Direct Bundle Estimation for Recovery of Shape, Reflectance Property and Light Position

Tsuyoshi Migita, Shinsuke Ogino, and Takeshi Shakunaga

Department of Computer Science, Okayama University
{migita,ogino,shaku}@chino.cs.okayama-u.ac.jp

Abstract. Given a set of images captured with a fixed camera while a point light source moves around an object, we can estimate the shape, reflectance property and texture of the object, as well as the positions of the light source. Our formulation is a large-scale nonlinear optimization that allows us to adjust the parameters so that the images synthesized from all of the parameters optimally fit the input images. This type of optimization, which is a variation of the bundle adjustment for structure and motion reconstruction, is often employed to refine a carefully constructed initial estimation. However, the initialization task often requires a great deal of labor, several special devices, or both. In the present paper, we describe (i) an easy method of initialization that does not require any special devices or a precise calibration and (ii) an efficient algorithm for the optimization. The efficiency of the optimization method enables us to use a simple initialization. For a set of synthesized images, the proposed method decreases the residual to zero. In addition, we show that various real objects, including toy models and human faces, can be successfully recovered.

1 Introduction

In the present paper, we present a method for estimating the three-dimensional shape and bidirectional reflectance distribution function (BRDF) of an object from a set of images, based on the appearance changes that occur with respect to the changing position of a point light source. The proposed method should fulfill the following two criteria: (i) it should not require any special devices, except for a camera, a darkened room, a light source, and a computer, and (ii) it should not be too theoretically complicated. Although the method proposed herein and that proposed in [1,2] are similar with respect to the formulation and the input data set, we would like to solve the problem using a much simpler framework. The theoretical simplicity requirement enables the method to be easily extended to more complicated models, even though this task will not be examined in the present paper. On the other hand, the minimal requirement of the proposed method is important to be usable by non-professionals who wish to conveniently create three-dimensional models.

Several methods for recovering the shape and BRDF of an object were proposed in the literature [1,2,3,4,5,6,7,8,9,10,11,12,13,14]. Some of these methods,

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part III, LNCS 5304, pp. 412–425, 2008.

as well as some earlier methods, proposed the recovery of an object shape by assuming that the lighting conditions are known in a computer-controlled lighting system, or by using a mirror spherical probe. Without prior knowledge of the lighting conditions, a typical method first estimates the shape of an object by using the silhouette intersection method, or by using a range finder and then estimating the reflectance properties. In the proposed method, the three-dimensional positions of the light source, as well as the object shape and reflectance properties of the object are estimated. Note that, in the previous proposed methods, a distant light source (which only has two degrees-of-freedom) is typically assumed. However, multiple viewpoints and a complicated lighting environment are beyond the scope of the present paper.

The proposed formulation and solution method is a variation of the bundle adjustment [15] used for structure and motion reconstruction. In other words, it is a large-scale nonlinear optimization that adjusts the parameters so that the images synthesized from all parameters optimally fit the input images. In terms of the given cost function, no other method attains a more accurate result. However, one of the difficulties is that the optimization requires a *reasonable* initial estimation. It is possible to use a sophisticated method, such as that described in Refs. [4,6,8], to initialize the optimization process. However, the required accuracy for the initialization is much lower. In fact, we use a flat plane for our initial shape parameters.

Once the initialization is finished, an efficient algorithm is required to perform the optimization. Since we assume the typical number of parameters to be approximately 10^5, a naive optimization method, such as the steepest descent method, is insufficient, and the Levenberg-Marquardt method [16], which exploits the second-order derivative, or a Hessian matrix, is required. For solving a large-scale linear equation system with a sparse coefficient matrix (Hessian matrix) for each iteration, the preconditioned conjugate gradient method is more suitable in that it allows us to solve the problem within a limited memory requirement and at a reasonable computational cost. The algorithm can attain an almost zero residual for an input set of synthesized images, which is not possible using naive methods. However, since extra care should be taken to avoid local minima, we gradually increase the number of parameters to be estimated and worked to detect abnormally estimated parameters that must be corrected.

The methods proposed in [1,2] employ a cost function similar to the proposed function and a very different approach for minimization in order to achieve a feasible computational cost. These methods require that the parameters be updated one by one, using several different algorithms (such as the steepest descent method, Newton's method, DCT, and SVD) based on several different aspects of the reflectance model. However, with our proposed method, all of the parameters are updated simultaneously.

