

# Specular-Free Residual Minimization for Photometric Stereo with Unknown Light Sources

Tsuyoshi Migita, Kazuhiro Sogawa, and Takeshi Shakunaga

Okayama University

**Abstract.** We address a photometric stereo problem that has unknown lighting conditions. To estimate the shape, reflection properties, and lighting conditions, we employ a nonlinear minimization that searches for parameters that can synthesize images that best fit the input images. A similar approach has been reported previously, but it suffers from slow convergence due to specular reflection parameters. In this paper, we introduce specular-free residual minimization that avoids the negative effects of specular reflection components by projecting the residual onto the complementary space of the light color. The minimization process simultaneously searches for the optimal light color and other parameters. We demonstrate the effectiveness of the proposed method using several real and synthetic image sets.

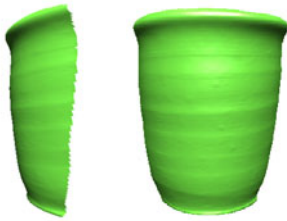
## 1 Introduction

Photometric stereo is a method for recovering the shape and albedo of an object from a set of images, when the object and the camera are fixed but the lighting conditions vary between images. A classical formulation assumes that the object is a Lambertian surface and that the lighting conditions are known. Several recently developed methods consider unknown lighting conditions and/or non-Lambertian surfaces (e.g., [4,14,3,10,2,11,12]).

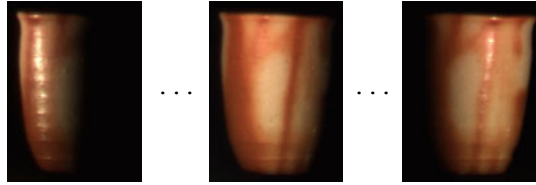
We have developed a method based on the formulation of Migita et al. [6], that directly minimizes a cost function to estimate the shape and the reflection properties of the object and the light position for each image. The cost function is the difference between input and synthesized images based on the Torrance-Sparrow model [13]. In this study, we modify the cost function to improve the performance because the nonlinearity of the Torrance-Sparrow model causes the cost function to be highly nonlinear and thus convergence tends to be slow.

This method involves decomposing the residual (an RGB vector of the difference for each pixel) into two subspaces: a 1d space parallel to the light color and a 2d space orthogonal to the light color. This decomposition is similar to that proposed by Zickler et al. [15]. By using the latter component only, the specular term is removed so that the estimation should be faster and more accurate. In addition, removing specular component reduces the number of parameters to be estimated.

However, there are some disadvantages with removing nonlinear components from the input images. For example, the method cannot be used for monochrome



**Fig. 1.** Reconstructed shape



**Fig. 2.** Images under various lighting conditions

images. Another problem is the generalized bas-relief (GBR) ambiguity [1,7,11]. Since our method removes the specular component, which may help resolve the ambiguity, we have to rely only on the effect of near light sources to resolve the ambiguity.

The new formulation does not use the Torrance-Sparrow model. Instead the dichromatic reflection model [9,5] is used. This model is commonly used to analyze a color histogram to estimate the light source color. However, in our method, estimation of the light source color is integrated with the minimization process so that it is not a separate process.

We verify the effectiveness of our method using several real and synthetic image sets.

## 2 Formulation

This section describes our formulation, which is referred to as the specular-free residual (SR) minimization hereafter. We describe the similarities and differences between our method and another method [6], which we call the full-color residual (FR) minimization.

The method is essentially a photometric stereo method, which recovers the object shape (e.g., Fig. 1) from images obtained under various lighting conditions, such as Fig. 2. Input images are of a static object and are obtained by a static camera. We also assume that the lighting is a single point light source in the near distance. We need to estimate the shape and the reflection properties of the object and the light position for each image. The reconstruction is performed by nonlinear minimization by comparing input with images synthesized using an image generation model (see below).

### 2.1 Full-Color Residual Minimization

Each pixel in the input image corresponds to a surface element of the object, and its intensity is described by a certain reflection model.

