

Separation of Reflection Components by Sparse Non-negative Matrix Factorization

Yasuhiro Akashi, Takayuki Okatani

Tohoku University, Japan

Abstract. This paper presents a novel method for separating reflection components in a single image based on the dichromatic reflection model. Our method is based on a modified version of sparse non-negative matrix factorization (NMF). It simultaneously performs the estimation of body colors and the separation of reflection components through optimization. Our method does not use a spatial prior such as smoothness of colors on the object surface, which is in contrast with recent methods attempting to use such priors to improve separation accuracy. Experimental results show that as compared with these recent methods that use priors, our method is more accurate and robust. For example, it can better deal with difficult cases such as the case where a body color is close to the illumination color.

1 Introduction

This paper considers the problem of separating reflection components (i.e., specular and diffuse reflections) in a single image. It is useful for several purposes. One is the use with photometric methods, such as shape-from-shading [1, 2] and photometric stereo [3]. As these methods often assume the surfaces of objects to be perfectly diffuse, it is necessary to eliminate specular component before applying them to real objects having specular reflectance. The separation of reflection components are also useful for the visual recognition of materials of objects; the highlights extracted from images are used as features for the recognition.

A large number of studies have been conducted to develop a method for accurately and robustly separating reflection components in a single image [4–13]. Most of them assume the dichromatic reflection model, which states that the light reflected on an object surface is given by a linear sum of a specular component and a diffuse component [4]. Specifically, the 3-vector \mathbf{i}_p containing the RGB values of a pixel p is given by

$$\mathbf{i}_p = \alpha_p \mathbf{i}_s + \beta_p \mathbf{i}_d, \quad (1)$$

where \mathbf{i}_s is the color of the only illumination existing in the scene and \mathbf{i}_d is the body color (i.e., the color caused by diffuse reflection) of the object surface.

If multiple pixels p share the same illumination color \mathbf{i}_s and the same body color \mathbf{i}_d , then Eq.(1) gives constraints on the variables on the right hand side. This is the principle on which color-based methods for separating components in a single image rely. More specifically, they commonly consider the following setting:

- The object surface consists of multiple regions with different body colors, each of which consists of a number of pixels with a single color \mathbf{i}_d .
- The illumination color \mathbf{i}_s is known. The body color \mathbf{i}_d of each region and also which region each pixel belongs to are unknown. The coefficients α_p and β_p are different for each pixel p , both of which are also unknown.

Early studies [7, 8, 14] attempt to solve the problem within this setting. More recent studies [9, 15, 10–13] attempt to utilize spatial information to improve separation accuracy. To do so, they incorporate spatial priors such as the smoothness of the body colors and/or the specular reflections on the object surface.

Our method separates reflection components based on sparse non-negative matrix factorization. It simultaneously performs the estimation of body colors and the separation of reflection components through optimization. It is notable that our method does not use an additional prior or assumption as those used in the recent studies. In this respect, our study runs counter to the recent trend of research, which is also the argument we make in this paper. That is, the above setting with the dichromatic model (1) alone might be more sufficient than expected for accurate separation of reflection components. In fact, as shown in the experimental results, our method is more accurate and robust than the state-of-the-art that uses additional priors. As an additional prior is not necessary, our method is free from tuning a number of hyper parameters.

2 Related work

The early approach to the problem is to determine body colors by analyzing the color space. A number of studies [7, 8, 14] are fallen in this category, and they solve the problem in two steps: (i) they determine the body colors first by analyzing the color space onto which all the image pixels are projected, and (ii) then determine the other unknowns using the results. The method of Klinker et al. [7] performs clustering of all the pixels in the RGB color space to determine body colors. Bajscy et al. [8] used the Hue-Saturation-Lightness color space instead of the RGB space. The method of Tan and Ikeuchi [14] projects pixel colors along the direction of the illumination color to a point of the lowest observed intensity to determine the body color. However, all these methods tend to be vulnerable to the clutters in the color space, such as image noises and color blending along the border of body colors.

To cope with this difficulty, more recent methods attempt to utilize spatial information in the image [9, 15, 10–13]. Instead of determining body colors first, most of these methods search for all the parameter values simultaneously through optimization. Some of them use a specular-free image (or its extension), an image free from specular components but with distorted diffuse components. It is created from the input image usually by a simple, pixel-wise operation.

Tan and Ikeuchi [9] first proposed a method of this category. They showed a method of creating the specular-free image by setting the maximum chromaticity of each pixel to an arbitrary value. Based on this, they presented a method that iteratively separates the reflection components by using a relation of two neighboring pixels. In their method,

body colors are estimated gradually in such a way that information propagates from outside highlight regions to inside them. This propagation often fails on the boundary of body colors. It also cannot correctly deal with body colors having the same hue but different saturation. To solve these problems, Tan et al. [10] proposed a method for recovering diffuse components by using the texture information around highlights, but it requires the positions of highlights to be known.