Using our proposed method, we demonstrate that various real objects, including a wooden figure, model toys and a human face, can be successfully recovered.

2 Shape Recovery Method

2.1 Image Formation/Reconstruction Model

A set of input images can be described as a collection of the following *measurement vectors*:

$$\boldsymbol{m}_{fp} := (r_{fp}\ g_{fp}\ b_{fp})^T \tag{1}$$

which is a three-vector containing red, green, and blue components of the image intensity at the p-th pixel in the f-th image.

Since we assume that the object and the camera are fixed and that a point light source will be moving, $\boldsymbol{m}_{fp}$ is a function of the position of the light $\boldsymbol{l}_f$. It is also a function of the object shape and its reflectance property. We approximate the surface reflectance by use of the Simplified Torrance-Sparrow Model described in Ref. [3] to describe the measurement as follows:

$$\boldsymbol{m}_{fp} = \eta_f \left(\begin{bmatrix} w_{1p} \\ w_{2p} \\ w_{3p} \end{bmatrix} \cos\beta_{fp} + w_{4p} \begin{bmatrix} s_R \\ s_G \\ s_B \end{bmatrix} \frac{\exp\left(\rho\alpha_{fp}^2\right)}{\cos\gamma_p} \right) \tag{2}$$

where

$$\cos\beta_{fp} = \boldsymbol{n}_p^T \mathcal{N}\left[\boldsymbol{l}_f - \boldsymbol{x}_p\right], \tag{3}$$

$$\cos\gamma_p = \boldsymbol{n}_p^T \mathcal{N}\left[\boldsymbol{v} - \boldsymbol{x}_p\right], \tag{4}$$

$$\cos\alpha_{fp} = \boldsymbol{n}_p^T \mathcal{N}\left[\mathcal{N}\left[\boldsymbol{l}_f - \boldsymbol{x}_p\right] + \mathcal{N}\left[\boldsymbol{v} - \boldsymbol{x}_p\right]\right], \tag{5}$$

η_f is the emittance of the light source for the f-th image, $(w_{1p}, w_{2p}, w_{3p})^T$ and w_{4p} are the intrinsic color and the reflectance of the specular reflection at the p-th pixel. We refer to w_{mp} as the *weight*. Then, $\boldsymbol{l}_f$ is the position of the light source for the f-th image, $(s_R, s_G, s_B)^T$ is the color of the light source, $\boldsymbol{v}$ is the camera position, $\mathcal{N}$ is the normalization operator such that $\mathcal{N}[\boldsymbol{x}] := \boldsymbol{x}/|\boldsymbol{x}|$, and $\boldsymbol{x}_p$ is the three-dimensional position of the object at the p-th point. Note that the object shape is represented by its depth d_p from the camera for each pixel, not by triangular meshes. In addition, $\boldsymbol{n}_p$ is a unit normal, which is calculated from the three-dimensional positions of neighboring pixels. Finally, ρ is the surface roughness, which is shared by each pixel. However, this constraint does not mean that the object consists of just one material, because the specular reflection of w_{4p} can change from pixel to pixel.

Although Eqs. (3)–(5) assume that the light source is near the object, Eq. (2) does not take into account the attenuation with respect to the distance between the object and the light source. However, this effect can be approximated by considering that the light source emittance η_f decreases as the distance grows. Strictly speaking, the attenuation varies from pixel to pixel, but when the object is sufficiently small, compared to the distance, this effect is small and thus the approximation is sufficient for our experimental setup.

Using the image formation model, we can formulate the simultaneous recovery as a nonlinear optimization problem:

$$\arg\min_{\boldsymbol{u}} E(\boldsymbol{u}) \ , \quad \text{where} \quad E(\boldsymbol{u}) = \sum_{fp} \boldsymbol{r}_{fp}^T \boldsymbol{r}_{fp} \tag{6}$$

where $\boldsymbol{r}_{fp}$ is the difference between the measured intensity $\boldsymbol{m}_{fp}$ and the synthesized intensity (i.e., the right-hand side of Eq. (2)), and $\boldsymbol{u}$ is a vector containing all the parameters to be estimated, namely, depths d_p's and weights w_{mp}'s for all p and m, as well as emittances η_f's and positions $\boldsymbol{l}_f$'s for all f, in addition to specular parameters ρ and $\boldsymbol{s}$. Let N denote the dimension of $\boldsymbol{u}$; we typically assume $N \approx 10^5$.