The FR minimization uses a simplified version of the Torrance-Sparrow model [13] to describe the input intensity  $e_{fp}$  (an RGB vector for the  $p$ 'th pixel in the  $f$ 'th image) as follows:

$$\mathbf{e}_{fp} = \eta_f \mathbf{d}_p \cos \beta_{fp} + \eta_f s_p \mathbf{S} \frac{1}{\cos \gamma_p} \exp(\rho \alpha_{fp}^2) + \mathbf{r}_{fp} \quad (1)$$

where

$$\begin{aligned} \cos \beta_{fp} &= \mathcal{N}[\mathbf{n}_p]^T \mathcal{N}[\mathbf{l}'_{fp}], & \mathbf{l}'_{fp} &= \mathbf{l}_f - \mathbf{x}_p, \\ \cos \gamma_p &= \mathcal{N}[\mathbf{n}_p]^T \mathcal{N}[\mathbf{v}'_p], & \mathbf{v}'_p &= (\mathbf{v} - \mathbf{x}_p)/\ell, \\ \cos \alpha_{fp} &= \mathcal{N}[\mathbf{n}_p]^T \mathcal{N}[\mathcal{N}[\mathbf{l}'_{fp}] + \mathcal{N}[\mathbf{v}'_p]], & \eta_f &= |\mathbf{l}_f - \mathbf{x}_p|^{-2}, \end{aligned} \quad (2)$$

and  $\mathbf{r}_{fp}$  is the residual,  $\mathbf{n}_p$  is the normal vector,  $\mathbf{l}_f$  is the light position,  $\mathbf{v}$  is the camera position,  $\ell$  is the focal length of the camera,  $\mathbf{d}_p$  is an RGB vector describing the diffuse reflectance,  $s_p$  is the specular reflectance,  $\mathbf{S}$  is an RGB vector describing the light color,  $\eta_f$  is a coefficient describing the attenuation of the light intensity due to the distance between the light and the object,  $\rho$  is a specular parameter, and  $\mathcal{N}[\cdot]$  is an operator that normalizes the norm of a vector to 1.

Note that for the first term in eq. (1),  $|\beta_{fp}| \geq \pi/2$  implies that the surface element is in an attached shadow region, which means the light is not positioned in front of the surface element. Consequently, this term is replaced with 0 in this case.

In the FR minimization, the shape, the reflection properties and the lighting conditions are reconstructed by minimizing the following cost function.

$$E(\mathbf{p}) = \frac{1}{2} \sum_{f=1}^F \sum_{p=1}^P |\mathbf{r}_{fp}|^2 \quad (3)$$

where  $\mathbf{r}_{fp}$  is the residual term in eq. (1). Note that, when the surface element for the  $p$ 'th pixel in the  $f$ 'th image is judged to be saturated or too dark (may be due to a cast shadow), we set  $\mathbf{r}_{fp} = (0, 0, 0)^T$ . The minimization is performed by Levenberg-Marquardt (LM) method [8], the details of which are given in section 3.

## 2.2 Specular-Free Residual Minimization

In the FR minimization, the specular term (i.e. the second term in eq. (1)) strongly reduces the convergence rate of the LM minimization process. Furthermore, it is computationally very expensive to calculate its derivative function, which is required for the minimization. Thus, the basic idea of the present study is to remove the specular term by using the following cost function instead of eq. (3):

$$E(\mathbf{p}) = \frac{1}{2} \sum_{f=1}^F \sum_{p=1}^P \left| \frac{\mathbf{S}_\times}{|\mathbf{S}|} \mathbf{r}_{fp} \right|^2 \quad (4)$$

where

$$\mathbf{S}_\times = \begin{bmatrix} 0 & -S_3 & S_2 \\ S_3 & 0 & -S_1 \\ -S_2 & S_1 & 0 \end{bmatrix} \quad (5)$$

with  $\mathbf{S} = (S_1, S_2, S_3)^T$ . Since  $\mathbf{S}_\times \mathbf{S} = 0$ , we can calculate the specular-free residual  $\mathbf{S}_\times \mathbf{r}_{fp}$  without calculating the second term in eq. (1). Thus, in our method, specular reflection is not limited to the Torrance-Sparrow model. Instead, the method employs a dichromatic reflection model; i.e., the specular color is same in every pixel, although its scale can differ between pixels.

However, this formulation has several drawbacks, because the cross product operation removes a part of the diffuse component in addition to the specular term. Consequently, some important information is lost. For example, it cannot process monochrome input images. Full-color input images are required to decompose the residual vectors into components parallel and orthogonal to the light color. Specifically the object must have two or more colors besides the light color to avoid a local minima that causes the estimator for  $\mathbf{S}$  to converge to the diffuse color rather than the specular color.