Yang et al. [12] extends the method of Tan and Ikeuchi by incorporating a more explicit prior that the body colors should be smooth on the object surface. Their method separates reflection components by applying a bilateral filter to the image of chromaticity. The bilateral filter, whose range filter is determined according to chromaticity information, smoothes out specular reflections while maintaining the edges in the chromaticity image.

There are more studies that follow a similar approach. Shen et al. [11] proposed another specular-free image that is obtained by subtracting the minimum of the RGB values from them for each pixel. They proposed a simple separation method based on it and also on an incorporated prior that the body color changes smoothly around highlights. Although it is simple and fast, their method is less accurate than the above methods, as it simplifies the problem too much, resulting in that Eq.(1) will no longer be satisfied. Kim et al.'s [13] have recently proposed an optimization-based approach that uses three different priors (i.e., the spatial smoothness of specular reflections and body colors, and the number of body colors being as small as possible). They also propose to apply the dark channel prior [16] to obtain another specular-free image, although it is exactly the same as the one of Shen et al. [11]. Their method alternately performs the following steps in an iterative manner: (i) cluster image pixels based on the latest estimate of their chromaticity and (ii) apply an edge-preserving filter to the result, followed by reassignment of labels. However, it remains unclear how accurate their method is, since their experiments compare mostly with the method of Tan and Ikeuchi [9] alone and not with that of Yang et al. [12], which is more close to their method in that the smoothness of body colors is assumed and an edge-preserving filter is used. Moreover, their method requires a number of hyperparameters (and it is unclear how to choose them) and also the assumed three priors are too much and could narrow the range of applicability.

3 Non-negative Matrix Factorization (NMF)

Our method is based on sparse non-negative matrix factorization (sparse NMF). Before describing our method, this section briefly summarizes sparse NMF and its numerical algorithm.

3.1 Basic NMF

NMF is a general-purpose method for multi-variate analysis. For data consisting of non-negative values such as images and speech signals, it factorizes the data into additive components. To be specific, a $M \times N$ matrix \mathbf{V} containing only non-negative elements

is factored into a product of a $M \times R$ matrix \mathbf{W} and a $R \times N$ matrix \mathbf{H} , both of which similarly contain only non-negative elements:

$$\mathbf{V} \simeq \mathbf{W}\mathbf{H}, \quad (2)$$

or the j -th column vector \mathbf{v}_j of \mathbf{V} is represented by a linear combination of the column vectors \mathbf{w}_k 's of \mathbf{W} weighted by the (k, j) element $H_{k,j}$ of \mathbf{H} as

$$\mathbf{v}_j \simeq \sum_{k=1}^R \mathbf{w}_k H_{k,j}. \quad (3)$$

The factorization is obtained by minimizing some cost $D(\mathbf{W}, \mathbf{H})$ measuring the difference between \mathbf{V} and its reproduction $\mathbf{W}\mathbf{H}$. For $D(\mathbf{W}, \mathbf{H})$, L_2 norm

$$D(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_2^2 \quad (4)$$

is widely used for general purposes, so is in our method. The generalized KL divergence[17] and Itakura-Saito divergence[18] are sometimes used depending on problems.

Since an efficient iterative algorithm was developed by Lee et al.[17], NMF has been applied to all sorts of problems, and various extensions have been made to the cost function depending on problems[18–24]

3.2 Sparse NMF

An important extension is the sparse NMF that incorporates the sparse regularization into the minimization[19]. It minimizes the following cost for the purpose of obtaining \mathbf{H} having as small a number of non-negative elements as possible.

$$F(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_2^2 + \lambda \sum_{i,j} H_{i,j}. \quad (5)$$

The second term on the right hand side follows the same relaxation as sparse coding[25] that L_0 norm is replaced by L_1 norm. The minimization of this cost results in that each data vector \mathbf{v}_j is represented by a linear combination of as small a number of bases (i.e., the column vector of \mathbf{W}) as possible, as in sparse coding. Its difference from sparse coding is that the resulting quantities are all non-negative.

A numerical algorithm for this sparse NMF, i.e., minimizing this cost under the constraints that \mathbf{W} and \mathbf{H} both have only non-negative entries is as follows. Starting from initial values \mathbf{W} and \mathbf{H} that are usually initialized in a random manner, it alternately iterates the following two updating rules until convergence:

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\bar{\mathbf{W}}^T \mathbf{V}}{\bar{\mathbf{W}}^T \bar{\mathbf{W}}\mathbf{H} + \lambda}, \quad (6a)$$

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{\mathbf{V}\mathbf{H}^T + \bar{\mathbf{W}} \odot \mathbf{A}\bar{\mathbf{W}}\mathbf{H}\mathbf{H}^T}{\bar{\mathbf{W}}\mathbf{H}\mathbf{H}^T + \bar{\mathbf{W}} \odot \mathbf{A}\bar{\mathbf{V}}\mathbf{H}^T}, \quad (6b)$$

where $\bar{\mathbf{W}}$ represents a matrix obtained by normalizing each column vector of \mathbf{W} ; \mathbf{A} is the $M \times M$ matrix whose entries are all 1; \odot indicates the Hadamard product (entry-wise product) and the division is similarly performed in the entry-wise manner. It is shown [19] that this iteration always reaches a local minimum in a finite counts of iterations.

