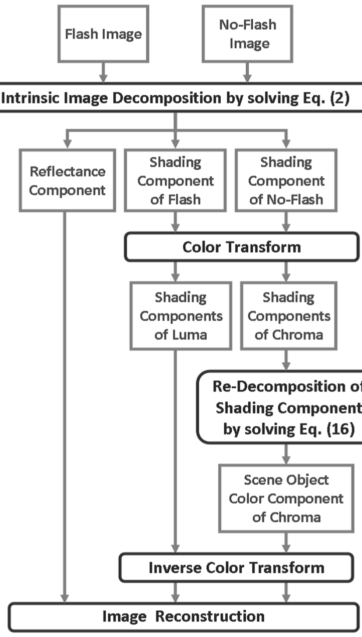

Fig. 1 The flow chart of the proposed method.

shading components by minimizing the cost function:

$$\min_{\mathbf{r},\mathbf{s}_1,\mathbf{s}_2} \|\mathbf{D}\mathbf{r}\|_{0,1} + \sum_{i=1}^{2} w_{si}\|\mathbf{L}\mathbf{s}_i\|_2^2 + \sum_{i=1}^{2} w_{fi}\|\mathbf{q}_i - (\mathbf{r} + \mathbf{s}_i)\|_2^2,$$
$$\text{s.t. } b_j \le r_j \le t_j,\quad b_j \le s_{1j} \le t_j,\quad b_j \le s_{2j} \le t_j,\ \text{for } \forall j, \tag{2}$$

where r_j, s_{1j}, and s_{2j} are a j-th elements of $\mathbf{r} = [\mathbf{r}_R^\top\ \mathbf{r}_G^\top\ \mathbf{r}_B^\top]^\top \in \mathbb{R}^{3N}$, $\mathbf{s}_1 \in \mathbb{R}^{3N}$ and $\mathbf{s}_2 \in \mathbb{R}^{3N}$ respectively. The two inputs $\mathbf{q}_1 \in \mathbb{R}^{3N}$ and $\mathbf{q}_2 \in \mathbb{R}^{3N}$ are the flash and no-flash images respectively. $\mathbf{L} = \mathrm{diag}\{\mathbf{L}', \mathbf{L}', \mathbf{L}'\} \in \mathbb{R}^{3N\times 3N}$ is a convolution matrix representing a laplacian operator $\mathbf{L}' \in \mathbb{R}^{N\times N}$, $\mathbf{D} = \mathrm{diag}\{\mathbf{D}', \mathbf{D}', \mathbf{D}'\} \in \mathbb{R}^{6N\times 3N}$ consists of the vertically concatenated first-order differential operators $\mathbf{D}' = [\mathbf{D}_h^\top\ \mathbf{D}_v^\top]^\top \in \mathbb{R}^{2N\times N}$ with horizontal $\mathbf{D}_h \in \mathbb{R}^{N\times N}$ and vertical operators $\mathbf{D}_v \in \mathbb{R}^{N\times N}$. The norm for the vectorized color image gradients $\|\mathbf{D}\mathbf{r}\|_{0,1}$ is defined by the operator $C(m)$, which returns 0 if m is 0, and 1 otherwise, by

$$\|\mathbf{D}\mathbf{r}\|_{0,1} = \sum_{n=1}^{N} C\Big(|\partial_x r_{Rn}| + |\partial_y r_{Rn}| + |\partial_x r_{Gn}| + |\partial_y r_{Gn}| + |\partial_x r_{Bn}| + |\partial_y r_{Bn}|\Big), \tag{3}$$

where n is a pixel index, r_{Rn}, r_{Gn} and r_{Bn} are the n-th RGB channels of $\mathbf{r}$ respectively. Our feature is that we relax the relationship (1) by allowing some reconstruction error and directly find the two shading components, and we use the $\ell_{0,1}$ norm[†] in the first term to treat the RGB channels simultaneously. To take account of the properties of the locally flat reflectance, we introduce the ℓ_0 based term in (2). Instead of the simple ℓ_0 norm, we use the $\ell_{0,1}$ norm considering the sparseness of the gradients of all the three color

[†]The $\ell_{0,1}$ norm is essentially same as $\ell_{0,2}$ norm introduced in [?,], and also a simple extended version of ℓ_0 norm which is utilized for image smoothing in [?,].

channels. By introducing the $\ell_{0,1}$ norm, fake color artifact due to violation of the color balance is relieved, which is important treatment especially for our white balance application. The second term is introduced to satisfy the properties of the shading whose gradient gradually varies. The third term penalizes the decomposition error. To obtain a meaningful solution for $\mathbf{r}$, $\mathbf{s}_1$ and $\mathbf{s}_2$, we consider the constrained problem with the specific range constraint for each pixel of the three images.