In addition, the estimation is greatly stabilized by normalizing the residual terms in eq. (4) as follows:

$$E(\mathbf{p}) = \frac{1}{2} \sum_{f=1}^F \sum_{p=1}^P \frac{|\mathbf{S}_\times \mathbf{r}_{fp}|^2}{|\mathbf{S}_\times \mathbf{y}_p|^2}, \quad \text{where } \mathbf{y}_p = \mathcal{N} \left[ \sum_f \mathbf{e}_{fp} \right]. \quad (6)$$

This normalization is interpreted as follows. Applying  $\mathbf{S}_\times$  from left, the input intensity is somewhat scaled and the scale factor is dependent on the color of the corresponding surface element. This scale factor should be compensated to accurately evaluate the residual  $\mathbf{r}_{fp}$ .

Another problem that needs to be considered is the GBR ambiguity. This ambiguity is described as follows. The Lambertian component is expressed in a bilinear form  $\mathbf{l}^T \mathbf{n}$ , which could be transformed into  $(\mathbf{A}^{-T} \mathbf{l})^T (\mathbf{A} \mathbf{n})$  by any nonsingular  $3 \times 3$  matrix  $\mathbf{A}$ . Using the integrability constraint,  $\mathbf{A}$  can be specified up to three degrees of freedom [1]. To determine the remaining three parameters, previous studies have used several nonlinearities [7,2,10]. The FR minimization has some such nonlinearities. However, since our SR minimization removes the specular component, we have to use another clue to resolve the ambiguity. The most important clue is the attenuation with distance between the light source and the object, as shown in [7]. Below, we present an experimental result that demonstrates that this nonlinearity can resolve the ambiguity.

### 2.3 Estimation Parameters

The parameters to be estimated form a large vector  $\mathbf{p}$ . The  $\mathbf{p}$  consists of four kinds of parameters: the object shape, the reflection properties, the global parameters, and the light positions as

$$\mathbf{p} = (\mathbf{p}_w^T, \mathbf{p}_s^T, \mathbf{p}_m^T, \mathbf{p}_l^T)^T. \quad (7)$$

**Object Shape.** The object shape is described by the depth  $\lambda_p$  for each pixel. A vector containing all the depths is  $\mathbf{p}_s = (\lambda_1, \dots, \lambda_P)^T$ , where  $P$  is the number of pixels to be estimated.

From the depths, the 3d coordinates  $\mathbf{x}_p$  for the  $p$ 'th surface element are calculated using the following formula:

$$\mathbf{x}_p = \lambda_p \left( \frac{u_p}{\ell}, \frac{v_p}{\ell}, 1 \right)^T + (u_p, v_p, 0)^T \quad (8)$$

where  $(u_p, v_p)^T$  is the 2d coordinates of the pixel with respect to the image center and  $\ell$  is the focal length of the camera. For affine camera model,  $\ell$  is set to infinity, while for projective camera model,  $\ell$  is set to a finite value. The world origin is located at the center of the image plane and the camera position  $\mathbf{v}$  is  $(0, 0, -\ell)^T$ .

The surface normal  $\mathbf{n}_p$  for the  $p$ 'th pixel is calculated from the 3d coordinates of its neighboring pixels,  $l, r, t, b$  (i.e., left, right, top and bottom, respectively), as follows:

$$\mathbf{n}_p = (\mathbf{x}_t - \mathbf{x}_b) \times (\mathbf{x}_r - \mathbf{x}_l). \quad (9)$$

The normal is not limited to being a unit vector. Instead, normalization is included in the cosine operations in eq. (2).

**Reflectance.** In our SR minimization, reflection parameters are the diffuse colors  $\mathbf{d}_p$  for each pixel. Unlike the FR minimization, specular reflectance  $s_p$  is not required. Therefore,  $\mathbf{p}_w = (\mathbf{d}_1^T, \dots, \mathbf{d}_P^T)^T$ .

**Global Parameters.** The only global parameter in the SR minimization is the light (or specular) color  $\mathbf{S}$ , whereas the FR minimization has an additional specular parameter,  $\rho$ .

**Light Positions.** The light position is estimated for each image. Thus,  $\mathbf{p}_l = (\mathbf{l}_1^T, \dots, \mathbf{l}_F^T)^T$ , where  $F$  is the number of input images.

### 3 Minimization

Letting the number of pixels to be  $P$ , there are more than  $4P$  elements in  $\mathbf{p}$ . Thus, the search space can typically have the dimension of about 100,000. An algorithm is required that can deal with a minimization problem on this scale.