The foreground and background are differentiated by thresholding the input images. In other words, constantly dark pixels throughout the images are considered to be in the background, which are not used for the estimation, and each foreground pixel has its own unique identifier p. Even if the p-th pixel is in the foreground, we exclude the error term $\boldsymbol{r}_{fp}$ from the cost function if the pixel is considered to be saturated or the pixel is in a shadow in the f-th image, i.e., all components of the pixel must be more than 0 and less than 255 when the intensity is 8 bits.

For each p-th pixel, the surface normal $\boldsymbol{n}_p$ is calculated by

$$\boldsymbol{n}_p = \mathcal{N}[(\boldsymbol{x}_{Rp} - \boldsymbol{x}_{Lp}) \times (\boldsymbol{x}_{Tp} - \boldsymbol{x}_{Bp})] \tag{7}$$

where Rp, Lp, Tp, and Bp indicate the indices of the pixel to the left, right, top, or bottom of the p-th pixel, respectively. However, when these pixels are outside the boundary of the object foreground, Rp, Lp, Tp, or Bp indicate p.

2.2 Optimization Method

To achieve an efficient search for the optimal parameter, we first describe the basic idea of the Levenberg-Marquardt (L-M) method, and then describe its efficient implementation by use of the preconditioned conjugate gradient (PCG) method. The selection of the initial parameter vector is discussed in the following section.

Letting $\boldsymbol{u}_k$ be the search vector at the k-th iteration, the L-M process is as follows:

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k - (H_k + \mu_k I)^{-1}(\nabla E) \tag{8}$$

where ∇E is the gradient of the cost function E, and H_k is the Hessian matrix of E, that is, $H_k := (\partial^2 E / \partial u_i \partial u_j)$. These are evaluated at $\boldsymbol{u}_k$, and μ_k is a constant for stabilization. For each iteration of the L-M process, we have to solve the large-scale linear equation system $(H_k + \mu_k I)\boldsymbol{q} = \nabla E$. Since the coefficient matrix is sparse, the PCG method is suitable. In this method, solving $\boldsymbol{q}$ so as

to satisfy $A\boldsymbol{q} = \boldsymbol{b}$ is equivalently transformed into the minimization of $f(\boldsymbol{q}) := (1/2)\boldsymbol{q}^T A\boldsymbol{q} - \boldsymbol{b}^T\boldsymbol{q}$, yielding the process [1]:

$$\boldsymbol{q}_k = \begin{cases} \text{initial guess} & (k=0) \\ \boldsymbol{q}_{k-1} - \alpha_{k-1}\boldsymbol{d}_{k-1} & (k>0), \end{cases} \quad \text{where} \quad \alpha_k = \arg\min_{\alpha} f\left(\boldsymbol{q}_k - \alpha\boldsymbol{d}_k\right) \tag{9}$$

$$\boldsymbol{d}_k = \begin{cases} \boldsymbol{C}^{-1}\boldsymbol{g}_0 & (k=0) \\ \boldsymbol{C}^{-1}\boldsymbol{g}_k + \beta_k\boldsymbol{d}_{k-1} & (k>0), \end{cases} \quad \text{where} \quad \beta_k = \frac{\boldsymbol{g}_k^T\boldsymbol{C}^{-1}\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^T\boldsymbol{C}^{-1}\boldsymbol{g}_{k-1}} \tag{10}$$

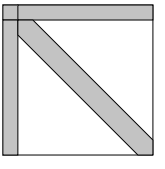
and

$$\boldsymbol{g}_k = \nabla f \qquad (\text{evaluated at } \boldsymbol{q}_k). \tag{11}$$

Here, $\boldsymbol{C}$ is called a preconditioning matrix, which is an approximation of the coefficient matrix A, such that $\boldsymbol{C}^{-1}\boldsymbol{g}_k$ is easily obtainable.

The structure of the Hessian matrix, or the coefficient matrix, becomes as follows:

$$\boldsymbol{C}_k = \tag{12}$$



when the parameters are ordered in such a manner that the former elements of $\boldsymbol{u}$ are the parameters independent of position p, and the latter elements are the shape and reflection parameters for each p. It is important that the bottom-right part has a band structure and that there are numerous zero elements inside the band. The reason for this will be described later. The topmost and the leftmost parts of the matrix are dense, but their heights and widths are small. Thus, although the size of the matrix is $N \times N$, the required memory size is $O(N)$, as is the computational complexity for approximating $\boldsymbol{C}_k^{-1}\nabla E$, which is obtained by a fixed number of iterations of the PCG process.