Since the cost function is non-convex due to the $\ell_{0,1}$ norm, and likewise there is an inequality constraint, it is impossible to solve it by conventional gradient-based methods. To solve the problem, we introduce auxiliary variables and adopt the penalty function method. By introducing the auxiliary variables $\mathbf{z}_i (i=1,2,3,4)$, the cost function to minimize in each iteration of the algorithm is given by

$$\begin{aligned}
&\min_{\mathbf{r},\mathbf{s}_{1,2},\mathbf{z}_{1,2,3,4}} f(\mathbf{r},\mathbf{s}_{1,2},\mathbf{z}_{1,2,3,4}), \quad \text{where} \\
&f(\mathbf{r},\mathbf{s}_{1,2},\mathbf{z}_{1,2,3,4}) = \|\mathbf{z}_1\|_{0,1} + \sum_{i=1}^{2} w_{si}\|\mathbf{L}\mathbf{s}_i\|_2^2 \\
&+ \sum_{i=1}^{2} w_{fi}\|\mathbf{q}_i - (\mathbf{r}+\mathbf{s}_i)\|_2^2 + \sum_{i=2}^{4} \iota(\mathbf{z}_i) + \alpha\|\mathbf{D}\mathbf{r} - \mathbf{z}_1\|_2^2 \\
&+\alpha\|\mathbf{r}-\mathbf{z}_2\|_2^2 + \alpha\|\mathbf{s}_1 - \mathbf{z}_3\|_2^2 + \alpha\|\mathbf{s}_2 - \mathbf{z}_4\|_2^2. \qquad (4)
\end{aligned}$$

$\iota(\cdot)$ is an indicative function, which is defined for each j-th elements of $\mathbf{x}\in\mathbb{R}^{3N}$:

$$\iota(x_j) = \begin{cases} 0, & \text{if } b_j \le x_j \le t_j \\ +\infty & \text{otherwise} \end{cases}. \qquad (5)$$

The indicative function guarantees that the optimal solution falls in the range $[b_j, t_j]$. The auxiliary variables $\mathbf{z}_i (i=1,2,3,4)$, are introduced for $\mathbf{Dr}$, $\mathbf{r}$, $\mathbf{s}_1$ and $\mathbf{s}_2$, respectively, and then we add the ℓ_2 penalty terms between the four pairs. Here, α is a weight that we increase during iterations of the algorithms. As α gets larger, the solution gets closer to the solution of the original cost function (2). We alternately minimize (4) w.r.t. each of the seven variables $\mathbf{r},\mathbf{s}_1,\mathbf{s}_2,\mathbf{z}_i (i=1,2,3,4)$ with other variables fixed. Overall of this algorithm is roughly shown in Algorithm 1. The solutions for all of these seven sub-problems are readily calculated as follows.

Algorithm 1 Algorithm for solving (4)

1: flash $\hat{\mathbf{q}}_1$, and no-flash image $\hat{\mathbf{q}}_2$ are given, and they are transformed to log domain $\mathbf{q}_1$ and $\mathbf{q}_2$.
2: set $k=0$, and choose the weights w_{si}, w_{fi} $(i=1,2)$ and α, η.
3: Choose $\mathbf{r}^{(0)}, \mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \mathbf{z}_i^{(0)} (i=1,2,3,4)$.
4: **while** a stop criterion is not satisfied **do**
5: $\quad\mathbf{s}_1^{(k+1)} = \arg\min_{\mathbf{s}_1} f(\mathbf{s}_1|\mathbf{r}^{(k)}, \mathbf{s}_2^{(k)}, \mathbf{z}_{1,2,3,4}^{(k)})$
6: $\quad\mathbf{s}_2^{(k+1)} = \arg\min_{\mathbf{s}_2} f(\mathbf{s}_2|\mathbf{r}^{(k)}, \mathbf{s}_1^{(k+1)}, \mathbf{z}_{1,2,3,4}^{(k)})$
7: $\quad\mathbf{r}^{(k+1)} = \arg\min_{\mathbf{r}} f(\mathbf{r}|\mathbf{s}_1^{(k+1)}, \mathbf{s}_2^{(k+1)}, \mathbf{z}_{1,2,3,4}^{(k)})$
8: $\quad\mathbf{z}_{1,2,3,4}^{(k+1)} = \arg\min_{\mathbf{z}_{1,2,3,4}} f(\mathbf{z}_{1,2,3,4}|\mathbf{r}^{(k+1)}, \mathbf{s}_1^{(k+1)}, \mathbf{s}_2^{(k+1)})$
9: $\quad\alpha = \eta\cdot\alpha, \quad k = k+1$
10: **end while**

NOTE: $f(a|b)$ indicates the function of the variable a with given b.

2.1.1 Optimal solution for $\mathbf{r}, \mathbf{s}_1$ and $\mathbf{s}_2$

The sub-problem w.r.t. $\mathbf{s}_1$ in Step 5 of Algorithm 1 is rewritten as follows (superscript k is omitted hereafter):

$$\begin{aligned}
&\min_{\mathbf{s}_1} f(\mathbf{s}_1|\mathbf{r},\mathbf{s}_2,\mathbf{z}_{1,2,3,4}), \quad \text{where} \\
&f(\mathbf{s}_1|\mathbf{r},\mathbf{s}_2,\mathbf{z}_{1,2,3,4}) = w_{s1}\|\mathbf{L}\mathbf{s}_1\|_2^2 + w_{f1}\|\mathbf{q}_1 - (\mathbf{r}+\mathbf{s}_1)\|_2^2 \\
&+\alpha\|\mathbf{s}_1 - \mathbf{z}_3\|_2^2. \qquad (6)
\end{aligned}$$

From (6), the problem w.r.t. $\mathbf{s}_1$ is a simple quadratic form. Thus, by setting the first-order derivative of (6) to zero, the optimal solution is determined by solving

$$(w_{s1}\mathbf{L}^\top\mathbf{L} + (w_{f1}+\alpha)\mathbf{I})\mathbf{s}_1 = w_{f1}(\mathbf{q}_1 - \mathbf{r}) + \alpha\mathbf{z}_3, \qquad (7)$$

where $\mathbf{I}\in\mathbb{R}^{3N\times 3N}$ is an identity matrix.