#### 3.1 LM Method

We use LM method for minimizing the cost function, as in the FR minimization, which is given by:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - (J_k^T J_k + \mu_k I)^{-1} J_k^T \mathbf{r}_k \quad (10)$$

where  $\mathbf{r}_k$  is a vector containing all the (specular-free) residual vectors, and  $J$  is the Jacobian matrix. The subscript  $k$  indicates the value is dependent on the parameter  $\mathbf{p}_k$ . The initial value  $\mathbf{p}_0$  is discussed in section 3.3.





Fig. 3. Initial shape

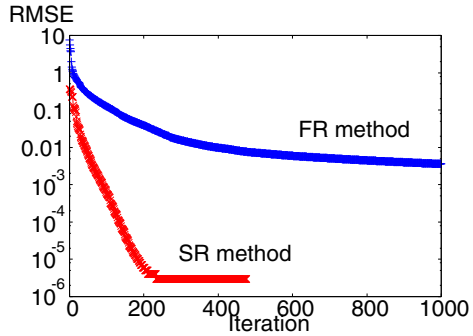


Fig. 4. Residual w.r.t. iteration

### 3.3 Initialization

The initial parameter of the LM process is as follows.

**Object Shape.** We initialize the shape as a simple paraboloid described by the following formula by an appropriate set of parameters  $(a, b)$ :

$$\lambda_p = au_p^2 + bv_p^2 \quad (15)$$

Fig. 3 shows an example. Using an appropriate curvature is effective for avoiding local minima. Otherwise a convex shape may be recovered as a concave object.

**Lighting.** There are two types of lighting parameters: color, which is initialized to  $(1, 1, 1)^T$  (i.e., a white light source), and the light position for each image, which is initialized to  $(0, 0, -d)^T$ , which is independent of  $f$ , where  $d$  is an appropriate value.

Although this is a very rough initialization, the nonlinear optimization algorithm can search for a reasonable shape and light positions. A more elaborate initialization such as a method based on SVD [14] may improve the convergence.

**Reflectance.** The Lambertian parameter  $\mathbf{d}_p$  is an RGB vector that is a scalar multiple of the reflectance. It is initialized by taking an average of the input images as in the following formula.

$$\mathbf{d}_p = \frac{1}{F} \sum_{f=1}^F \mathbf{e}_{fp} \quad (16)$$

## 4 Experiments

The SR minimization was tested on several real and synthetic image sets.

In several of the experiments, affine and projective camera models were tested and the results obtained were similar. The results given below are based on the

affine model (i.e., infinite  $\ell$ ). Note that the computational cost for infinite  $\ell$  could be significantly smaller than for a finite  $\ell$ , although this is not the focus of this paper.

#### 4.1 Convergence Speed

The SR minimization removes the specular components to speed up convergence. Fig. 4 shows the convergence speeds for FR and SR minimizations for an synthetic image set. The SR minimization converges much faster than the FR minimization. Although comparison of absolute RMSE values are not so meaningful because the definitions of the residuals are different, the RMSE cannot be significantly lower than  $10^{-6}$ , since the input images are given in 32-bit floating point format. The SR minimization reached this limit after about 200 iterations, but we doubt that the FR minimization will reach this limit.

#### 4.2 Resistance for GBR Ambiguity

We verified that the nearby lighting can resolve GBR ambiguity in our formulation. To demonstrate this, we used two sets of synthetic images. One is rendered using near light sources and the other using distant light sources. Several minimization trials were conducted using various initial values that were created by applying different GBR transformations to the same converged parameters. Fig. 5 shows the results. The bottom row corresponds to the cases for near light sources. The same distinctive shape was obtained from several initial values. In contrast, the reconstructed shapes were not fully corrected for distant light sources (the top row).

#### 4.3 Real Image Sets

Here, we present several experiments on real image sets. We stopped the search after 200 iterations. The results fully converged for several image sets, but not for other image sets.

In these experiments, we used multiple light sources that had different intensities, despite the fact that eq. (1) assumes that every light source has the same intensity. This might generate some error in the distance between the light source and the target object.

**Apple.** A fresh apple was imaged in a dark room. The light positions are shown in Fig. 6. We obtained 24 images, several examples of which are shown in Fig. 7. The size of each image is  $125 \times 133$  pixels, and 11,098 points are estimated. The FR and SR minimizations produced the results shown in Fig. 8 and Fig. 9, respectively. In Fig. 8, the estimation produced a concave object, which is apparently a failure. This is due to an intense specular reflection near the image center. In Fig. 9, on the other hand, the estimated shape appears smooth and apple-like. This result is an example of the SR minimization outperforming the FR minimization.