4 Separation of specular components by sparse NMF

4.1 Problem formulation

In this section, we present our formulation of the problem of separating reflection components in a given image. Letting $R - 1$ be the number of body colors in the image (or equivalently, R be the body colors plus one illumination color), we denote the body colors by \mathbf{i}_k for $k = 1, \dots, R - 1$. We assume that there is only a single illumination color and denote it by a normalized vector \mathbf{i}_s (i.e., $\|\mathbf{i}_s\|_2 = 1$). Following the dichromatic model (1), the color \mathbf{i}_p of a pixel p can be represented as

$$\mathbf{i}_p = \alpha_p \mathbf{i}_s + \sum_{k=1}^{R-1} \beta_{k,p} \mathbf{i}_k, \quad (7)$$

where $\beta_{k,p}$ has a nonzero value only for a unique k in $[1, \dots, R - 1]$ and is zero for all other k 's, as the pixel color should be a linear sum of one body color and the illumination color. Using L_0 (counting) norm, this can be represented as

$$\sum_{k=1}^{R-1} \|\beta_{k,p}\|_0 = 1, \quad (8)$$

for any p . Note that the entries of \mathbf{i}_p , \mathbf{i}_s , and \mathbf{i}_k as well as α_p and $\beta_{k,p}$ have non-negative values.

For the moment, we assume that the number of body colors to be known, or equivalently, that R is known. Then, the separation problem is to estimate all the variables but \mathbf{i}_s on the right hand side of Eq.(7), under the constraint of Eq.(8) along with the non-negativeness of these variables, given \mathbf{i}_p for a number of pixels $p = 1, \dots$

This problem can be expressed compactly in a matrix form as the following constrained non-negative matrix factorization:

$$\mathbf{V} = \mathbf{W}\mathbf{H}, \quad (9a)$$

$$\text{s.t. } \sum_{i=2}^R \|H_{i,j}\|_0 = 1 \quad (j = 1, \dots, N), \quad (9b)$$

where \mathbf{V} a $3 \times N$ matrix (N is the number of image pixels) whose columns store the pixel colors \mathbf{i}_p 's for $p = 1, \dots, N$; \mathbf{W} is a $3 \times R$ matrix defined to be

$$\mathbf{W} = [\mathbf{i}_s, \mathbf{i}_1, \dots, \mathbf{i}_{R-1}]; \quad (10)$$

\mathbf{H} is a $R \times N$ matrix storing α_p and $\beta_{k,p}$'s in its columns as

$$\mathbf{H} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \\ \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{R-1,1} & \beta_{R-1,2} & \cdots & \beta_{R-1,N} \end{bmatrix}, \quad (11)$$

where the order of the columns (i.e., the pixels) is the same as \mathbf{V} ; $H_{i,j}$ is the (i, j) entry of \mathbf{H} . Note that the second equation is merely a rewritten version of Eq.(8). Note also that the constraints of non-negative values are naturally implemented in the inherent requirement of NMF. However, there is also a difference from usual NMF, which is that not all the entries of \mathbf{W} are unknown; when separating \mathbf{W} into the body and illumination colors and denoting it by

$$\mathbf{W} = [\mathbf{i}_s, \mathbf{W}_d], \quad (12)$$

we are to determine only the submatrix \mathbf{W}_d .

In the presence of noises, it is natural to minimize

$$F(\mathbf{W}_d, \mathbf{H}) \equiv \|\mathbf{V} - \mathbf{WH}\|_2^2 \quad (13)$$

under the constraint of (9b). The choice of L_2 norm is rationalized if we assume i.i.d. additive Gaussian noises with zero mean for each RGB value of each pixel. Then, the problem turns to finding a solution to this constrained NMF.

4.2 Relaxation of the problem

The presence of the constraint with L_0 norm makes it hard to directly solve the constrained minimization. Instead, we consider

$$F(\mathbf{W}_d, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_2^2 + \lambda \sum_{j=1}^N \sum_{i=2}^R \|H_{i,j}\|_0. \quad (14)$$

If there were no noise, the minimizer to this clearly gives the correct solution (i.e., the minimizer (\mathbf{W}, \mathbf{H}) yields the exact $\mathbf{V}(= \mathbf{WH})$ and \mathbf{H} satisfies the constraint (9b)), as long as we choose a small λ . (We exclude here the case where there are two or more possible factorizations for the given \mathbf{V} .) The same argument basically remains true in the presence of noises unless they are very large. However, one problem emerges in that case, which originates from the fact that the illumination color can be used 'for free.' For example, the illumination color could be wrongly used to explain noises. Even worse, it could also be used to explain pure diffuse colors, particularly for pixels having dark colors. If a pixel's color is dark (i.e., it is close to the origin in the RGB space), then which color to choose tends not to change the L_2 error much.