Similarly, the optimal solution of $\mathbf{s}_2$ is obtained by solving the following equation,

$$(w_{s2}\mathbf{L}^\top\mathbf{L} + (w_{f2}+\alpha)\mathbf{I})\mathbf{s}_2 = w_{f2}(\mathbf{q}_2 - \mathbf{r}) + \alpha\mathbf{z}_4. \qquad (8)$$

The sub-problem w.r.t. $\mathbf{r}$ in Step 7 of Algorithm 1 is rewritten as:

$$\begin{aligned}
&\min_{\mathbf{r}} f(\mathbf{r}|\mathbf{s}_1,\mathbf{s}_2,\mathbf{z}_{1,2,3,4}), \quad \text{where} \\
&f(\mathbf{r}|\mathbf{s}_1,\mathbf{s}_2,\mathbf{z}_{1,2,3,4}) = \sum_{i=1}^{2} w_{fi}\|\mathbf{q}_i - (\mathbf{r}+\mathbf{s}_i)\|_2^2 \\
&+\alpha\|\mathbf{D}\mathbf{r} - \mathbf{z}_1\|_2^2 + \alpha\|\mathbf{r} - \mathbf{z}_2\|_2^2. \qquad (9)
\end{aligned}$$

From (9), the problem w.r.t. $\mathbf{r}$ is also a simple quadratic form. Thus, by setting the first-order derivative of (9) to zero, the optimal solution is determined by solving

$$\begin{aligned}
&(\alpha\mathbf{D}^\top\mathbf{D} + (w_{f1} + w_{f2} + \alpha)\mathbf{I})\mathbf{r} = \\
&\qquad \alpha\mathbf{D}^\top\mathbf{z}_1 + \sum_{i=1}^{2} w_{fi}(\mathbf{q}_i - \mathbf{s}_i) + \alpha\mathbf{z}_2. \qquad (10)
\end{aligned}$$

As described above, the sub-problem for each of $\mathbf{r}, \mathbf{s}_1$ and $\mathbf{s}_2$ is a simple least squares problem whose solutions can be found by solving the linear equation of the form $\mathbf{Ax} = \mathbf{b}$. Since the matrix to solve for the sub-problems (7), (8) and (10) is a block circulant matrix with circulant blocks (BCCB), it is diagonalized by FFT, thereby the solution can be quickly calculated.

2.1.2 Optimal solution for $\mathbf{z}_1$

Note again that the vector $\mathbf{r} = [\mathbf{r}_R^\top \;\mathbf{r}_G^\top \;\mathbf{r}_B^\top]^\top \in \mathbb{R}^{3N}$ is composed of the *RGB* channels of the image, and $\mathbf{Dr}$ is a $6N$-dimensional vector composed of the derivative w.r.t. the horizontal/vertical directions of $\mathbf{r}$. The auxiliary vector $\mathbf{z}_1\in\mathbb{R}^{6N}$, which is introduced in place of $\mathbf{Dr}$, is of the form:

$$\mathbf{z}_1 = [\mathbf{z}_{1R}^{(h)\top} \;\mathbf{z}_{1R}^{(v)\top} \;\mathbf{z}_{1G}^{(h)\top} \;\mathbf{z}_{1G}^{(v)\top} \;\mathbf{z}_{1B}^{(h)\top} \;\mathbf{z}_{1B}^{(v)\top}]^\top.$$

$\mathbf{z}_{1R}^{(h)}$, and $\mathbf{z}_{1R}^{(v)}$ are the red channel of $\mathbf{z}_1$ and so on. The superscript $\cdot^{(h)}$ and $\cdot^{(v)}$ indicates that they corresponds to the horizontal and vertical derivatives, respectively.

From (4). the sub-problem w.r.t. $\mathbf{z}_1$ is rewritten as,

$$\min_{\mathbf{z}_1} f(\mathbf{z}_1|\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{z}_{2,3,4}), \quad \text{where}$$
$$f(\mathbf{z}_1|\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{z}_{2,3,4}) = \|\mathbf{z}_1\|_{0,1} + \alpha\|\mathbf{D}\mathbf{r} - \mathbf{z}_1\|_2^2. \quad (11)$$

The optimal solution of (11) is found for each element individually by applying group hard shrinkage to the total sum of the square of gradients in the *RGB* channels. The problem (11) is equivalent to solving the following cost function for all pixels $n = 1, \cdots, N$.