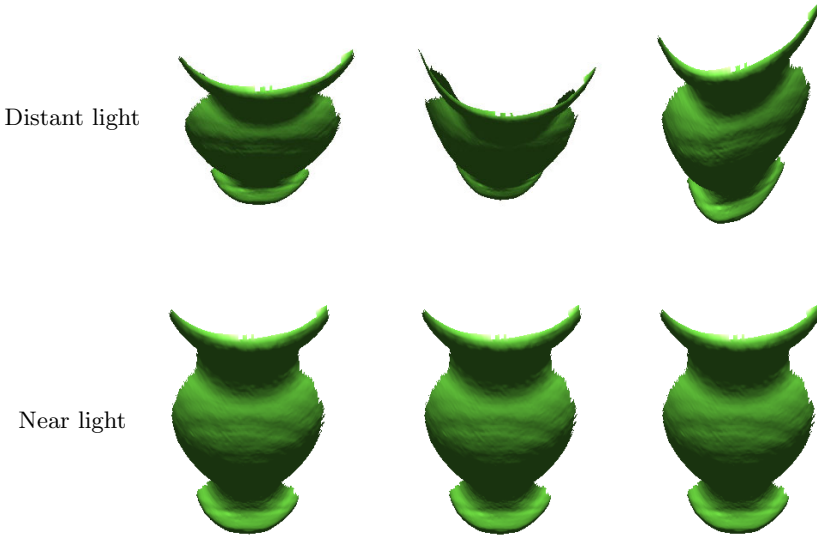


Fig. 5. Shapes reconstructed from several initial values

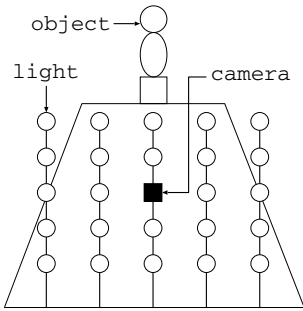


Fig. 6. Configuration

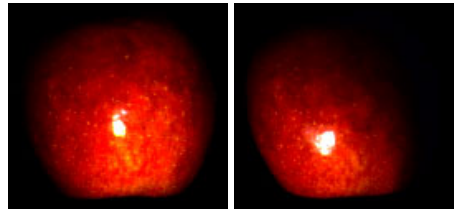


Fig. 7. Example images of an apple

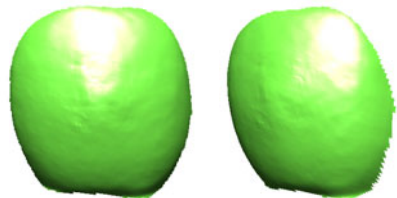
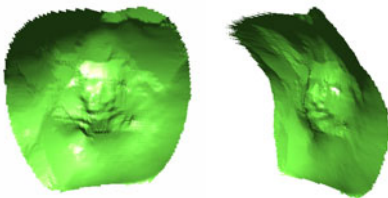
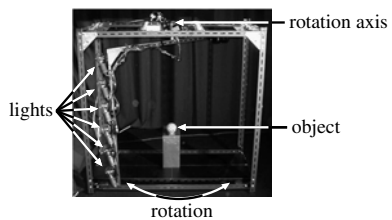


Fig. 8. Apple results by FR minimization Fig. 9. Apple results by SR minimization



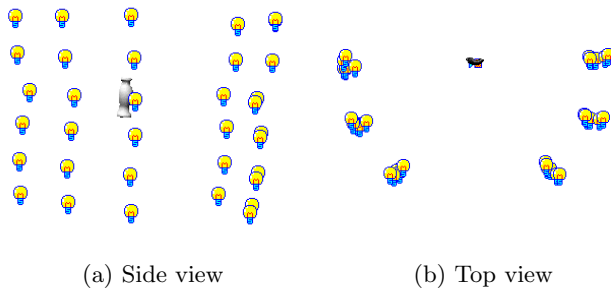
**Fig. 10.** Lighting system



**Fig. 11.** Example images of a wooden figure



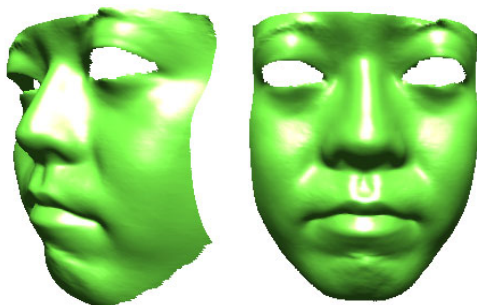
**Fig. 12.** Results for a wooden figure