This necessitates imposing a certain penalty on choosing the illumination color. For this purpose, we employ the same regularization for the illumination color and rewrite the cost into

$$F(\mathbf{W}_d, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_2^2 + \lambda_s \sum_{j=1}^N \|H_{1,j}\|_0 + \lambda_d \sum_{j=1}^N \sum_{i=2}^R \|H_{i,j}\|_0. \quad (15)$$

The newly added term requires as small a number of pixels as possible to have specular components. Although the regularization parameters λ_s and λ_d can be arbitrarily chosen, they have the same nature that they should be determined based on the comparison against the strength of the noises. Thus, we set

$$\lambda = \lambda_s = \lambda_d. \quad (16)$$

Eq.(15) is still difficult to directly minimize, because of the existence of L_0 norm. Following the popular relaxation strategy of replacing L_0 norm with L_1 norm, we consider the minimization of

$$F(\mathbf{W}_d, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_2^2 + \lambda \sum_{j=1}^N \sum_{i=1}^R \|H_{i,j}\|_1. \quad (17)$$

Then, this is mostly the same as the cost for sparse NMF (Eq.(5)), although there is a difference that not all entries of \mathbf{W} are unknown in the above cost.

4.3 Modifying sparse NMF algorithm for separating reflection components

To deal with the difference of our problem from sparse NMF, we modify the algorithm so that only the unknown submatrix \mathbf{W}_d in \mathbf{W} will be updated. To be specific, we revise Eq.(6b) as

$$\mathbf{W}_d \leftarrow \mathbf{W}_d \odot \frac{\mathbf{V}'\mathbf{H}_d^\top + \bar{\mathbf{W}}_d \odot \mathbf{A}\bar{\mathbf{W}}_d\mathbf{H}_d\mathbf{H}_d^\top}{\bar{\mathbf{W}}_d\mathbf{H}_d\mathbf{H}_d^\top + \bar{\mathbf{W}}_d \odot \mathbf{A}\mathbf{V}'\mathbf{H}_d^\top}, \quad (18)$$

where

$$\mathbf{V}' = \mathbf{V} - \mathbf{i}_s \mathbf{h}_s, \quad (19a)$$

$$\mathbf{h}_s = [\alpha_1 \cdots \alpha_N], \quad (19b)$$

and

$$\mathbf{H}_d = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,N} \\ \vdots & \ddots & \vdots \\ \beta_{R-1,1} & \cdots & \beta_{R-1,N} \end{bmatrix}. \quad (20)$$

Here, \mathbf{A} is the $M \times M$ matrix of all 1's. This revised updating rule does not change \mathbf{i}_s , for which we set the known illumination color.

4.4 Determining R

We have assumed R (i.e., the number of body colors plus one) to be known so far. It is unknown in reality, and thus needs to be determined.

Our problem is originally to factorize a matrix containing the image into a product of non-negative matrices under the constraint (8); the constraint is such that a pixel color should be given by a linear combination of the illumination color and a single body color chosen from multiple candidates. This is relaxed to the minimization of

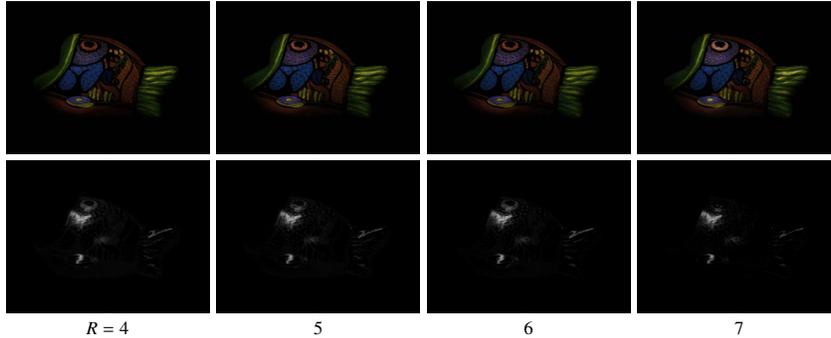


Fig. 1. The separation results obtained by our method for different R 's. Upper row: Diffuse components. Lower row: Specular components. The original image is shown in the top-left corner of Fig.6.

(17), which we expect to give a good solution. If it indeed gives a good solution, the solution will satisfy the constraint (8) at least approximately.

Thus, we incorporate a measure of how well a solution satisfies this constraint and search for R in an exhaustive manner using the measure. Specifically, computing a number of solutions for multiple different R 's, we choose the solution with the highest measure. For this measure, we define

$$\text{score}(\mathbf{H}) = \frac{1}{N} \sum_{j=1}^N \frac{\max_{i \in [2, R]} H_{i,j}}{\sum_{i=2}^R H_{i,j}}. \quad (21)$$

This score is in the range $[0,1]$ and returns 1 if the above constraint is fully satisfied. For a dark pixel without specular component, however, its diffuse component could be zero, making the score of the pixel also zero. Therefore, we choose the solution returning the highest score.

It seems difficult to determine the number of body colors for real images, because a certain amount of ambiguity has to be involved in the determination. Fortunately, our method tends not to be sensitive to the choice of R . As shown in Fig.1, it yields almost the same separation result for a certain range of R .