$$E_n = C\left(\sum_{l=0}^{5} |z_{1(n+lN)}|\right) + \alpha \sum_{l=0}^{5} \left(g_{n+lN} - z_{1(n+lN)}\right)^2, \quad (12)$$

where $z_{1(n+lN)}$ is a $(n+lN)$-th element of $\mathbf{z}_1$, and $z_{1(n+lN)}$, $(l = 0, 1, \cdots, 5)$ corresponds to the derivatives at a pixel n w.r.t. the both directions in the *RGB* channels. g_{n+lN} is the $(n+lN)$-th element of $\mathbf{g} = \mathbf{D}\mathbf{r} \in \mathbb{R}^{6N}$. The first term of (12) returns 0 if all of the derivatives at a pixel n w.r.t. the both directions in all the three channels are 0, and otherwise, returns 1. The second term ensures that the auxiliary valuables $\mathbf{z}_1$ approximate $\mathbf{g}$ at a pixel n. The optimal solution for (12) is obtained by group hard shrinkage (its derivation is found in Appendix):

$$z^*_{1(n+lN)} = \begin{cases} 0, & \text{if } \sum_{l=0}^{5} g^2_{n+lN} \le 1/\alpha \\ g_{n+lN}, & \text{otherwise} \end{cases}. \quad (13)$$

Applying the thresholding operation to every pixel, we obtain the solution of the sub-problem.

2.1.3 Optimal Solution for $\mathbf{z}_{2,3,4}$

To avoid the trivial solution, we add the range constraints for each $\mathbf{r}, \mathbf{s}_1$ and $\mathbf{s}_2$ to the cost function (2). To solve (2), we introduced the indicative function (5) to construct the unconstrained minimization problem (4). Since the indicative function $\iota(\cdot)$ is not differentiable w.r.t. $\mathbf{r}, \mathbf{s}_1$ and $\mathbf{s}_2$. The auxiliary variables $\mathbf{z}_{2,3,4}$ are introduced to solve this problem for variables $\mathbf{r}, \mathbf{s}_1$ and $\mathbf{s}_2$. Here we describe the solution for the case of $\mathbf{z}_2$. The same procedure can be applied to $\mathbf{z}_3$ and $\mathbf{z}_4$. From (4), the sub-problem w.r.t. $\mathbf{z}_2$ is rewritten as follows:

$$\min_{\mathbf{z}_2} f(\mathbf{z}_2|\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{z}_{1,3,4}), \quad \text{where}$$
$$f(\mathbf{z}_2|\mathbf{r}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{z}_{1,3,4}) = \iota(\mathbf{z}_2) + \alpha\|\mathbf{z}_2 - \mathbf{r}\|_2^2. \quad (14)$$

The optimal solution of (14) is found for each j-th element individually, that is,

$$z^*_{2j} = \begin{cases} t_j, & \text{if } r_j > t_j \\ r_j, & \text{if } b_j \le r_j \le t_j \\ b_j, & \text{if } r_j < b_j \end{cases}. \quad (15)$$

Similarly, by replacing the variable r_j with s_{1j} or s_{2j}, the

(a) Example 1: Puppets (b) Example 2: Toys (c) Example 2: Flower

Fig. 2 Scenes with multiple light sources used in our experiment.

optimal solution of z_{3j} or z_{4j} is also introduced. As aforementioned, $\mathbf{z}_{2,3,4}$ can be quickly updated. One can refer the literatures [**?**], [**?**], [**?**,] for the detail of the sub-problems.

2.2 White Balance Correction

In the previous section, the reflectance and shading components are calculated by solving the proposed decomposition problem. However, due to its insufficient estimation accuracy, object colors may still remain in the estimated shading components. Next, we discuss about the detail of the proposed white balance correction. We assume that the scene illumination contains one or a few dominant colors, and the chrominance of the shading has one or a few dominant values. Based on this assumption, we attempt to remove undesired colors from the shading components. We transform the shading components in *the linear scale* $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ to the *YUV* color space. The two chrominance components in the *YUV* color space are denoted by $\hat{\mathbf{s}}_1^U$, $\hat{\mathbf{s}}_1^V$, and so on. Then we decompose each of the U and V components by using

$$\min_{\mathbf{d}^U, \bar{\mathbf{s}}_1^U, \bar{\mathbf{s}}_2^U} \|\mathbf{d}^U\|_0 + \sum_{k=1}^{2} \lambda_k \|\mathbf{L}\bar{\mathbf{s}}_k^U\|_2^2 + \sum_{k=1}^{2} \beta_k \|\hat{\mathbf{s}}_k^U - (\mathbf{d}^U + \bar{\mathbf{s}}_k^U)\|_2^2, \quad (16)$$

where $\mathbf{d}^U \in \mathbb{R}^N$ is the chrominance component of a scene object, $\bar{\mathbf{s}}_1^U, \bar{\mathbf{s}}_2^U \in \mathbb{R}^N$ are the chrominance components of the shading component for a flash and a no-flash image respectively. $\hat{\mathbf{s}}_k^U (k = 1, 2)$ is the linear-domain version of $\mathbf{s}_k^U$ obtained in Sect.2.1. Here, we assume that the estimated shading component includes an object color information, and we remove it by using ℓ_0-based smoothing. We introduce the second term based on the prior on the shading component. The third term guarantees that the decomposition error is satisfactory small. The same procedure is applied for V components to obtain $\mathbf{d}^V$, $\bar{\mathbf{s}}_1^V$, and $\bar{\mathbf{s}}_2^V$. This cost function is also non-convex due to the ℓ_0 norm, and thus we solve it by the penalty function method, which is similar to the procedure in the previous section. The solution is quickly obtained by iteratively applying hard shrinkage and the least squares method implemented with FFT.