**Fig. 13.** Light positions and a wooden figure



**Fig. 14.** An image of a human face



**Fig. 15.** Results for a human face

**Wooden Figure.** A wooden figure was imaged using the lighting system shown in Fig. 10, which has six light bulbs on a rotating arm. We obtained 36 images (Fig. 11) that were  $128 \times 296$  in size and 25,480 pixels were used for the estimation. The resulting shape is shown in Fig. 12, and the light positions are shown in Fig. 13 along with the reconstructed object. We used this sequence because the light positions are known to form a cylinder. Fig. 13 shows slightly distorted cylinder. One possible cause for this distortion is that the light bulbs have different intensities. On the other hand, the angles between the light columns are estimated well, as the top view image shows. The radius and the height of the estimated cylinder divided by the object height are approximately 3.2–4.2 and 4.6–5.5, respectively. There are considered to be reasonable results compared to their actual values of 3.7 and 5.0, respectively.

**Human Face.** We show an example of applications to human faces. The input images, an example of which is shown in Fig. 14, were obtained in the environment shown in Fig. 6. There are 24 images with a size  $185 \times 220$ , and 35,629 pixels of which were used for the estimation. The eye regions were manually removed because they were too glossy. The result is shown in Fig. 15. It is a satisfactory result, since it does not exhibit any severe degradation due to specular reflection and/or shadows cast around the nose.

## 5 Conclusions

This study considered a photometric stereo problem with unknown lighting conditions. It has been reported [3,6] that the shape and the reflection properties of the object and the lighting conditions can be recovered by minimizing the differences between input and estimated images based on a reflection model. In this paper, the cost function of the minimization is not constructed from the full-color residual, but from a specular-free projection of the residual in a space orthogonal to the light color. Using several real and synthetic image sets, we demonstrated that the specular-free residual (SR) minimization exhibits better performance than the full-color residual (FR) minimization for several cases.

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant-in-Aid for Young Scientists (B) (No. 22700181).

## References

1. Belhumeur, P.N., Kriegman, D.J., Yuille, A.L.: The bas-relief ambiguity. *Intl. J. of Computer Vision* 35(1), 33–44 (1999)
2. Chandraker, M., Kahl, F., Kriegman, D.: Reflections on the generalized bas-relief ambiguity. In: *Proc. CVPR 2005*, vol. I, pp. 788–795 (2005)
3. Georghiadis, A.S.: Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In: *Proc. ICCV 2003*, vol. 2, pp. 816–823 (2003)

4. Ikeuchi, K.: Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-3*(6), 661–669 (1981)
5. Klinker, G., Shafer, S., Kanade, T.: The measurement of high-lights in color images. *Intl. J. Computer Vision* 2(1), 7–32 (1988)
6. Migita, T., Ogino, S., Shakunaga, T.: Direct Bundle Estimation for Recovery of Shape, Reflectance Property and Light Position. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part III*. LNCS, vol. 5304, pp. 412–425. Springer, Heidelberg (2008)
7. Okabe, T., Sato, Y.: Does a nearby point light source resolve the ambiguity of shape recovery in uncalibrated photometric stereo? In: *Proc. MIRU 2007*. pp. 881–886 (2007) (in Japanese)
8. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes*, 3rd edn. Cambridge University Press (2007)
9. Shafer, S.A.: Using color to separate reflection components. *COLOR Research and Application* 10(4), 210–218 (1985)
10. Tan, P., Mallick, S.P., Quan, L., Kriegman, D., Zickler, T.: Isotropy, reciprocity and the generalized bas-relief ambiguity. In: *Proc. CVPR 2007*, pp. 1–8 (2007)
11. Tan, P., Zickler, T.: A projective framework for radiometric image analysis. In: *Proc. CVPR 2009*, pp. 2977–2984 (2009)
12. Tan, P., Lin, S., Quan, L.: Resolution-Enhanced Photometric Stereo. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part III*. LNCS, vol. 3953, pp. 58–71. Springer, Heidelberg (2006)
13. Torrance, K., Sparrow, E.: Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.* 57, 1105–1114 (1967)
14. Yuille, A., Snow, D., Epstein, R., Belhumeur, P.: Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *Intl. J. of Computer Vision* 35(3), 203–222 (1999)
15. Zickler, T., Mallick, S.P., Kriegman, D.J., Belhumeur, P.N.: Color subspaces as photometric invariants. *Intl. J. of Computer Vision* 79(1), 13–30 (2008)