Note that, a naive method requires $O(N^2)$ memory and $O(N^3)$ computation and is hardly applicable for large N.

Implementation. The matrix $\boldsymbol{C}_k$ is calculated as follows, based on the approximation of the Hessian matrix used in the Gauss-Newton algorithm:

$$\boldsymbol{C}_k = \sum_{fp} \boldsymbol{J}_{fp}^T\boldsymbol{J}_{fp} + \mu_k\boldsymbol{I} \tag{13}$$

where

$$\boldsymbol{J}_{fp} = \begin{pmatrix} \dfrac{\partial \boldsymbol{r}_{fp}}{\partial u_1} & \cdots & \dfrac{\partial \boldsymbol{r}_{fp}}{\partial u_N} \end{pmatrix}. \tag{14}$$

[1] The symbols $\alpha_k, \beta_k, \boldsymbol{d}_k$ in the following algorithm are not the same as those in the reflection model.

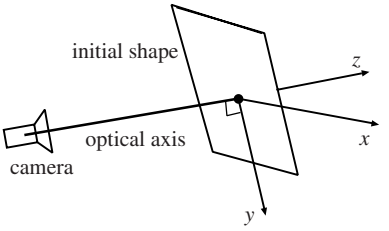


Fig. 1. Initial shape

Most elements in this Jacobian matrix are zero, because the residual vector $\boldsymbol{r}_{fp}$ is affected only by the three-dimensional positions and the weights of the p-th pixel and its direct neighbors, the lighting parameters of the f-th image, and several global parameters. Using this definition, it is easy to show that the Hessian matrix has the structure shown in Eq. (12).

To avoid the search for μ_k required for each iteration, as proposed in the original L-M algorithm [16], we again use the PCG algorithm. Note that, if α_k in Eq. (9) is 1, and β_k in Eq. (10) is 0, then the PCG method is exactly the same as the L-M method. Instead, we fix μ_k and search for α_k and calculate β_k for each iteration. In other words, we use a two-layered PCG algorithm, where the upper layer minimizes E in Eq. (6), and the lower layer calculates $\boldsymbol{C}_k^{-1}\nabla E$ for each (k-th) iteration of the upper layer. The preconditioning matrix $\boldsymbol{C}$ used for the upper layer is $\boldsymbol{C}_k$, and, for the lower layer, we use the block diagonalized version of $\boldsymbol{C}_k$, which is constructed by simply omitting the off-diagonal blocks of $\boldsymbol{C}_k$.

2.3 Initial Parameters

It is normally necessary to prepare an initial parameter carefully to ensure that the nonlinear optimization converges successfully. Special devices or sophisticated algorithms might be used to obtain the initial parameters. However, since the optimization method described in Section 2.2 is fast and powerful, it can recover the parameters from a relatively crude initial estimation. The initialization method used herein is described below.

We use a plane perpendicular to the optical axis of the camera as an initial shape (see Fig.1). Even using such crude initial parameters, the shape quickly converges into an appropriate shape if the light positions are reasonable. In order to prepare the light positions, we use the Lambertian reflection property based on Ref. [17]. If a Lambertian surface is lit by a distant light source, it is observed as a three-vector $\boldsymbol{m}^T$, which is described by the product of the intensity η and the direction $\boldsymbol{l}$ of the light, and the normal $\boldsymbol{n}$ and the albedo $\boldsymbol{d}$ (RGB 3-vector) of the surface as $\boldsymbol{m}^T = \eta \boldsymbol{l}^T \boldsymbol{n} \boldsymbol{d}^T$. By collecting measurements $\boldsymbol{m}_{fp}$ at the p-th pixel in the f-th image, this becomes a matrix relation

$$\begin{bmatrix} \boldsymbol{m}_{11}^T & \cdots & \boldsymbol{m}_{1P}^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{m}_{F1}^T & \cdots & \boldsymbol{m}_{FP}^T \end{bmatrix} = \begin{bmatrix} \eta_1 \boldsymbol{l}_1^T \\ \vdots \\ \eta_F \boldsymbol{l}_F^T \end{bmatrix} \begin{bmatrix} \boldsymbol{n}_1 \boldsymbol{d}_1^T & \cdots & \boldsymbol{n}_P \boldsymbol{d}_P^T \end{bmatrix}, \tag{15}$$