Once the solution for (16) is obtained, the set of smoothed chrominance $\mathbf{d}^U$, $\mathbf{d}^V$, and the illuminance of $\hat{\mathbf{s}}_2$ is transformed to the *RGB* space (denoted by $\tilde{\mathbf{s}}_2$). Then a final white-balanced result $\tilde{\mathbf{q}}$ is obtained by the product of $\hat{\mathbf{r}}$ and $\tilde{\mathbf{s}}_2$, where $\hat{\mathbf{r}}$ is the image in the linear domain obtained in Sect.2.1.

(a) Input flash/no-flash image pair

(b) Estimated shading components

Fig. 3 Weiss [?,] + Retinex [?,]: the result is obtained using a flash/no-flash image pair.

3. Experimental Results

We show the validity of the proposed method by applying it to a variety of scenes taken by under the mixed lighting conditions, which are shown in Fig.2. Using multiple images taken by different lighting conditions with same scene makes the decomposition easier than a single image. Weiss [?,] also takes advantage of it, and Grosse et al.'s paper [?,] shows that the method [?,] with Retinex algorithm [?,] outperforms other conventional methods. However, as shown in Fig.3, which is derived by Weiss's algorithm with Retinex [?,][†], the algorithm often fails since only the two images are inadequate and require a large amount of images. Moreover, as the method handles only edges, it does not work well when input images have different colors like flash/no-flash images. Thus we adopt [?,] and a modified version of Li et al.'s method [?,] for comparison, which is described in the next example.

Since it requires a heavy effort (or even impossible in many cases) to obtain ground truth of reflectance components, it is difficult to precisely perform a quantitative comparison. Meanwhile, white balancing under colored and multiple light sources is an appropriate application to evaluate the preciseness of the intrinsic image decomposition, since a precise decomposition will cancel color artifacts caused by the light. In this section, we show some comparison with figures for the decomposition and white balancing.

3.1 Example 1

First we apply our method to a flash/no-flash image pair (shown in Fig.3(a)), which is used in [?,]. We compare our result with the recently proposed image decomposition method [?,]. Although this method [?,] is not designed for white balancing, it is reasonable to compare with it to show the validity of our algorithm. For fair comparison, we slightly change the method [?,] to handle two inputs, i.e., the minimization problem used in the method is modified to

$$\min_{\mathbf{r}} \rho(\mathbf{Dr}) + \lambda \sum_{i=1}^{2} \|\mathbf{L}(\mathbf{q}_i - \mathbf{r})\|_2^2, \text{ s.t. } b_j \le r_j \le t_j, \text{ for } \forall j,$$

where $\mathbf{r} \in \mathbb{R}^{3N}$ is a reflectance component, $\mathbf{q}_1 \in \mathbb{R}^{3N}$ and $\mathbf{q}_2 \in \mathbb{R}^{3N}$ are the flash and no-flash images respectively. $\mathbf{L} = \mathrm{diag}\{\mathbf{L}', \mathbf{L}', \mathbf{L}'\} \in \mathbb{R}^{3N\times 3N}$ is a convolution matrix representing a laplacian operator $\mathbf{L}' \in \mathbb{R}^{N\times N}$, $\mathbf{D} = \mathrm{diag}\{\mathbf{D}', \mathbf{D}', \mathbf{D}'\} \in \mathbb{R}^{6N\times 3N}$ consists of the vertically concatenated first-order difference operators. Li and Brown [?,] assumed that the gradient probability of a reflectance component can be approximated by the Gaussian-like distribution with a long tail, and introduced the function $\rho(x) = \min\{x^2/k, 1\}$, where x is a gradient value and k is a small constant value (they set 10^{-4} to k). This problem is essentially the two-input version of [?,]. The first term is a regularization term of a reflectance component with $\rho(\cdot)$. The second term is a regularization term of each shading component (from (1), $\mathbf{s}_i = \mathbf{q}_i - \mathbf{r}$). After this optimization, we obtain results of the conventional method by adopting only the luminance of the obtained shading component (i.e., the color components U, V are eliminated in the YUV color space) and then adding it back to the reflectance.

Figure 4 shows the results of the reflectance $\hat{\mathbf{r}}$, and the shading $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$. The estimated reflectance component by our method has the vivid object color without the illumination color. In contrast, the result of the conventional method has the reddish illumination color, especially the face of puppet on the right. The proposed method has a better decomposition performance than the conventional method.

In our white-balance technique, we extract the scene object color from the estimated shading components, and then we reconstruct the final white-balanced image by the only luminance of the shading component, while keeping the extracted scene color in the second step. Here we show our second step results in Fig.5. From our decomposition results of Fig.4, we can find that the red color of the puppet's cheek on the right is included in the shading components. In Fig.5, this red color components, which must be kept in the white-balancing process, has been extracted as the chrominance component of a scene object (i.e. $\mathbf{d}^U$ and $\mathbf{d}^V$).