5 Experimental results

We conducted several experiments to examine the effectiveness of the proposed method and compare its performance with existing methods. For them, we choose Tan-Ikeuchi [9] and Yang et al. [12]; we used the authors' code^{1 2} in the experiments. We exclude Kim et al. [13], since its repeatability is limited due to lack of information. The method

¹ <http://www.staff.science.uu.nl/tan00109/code.html>

² <http://www.cs.cityu.edu.hk/qiyang/publications.html>

has a lot of tunable parameters, and the paper does not show how to choose them. (For example, the threshold of chromaticity distance between regions to be merged, after k-means clustering, in Sec.6.1 of [13]; it is unclear what edge-preserving filter is used for computing λ in Sec.6.2; the convergence threshold κ is not shown, and so on.) Note also that no authors' implementation is available.

5.1 Parameters and initial values

We set the weight of the sparse regularization $\lambda = 3$ throughout the experiments. For each image, we iterated the modified sparse NMF until it converges. The convergence is judged by

$$|F_t - F_{t-1}| < \epsilon |F_t|, \quad (22)$$

where F_t and F_{t-1} are the values of the cost (17) at iteration count t and $t - 1$, respectively; we set ϵ to be $\exp(-18)$.

The initial values of \mathbf{H} and \mathbf{W}_d are generated by uniform random numbers over $[1, 255]$, and each column vector of \mathbf{W}_d is normalized to have length 1. We set the illumination color \mathbf{i}_s to be $\frac{1}{\sqrt{3}}[1, 1, 1]^\top$ in all the experiments. As is described in Section 4.4, we choose the solution yielding the best score (21) among those obtained for different R in the range $[3, 12]$. The (modified) sparse NMF is not guaranteed to find the global optimum. In fact, we confirmed in our experiments that it was occasionally trapped in apparent local minima. Thus, we run it three times with different initial values for each value of R , and choose the solution yielding the best score among all the solutions thus obtained.

From \mathbf{W}_d and \mathbf{H} of the best solution, the specular and diffuse components are reconstructed by

$$\mathbf{I}_s = \mathbf{i}_s \mathbf{h}_s, \quad (23)$$

and

$$\mathbf{I}_d = \mathbf{W}_d \mathbf{H}_d, \quad (24)$$

respectively.

5.2 Synthetic images

We first applied the three methods to a synthetic image shown in Fig.2 and its results also are shown in there. The parenthesis below each image show the error (RMSE) of the separated specular components. From the separation results, our method is the best in terms of accuracy; Tan-Ikeuchi is clearly erroneous, and Yang et al. is better than Tan-Ikeuchi but has an erroneous horizontal line between the two regions on the left. These are also confirmed by the error values in the parentheses.

When a body color is close to the illumination color, the separation will be difficult. We compare the behaviors of the three methods in such situations. We created the images of a sphere with a single body color $[R, G, B] = [1, 1, x]$ by varying $x = 0.1, 0.3$, and 0.9 . As the illumination color is $[1, 1, 1]$, the separation is more difficult as x goes to 1. Figure 3 shows the results for these images. Tan-Ikeuchi could not separate the

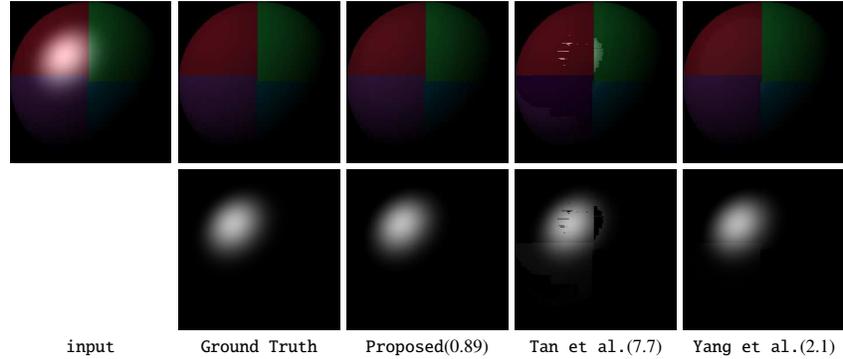


Fig. 2. Input image, ground truth, and separation results by the three methods. The numbers in the parentheses indicate the errors of the separated specular components.

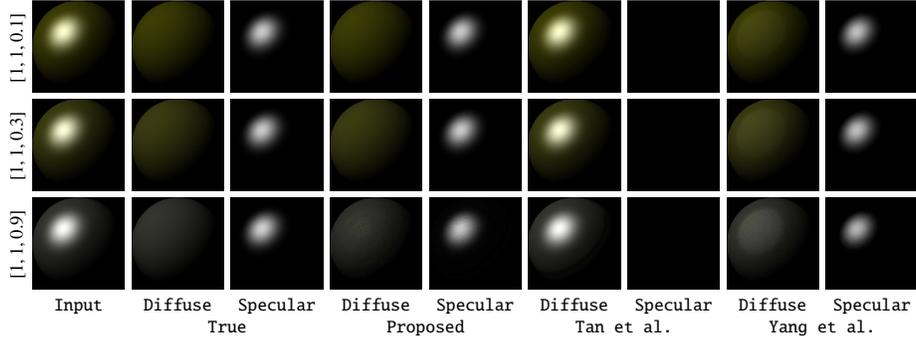


Fig. 3. The results when the body color is close to the illumination color. From top row to bottom, $[R, G, B] = [1, 1, 0.1]$, $[1, 1, 0.3]$, and $[1, 1, 0.9]$.

components at all, as they assume $R \neq G \neq B$ for every pixel. The results of Yang et al. tend to be worse with increasing x ; the separated specular components appear to be smaller than their true values. Our method yields clearly better separation results.