Figure 6 shows the final white-balanced result, and its close-up. In the conventional method, the scene object color is seriously faded, especially in the face of puppet on the right. Additionally, the reddish illumination color remains in the result. In contrast, our method realizes more natural white balancing.

Figure 7 illustrates the convergence plot for the values of the cost (2). From Fig.7, we confirm that the algorithm converges after a few dozens of iterations. In order to improve its numerical stability, we set $\eta = 1.6$ in Step 9 of the Algorithm 1.

3.2 Example 2

In this experiment, we prepare a pair of input images with

[†] We use the author-provided software [?,].

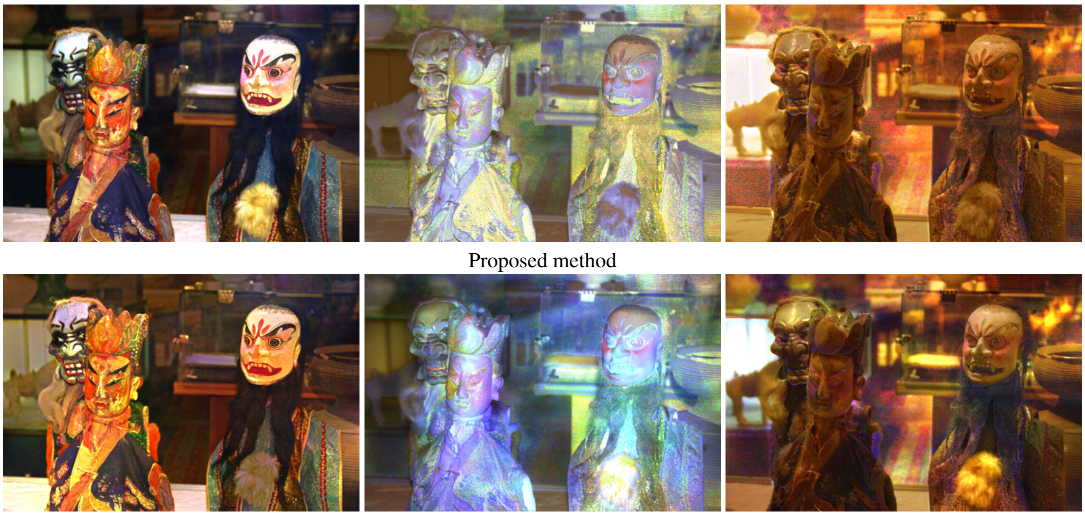

Fig. 4 The intrinsic image decomposition result of Example 1: (left to right) Reflectance, Shading of flash image, Shading of no-flash image.

no positional displacement by using CANON EOS 20D and a tripod. The white balance setting of the camera is fixed to the auto white balance (AWB) mode, except for the manual white balance results in Fig.9. In Fig.2(b)-(c), we show the images taken under the mixed lighting conditions. This scene has two or three different color light sources. From Fig.2(b)-(c), it can be seen that the AWB mode is inadequate under the complex lighting conditions.

We demonstrate the decomposition results by the proposed method in Fig.8. These scenes consist of the multiple colored illumination sources. The scene of the ”Toys” contains reddish illumination, bluish fluorescent lamps and sunlight. The scene of the ”Flower” contains reddish and greenish illumination. From Fig.8, the estimated shading components have these colored illumination. In contrast, the reflectance component retains the original color of the scene objects without the illumination colors. Our method can estimate the reflectance and shading components from the flash/no-flash image pair with high accuracy.

For comparison, we take an image with in-camera manual white balancing (MWB) mode, which estimates a reference point using an image of a white object photographed in advance. We also compared our white balancing results with the statistical color transfer of Pitié et al.’s method [?,]. In [?,], the color statistics of a source image are transformed to the color statistics of a reference image. To achieve the white balancing by [?,], we set the no-flash image as a source image and the flash image as a reference image. Figure 9 shows the results of the conventional method [?,], the MWB mode, [?,] and proposed method. The reddish color is remained in the results of the MWB, [?,] and [?,] for the ”Toys” result. Also, the greenish and reddish colors are remained in the results of the MWB and [?,] for the ”Flower” result. Especially, significant differences appear on the white petals and the body of the red color objects. Furthermore, in the result of [?,], the original color of the scene object is faded. In contrast, our method can eliminate the color of illumination more than the others, while maintaining the vivid color of objects.