5.3 Real images

We then show results for real images. We first show the results for the dataset of [9] that is available from the author's site³ and is widely used in previous studies.

We first tested robustness to image noises. We added zero-mean Gaussian noise with $\sigma = 0.1, 1.0,$ and 5.0 to each pixel (ranging in $[0, 255]$) of each channel of an input image. Figure 4 shows the results for different levels of the noise. It is seen that the three methods are almost equally robust to noises, although Tan et al. is slightly less robust than the other two.

³ <http://www.staff.science.uu.nl/tan00109/code.html>

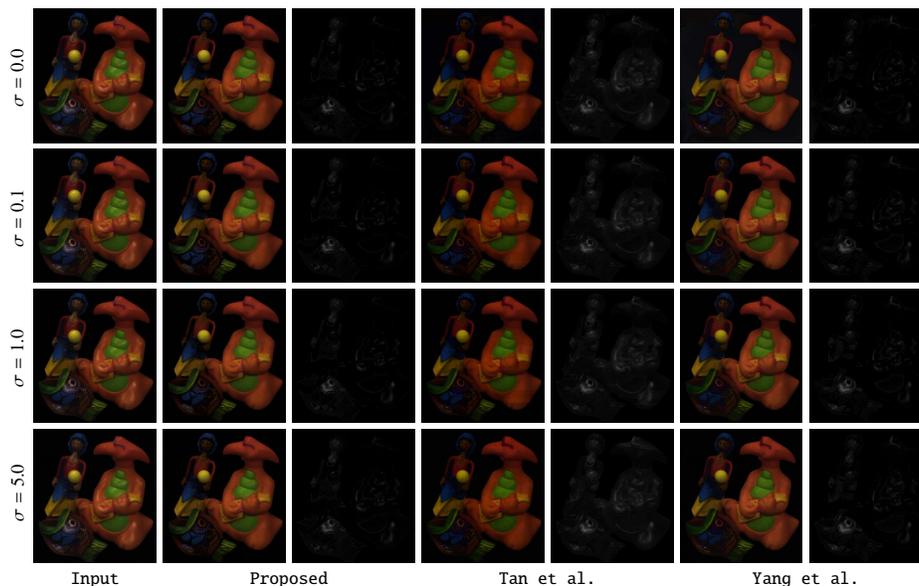


Fig. 4. Results for noisy input images. Zero-mean Gaussian noises with $\sigma = 0.1, 1.0,$ and 5.0 are added to each pixel of each channel of the original image of the top row.

Many existing methods assume a linear mapping between the image irradiance and the corresponding pixel value. Calibrating the response function of the camera will make the assumption true, but the calibration will include a certain amount of errors. We may simulate such errors by transforming each pixel value x into $x^{1/\gamma}$. To examine the performance of the methods in the presence of such calibration errors, we generated a number of images in this way and applied the three methods to the images. The results are shown in Fig.5.

We may consider a separation result for each γ value to be more accurate if it is closer to that for $\gamma = 1.0$ (the original image). Table 1 shows the errors in this accuracy measure. It is seen that our method is significantly better than Tan-Ikeuchi and is slightly better than Yang et al.

Table 1. The errors of separation results for each γ . Each number shows the difference of the separation result (i.e., specular component) from that for $\gamma = 1$ obtained by the individual method.

γ	Proposed	Tan-Ikeuchi[9]	Yang et al.[12]
1.0	0.0	0.0	0.0
1.1	0.77	5.3	1.8
1.3	1.9	9.5	4.0
1.5	3.3	14.7	5.7

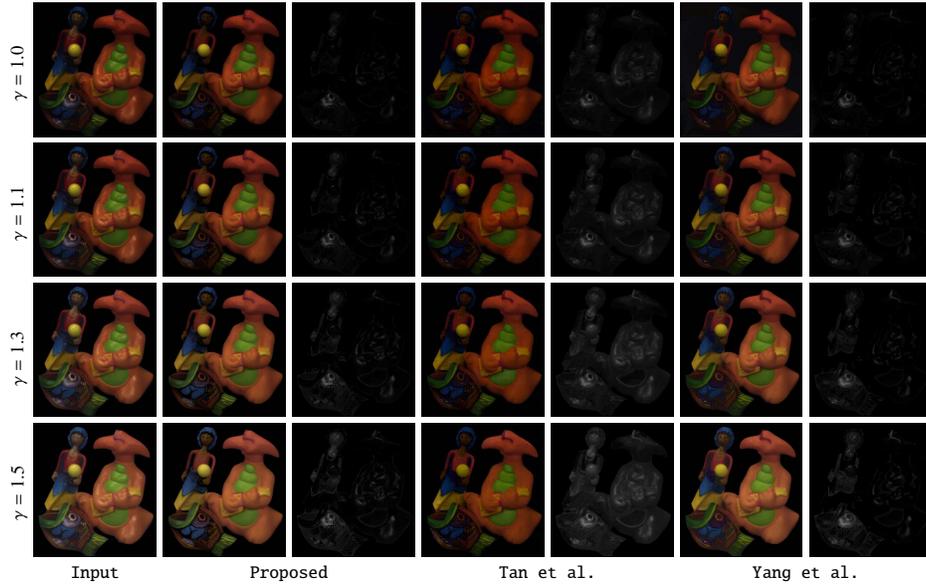


Fig. 5. Results for images whose brightness is distorted by transforming each pixel value x to $x^{1/\gamma}$. See Table 1 for quantitative comparisons.

Unlike other methods, our method does not assume a prior on spatial information such as the smoothness of body colors. To show the difference of our method from others, we intentionally destroyed the spatial relationship of pixels and applied the three methods to the resulting images. To be specific, we transform an image by dividing it into patches of $w \times w$ pixels in a grid manner and then randomly shuffle these patches to generate an image. Figure 6 shows the results. Note that the displayed images have been transformed back by the inverse of the above transformation (except for the input images). The three rows correspond to $w = \infty$ (no shuffle), 5×5 , and 1×1 (pixelwise), respectively. Tan-Ikeuchi is somewhat robust to this geometric transformation, as it uses only the relationship between the neighboring pixels, whereas Yang et al. is less robust; it completely fails separation at $w = 1$.

Finally, we show the results for other real images in Fig.7. Although there is no ground truth and thus it is difficult to evaluate separation accuracy, it is observed that Tan-Ikeuchi clearly yields excessive amount of specular components; the results of Yang et al. are slightly better, but its tendency is the same; our method provides much better results than the others. It should be noted that in the results from third to fifth rows, some of the pixels have zero diffuse components, which are caused by saturation of the brightness. It violates the assumed model, and it is no wonder that the correct results are not obtained. We nevertheless show the result to demonstrate how the methods can handle the real-world images having saturated pixels.

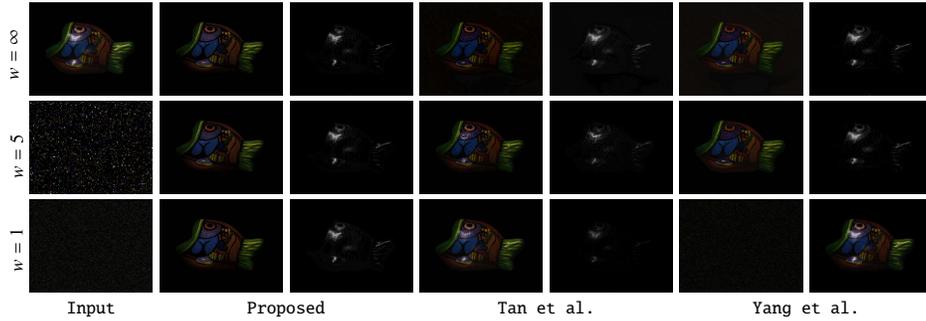


Fig. 6. The results for images undergoing patch-wise geometric shuffling. The image at the top-left corner is divided into $w \times w$ patches and randomly shuffled (the leftmost column). The separation results are transformed back by the inverse of the shuffling.

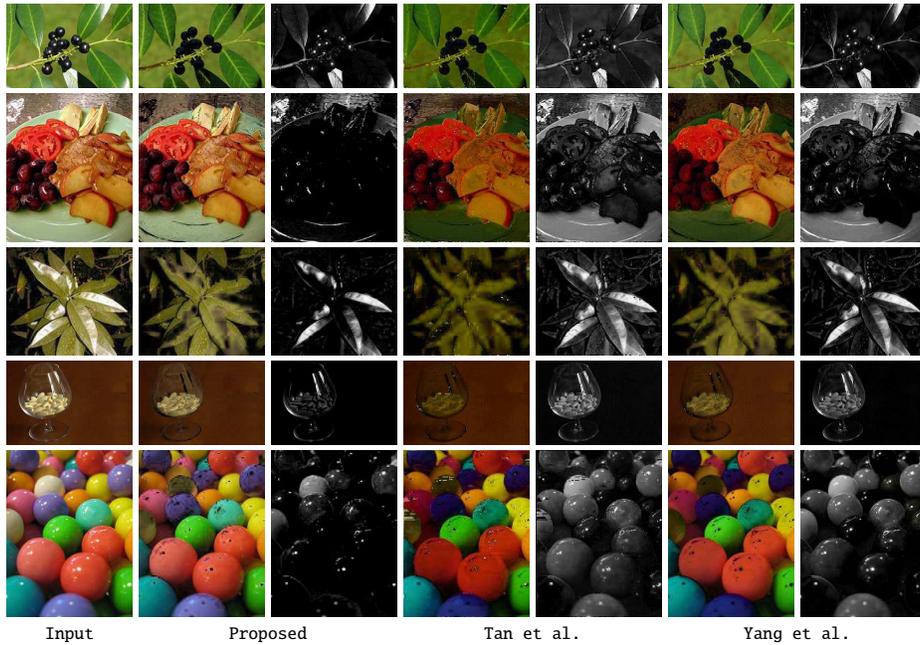


Fig. 7. The results for other real images.