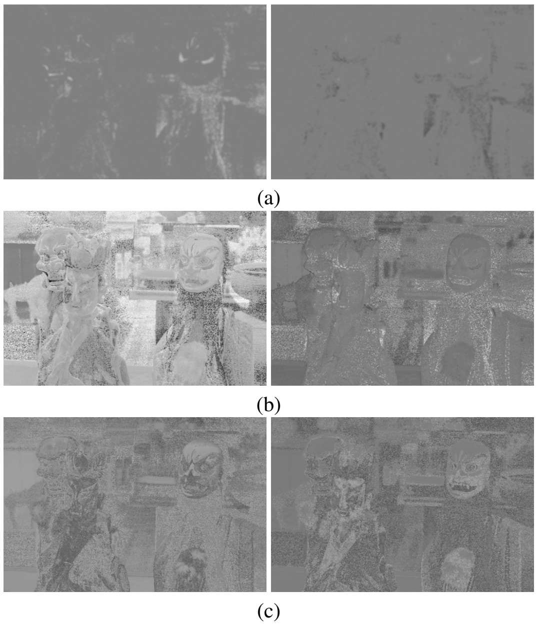

Fig. 5 The results of the second step: (a) $\mathbf{d}^U$ and $\mathbf{d}^V$, (b) $\bar{\mathbf{s}}_1^U$ and $\bar{\mathbf{s}}_1^V$, (c) $\bar{\mathbf{s}}_2^U$ and $\bar{\mathbf{s}}_2^V$.

4. Conclusion

In this paper, we proposed a novel white balance correction technique. The proposed method adopts a two-step approach. In the first step, we estimate the reflectance and shading components from the flash/no-flash image pair by applying intrinsic image decomposition. In the second step, we eliminate the color component of each estimated shad-

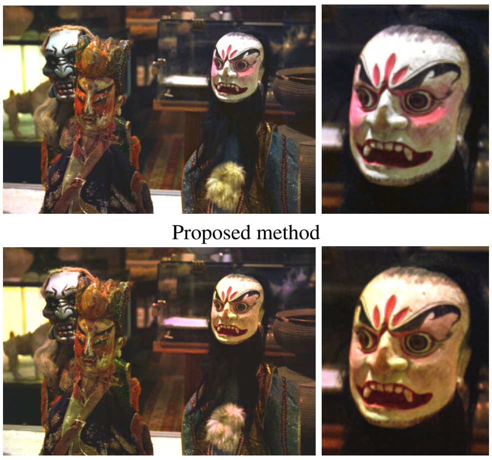

Fig. 6 The final white-balanced result of Example 1: (left to right) White-balanced result, and Close-up of the result.

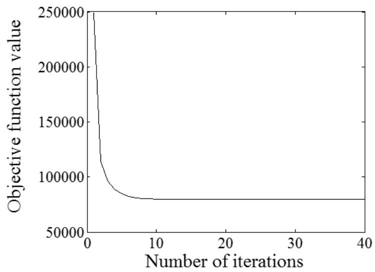

Fig. 7 Evolution of the objective function value of the proposed optimization problem (2) using the "Puppets" as input image pair.

Fig. 8 Example 2: (Left to right) Reflectance, Shading of flash image, and Shading of no-flash image obtained by our method.

ing component. Then we achieve the white balance correction by reconstructing the reflectance component and the appropriately corrected shading component. Experiments showed that the proposed method can achieve better performance under the mixed lighting conditions. Our method outperforms conventional intrinsic image decomposition technique and color correction technique.

Fig. 9 Example 2: The white-balanced results of (top to bottom) MWB, [?,], [?,], and Proposed method.

Acknowledgment

This work was supported in part by JSPS Grants-in-Aid (24560473) and KDDI foundation.

References

[1] G. Buchsbaum, "A spatial processor model for object colour perception," J. Franklin Inst., vol.310, pp.1–26, 1980.
[2] D.H. Brainard and W.T. Freeman, "Bayesian color constancy," Journal of the Optical Society of America A, vol.14, pp.1393–1411, 1997.
[3] J. van de Weijer and T. Gevers, "Color constancy based on the grey-edge hypothesis.," Proc. IEEE ICIP, Genua, Italy, pp.722–725, 2005.
[4] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics," IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2007.
[5] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance.," ACM Trans. Graph., vol.27, no.3, pp.70:1–70:7, 2008.
[6] I. Boyadzhiev, K. Bala, S. Paris, and F. Durand, "User-guided white balance for mixed lighting conditions," ACM Trans. Graph., vol.31, no.6, pp.200:1–200:10, 2012.
[7] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," ACM Transactions on

Graphics, vol.25, no.3, pp.646–653, July 2006.
[8] R. Kawakami, K. Ikeuchi, and R.T. Tan, "Consistent surface color for texturing large objects in outdoor scenes," 10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China, pp.1200–1207, 2005.
[9] H.G. Barrow and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images," Tech. Rep. 157, AI Center, SRI International, Apr 1978.
[10] B.V. Funt, M.S. Drew, and M. Brockington, "Recovering shading from color images," ECCV-92: Second European Conference on Computer Vision, pp.124–132, Springer-Verlag, 1991.
[11] Y. Weiss, "Deriving intrinsic images from image sequences," International Conference on Computer Vision, pp.68–75, 2001.
[12] M. Tappen, W. Freeman, and E. Adelson, "Recovering intrinsic images from a single image," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.27, no.9, pp.1459–1472, 2005.
[13] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," ACM Trans. Graph., vol.28, no.5, pp.130:1–130:10, 2009.
[14] E. Garces, A. Munoz, J. Lopez-Moreno, and D. Gutierrez, "Intrinsic images by clustering," Comp. Graph. Forum, vol.31, no.4, pp.1415–1424, June 2012.
[15] P.Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, "Coherent intrinsic images from photo collections," ACM Trans. Graph., vol.31, no.6, pp.202:1–202:11, Nov. 2012.
[16] Y. Li and M.S. Brown, "Single image layer separation using relative smoothness," Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, pp.2752–2759, IEEE Computer Society, 2014.
[17] X. Liu, L. Wan, Y. Qu, T.T. Wong, S. Lin, C.S. Leung, and P.A. Heng, "Intrinsic colorization," ACM Transactions on Graphics (SIGGRAPH Asia 2008 issue), vol.27, no.5, pp.152:1–152:9, 2008.
[18] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," ACM Trans. Graph., vol.23, no.3, pp.664–672, Aug. 2004.
[19] R. Matsuoka, T. Baba, M. Okuda, and K. Shirai, "High dynamic range image acquisition using flash image," Proc. IEEE ICASSP, pp.1612–1616, 2013.
[20] T. Baba, R. Matsuoka, S. Ono, K. Shirai, and M. Okuda, "Flash/no-flash image integration using convex optimization," Proc. IEEE ICASSP, pp.1185–1189, 2014.
[21] R. Grosse, M.K. Johnson, E.H. Adelson, and W.T. Freeman, "Ground-truth dataset and baseline evaluations for intrinsic image algorithms," International Conference on Computer Vision, pp.2335–2342, 2009.
[22] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," Signal Processing, IEEE Transactions on, vol.58, no.6, pp.3042–3054, June 2010.
[23] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," ACM Trans. Graph., vol.30, no.6, pp.174:1–174:12, 2011.
[24] M.V. Afonso, J.M. Bioucas-Dias, and M.A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," Image Processing, IEEE Transactions on, vol.20, no.3, pp.681–695, 2011.
[25] S. Ono, T. Miyata, I. Yamada, and K. Yamaoka, "Image recovery by decomposition with component-wise regularization," IEICE Transactions, vol.95-A, no.12, pp.2470–2478, 2012.
[26] F. Pitié, A.C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," Comput. Vis. Image Underst., vol.107, no.1-2, pp.123–137, 2007.

Appendix: Derivation of (13)

We denote again the cost function (12)

$$E_n = C\left(\sum_{l=0}^{5} |z_{1(n+lN)}|\right) + \alpha \sum_{l=0}^{5} \left(g_{n+lN} - z_{1(n+lN)}\right)^2 .$$

When $z_{1(n+lN)} \neq 0$ for $l = 0, 1, \ldots, 5$, the cost function is

$$E_n(z_{1(n+lN)}|_{l=0,1,\ldots,5}) = 1 + \alpha \sum_{l=0}^{5} \left(g_{n+lN} - z_{1(n+lN)}\right)^2 .$$

Then, it takes a minimal value when $z_{1(n+lN)} = g_{n+lN}$ for $l = 0, 1, \ldots, 5$, and the cost function is

$$E_n(z_{1(n+lN)}|_{l=0,1,\ldots,5}) = 1 + \alpha \sum_{l=0}^{5} \left(g_{n+lN} - g_{n+lN}\right)^2 = 1 \qquad \text{(A· 1)}$$

When $z_{1(n+lN)} = 0$ for $l = 0, 1, \ldots, 5$,

$$E_n(z_{1(n+lN)}|_{l=0,1,\ldots,5}) = 0 + \alpha \sum_{l=0}^{5} \left(g_{n+lN} - 0\right)^2 = \alpha \sum_{l=0}^{5} g_{n+lN}^2 \qquad \text{(A· 2)}$$

If the cost (A· 1) is larger than (A· 2), the optimal solution for $l = 0, 1, \ldots, 5$ is given by $z^*_{1(n+lN)} = 0$; otherwise, $z^*_{1(n+lN)} = g_{n+lN}$, which yields (13).

Ryo Matsuoka received the B.E., degree from The University of Kitakyushu, Kitakyushu, Japan, in 2013. He has been with the Graduate School of Environmental Engineering, The University of Kitakyushu. He is engaged in multiple image based image processing.

Tatsuya Baba received the B.E., degree from The University of Kitakyushu, Kitakyushu, Japan, in 2014. He has been with the Graduate School of Environmental Engineering, The University of Kitakyushu. He is engaged in guided image based image restoration.

Mia Rizkinia Mia Rizkinia received B.E., and M.E. degree from Universitas Indonesia in 2008 and 2011, respectively. Since 2011, she has worked for Universitas Indonesia as a junior lecturer. She is currently pursing a doctoral degree at the University of Kitakyushu. Her research interests lie in the remote sensing and signal/image processing.

Masahiro Okuda received the B.E., M.E.,

and Dr. Eng. Degrees from Keio University, Yokohama, Japan, in 1993, 1995 and 1998, respectively. He was with the University of California, Santa Barbara, and Carnegie Mellon University, Pittsburgh, PA as a Visiting Scholar in 1998 and 1999, respectively. He has been with Faculty of The University of Kitakyushu, Kitakyushu, Japan, as a Professor of environmental engineering since 2011. His research interests include filter design, vision/geometry coding, and multirate signal processing.