6 Summary and discussions

This paper has shown a method for separating reflection components based on sparse NMF. The method is characterized by that it simultaneously performs the estimation of body colors and the separation of reflection components through the NMF-based optimization. The experimental results show that the method yields more accurate separation results than the state-of-the-art that incorporates spatial priors such as the smooth-

ness of body colors. This shows that the basic problem setting with the dichromatic model alone might be more sufficient than expected for accurate separation of reflection components. Our sparse NMF-based method can successfully find an accurate solution through optimization, owing to the good numerical property of NMF.

The main contribution of this study is not necessarily the proposed method itself but rather the finding that the problem can be solved more simply than thought. Earlier studies use spatial priors, whereas our method does not use spatial information at all, and nevertheless yields reasonably better results. We suspect that our NMF formulation can better utilize the constraints given by the dichromatic model along with the implicit surface color model (i.e., surfaces consist of multiple uni-colored regions). This will be further investigated in our future study.

Acknowledgement. This work was supported by JSPS KAKENHI Grant Number 25135701.

References

1. Horn, B.K.P.: Obtaining shape from shading information. In: *The Psychology of Computer Vision*. Winston, H. P. and Horn, B. (1975) 115–155
2. Prados, E., Faugeras, O. In: *Shape from shading*. Springer (2006) 1–17
3. Woodham, R.: Photometric method for determining surface orientation from multiple images. *Optical engineering* **19** (1980) 139–144
4. Shafer, S.: Using color to separate reflection components. *Color Research & Application* **10** (1985) 43–51
5. Swaminathan, R., Kang, S., Szeliski, R.: On the motion and appearance of specularities in image sequences. In: *ECCV*. (2002)
6. Feris, R., Raskar, R., Turk, M.: Specular reflection reduction with multi-flash imaging. In: *17th Brazilian Symposium on Computer Graphics and Image Processing, IEEE Comput. Soc* (2004)
7. Klinker, G., Shafer, S., Kanade, T.: The measurement of highlights in color images. In: *IJCV*. (1988)
8. Bajcsy, B., Lee, S., Leonardis, A.: Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. In: *IJCV*. (1996)
9. Tan, R.T., Ikeuchi, K.: Separating reflection components of textured surfaces using a single image. In: *PAMI*. (2005)
10. Tan, P., Lin, S., Quan, L.: Separation of Highlight Reflections on Textured Surfaces. In: *CVPR*. (2006)
11. Shen, H.L., Cai, Q.Y.: Simple and efficient method for specular removal in an image. *Applied optics* **48** (2009) 2711–9
12. Yang, Q., Wang, S., Ahuja, N.: Real-time specular highlight removal using bilateral filtering. In: *ECCV*. (2010)
13. Kim, H., Jin, H., Hadap, s., Kweon, I.: Specular Reflection Separation Using Dark Channel Prior. In: *CVPR*. (2013)
14. Tan, R.T., Katsushi, I.: Reflection Components Decomposition of Textured Surfaces using Linear Basis Functions. In: *CVPR*. (2005)
15. Mallick, S., Zickler, T.: Specularity removal in images and videos: A PDE approach. In: *ECCV*. (2006)

16. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: CVPR. (2009)
17. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401** (1999) 788–791
18. Févotte, C., Bertin, N., Durrieu, J.: Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation* **21** (2009) 793–830
19. Eggert, J., Korner, E.: Sparse coding and NMF. *Neural Networks* **2** (2004) 2529–2533
20. Hoyer, P.: Non-negative matrix factorization with sparseness constraints. *Machine Learning Research* (2004)
21. Virtanen, T.: Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *Audio, Speech, and Language Processing* **15** (2007) 1066–1074
22. Schmidt, M.: Speech separation using non-negative features and sparse non-negative matrix factorization. *Computer Speech and Language* (2008)
23. Choi, S.: Algorithms for orthogonal nonnegative matrix factorization. *International Joint Conference on Neural Networks* (2008) 1828–1832
24. Bertin, N., Badeau, R., Vincent, E.: Enforcing Harmonicity and Smoothness in Bayesian Non-Negative Matrix Factorization Applied to Polyphonic Music Transcription. *Audio, Speech, and Language Processing* **18** (2010) 538–549
25. Olshausen, B., Field, D.J.: Sparse coding of sensory inputs. *Current opinion in neurobiology* **14** (2004) 481–487