or $M = LN$. Ideally, the measurement matrix M is easily constructed and decomposed into the product of two rank-3 matrices, which contain the light directions and the shape. Unfortunately, we could not retrieve correct information directly from this decomposition because the decomposition is not correct when the measurements contain specular reflections and/or shadows. Nor could we determine the distance between the light and the object. Moreover, $M = (LX)(X^{-1}N)$ is also correct for an arbitrary nonsingular 3×3 matrix X, that contains the bas-relief ambiguity [12]. Even so, we can use this decomposition to prepare the initial light positions, which will lead to a correct solution. The decomposition is performed via singular value decomposition, even if specular pixels or pixels in the shadow are included. Let the singular value decomposition be $M = LN$, where $L = (\boldsymbol{u}_1\boldsymbol{u}_2\boldsymbol{u}_3), N = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)(\boldsymbol{v}_1\boldsymbol{v}_2\boldsymbol{v}_3)^T$, and $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$. If the light source moves in front of the object, and the mean of the light positions is near the camera, the most significant singular vector $\boldsymbol{v}_1$ tends to be the mean of all of the input images, and thus $\boldsymbol{u}_1$ approaches a scalar multiple of $(1, 1, \cdots, 1)^T$ because all of the images are approximated by the summation of the mean ($\boldsymbol{v}_1$) and the relatively small deviations ($\boldsymbol{v}_2$ and $\boldsymbol{v}_3$). The structure of $\boldsymbol{u}_1$ implies that the ideal X has the form

$$X = \begin{bmatrix} & & \pm 1 \\ a & b & \\ c & d & \end{bmatrix}, \tag{16}$$

if the object is at the origin of the coordinate system and the camera is on the z-axis. The sign of ± 1 should be selected so that the light is always positioned between the object and the camera. Assuming the sign is positive, we examine the following candidates for X:

$$\begin{bmatrix} & & 1 \\ \pm 1 & & \\ & \pm 1 & \end{bmatrix}, \quad \begin{bmatrix} & & 1 \\ & \pm 1 & \\ \pm 1 & & \end{bmatrix}. \tag{17}$$

From the decomposition, only the direction of the light is obtained. Therefore, the positions are determined by projecting them onto a sphere or a flat plane.

We then conducted the optimization for each candidate X, and selected the best reconstruction. Note that the initial light positions only require qualitative correctness, which are then corrected quantitatively by the following optimization process.

Another possible initialization scheme is to reconstruct the shape by using a sophisticated method, such as those described in Refs. [4,6,8]. The computational cost of this scheme could be less than that using the planar initial shape.

The other parameters are relatively trivial. The light intensities η_f can be assumed to be uniformly 1, if we move a single light bulb around the object. The albedo at the p-th point, $(w_{1p}, w_{2p}, w_{3p})^T$, can be prepared as the mean of all observations of the images, $\sum_f \boldsymbol{m}_{fp}^T/F$. The specular weight w_{4p} can be chosen as 0, assuming that the specular region is relatively small compared with the entire image. The surface roughness ρ is chosen as -10, because it usually

converges at approximately -10 in our experiments. The specular color $\boldsymbol{s}$ is chosen as $(1, 1, 1)^T$, which assumes that the color of the light source is white.

2.4 Incremental Estimation

A complicated reflection model can cause the estimation to be unstable because such a model produces many local minima. To avoid local minima, we first use a coarse model with a limited number of parameters and then gradually upgrade the model and the estimation.

- **Step 1**
  We assume that there is no specular reflection and that all light emittances are uniform. As a result, we only estimate the shape, diffuse weights, and light positions without changing the specular reflection parameters and η_f.
- **Step 2**
  We then add the specular reflection parameters to the set of estimation parameters. Note that we can assume that these steps adjust the ambiguity X described previously, although we do not explicitly have X as a parameter.
- **Step 3**
  Finally, we estimate all of the parameters, including the light emittance for each image. This yields the final estimation result.

The required number of iterations differs for each step. We iterate the PCG process for a predetermined number of times, which is typically 100 for Step 1 and 200 for Steps 2 and 3. If the number of iterations is determined by analyzing the change in the cost function value for the last few iterations, unnecessary iterations would then be omitted or better accuracy would be attained.

2.5 Detection and Correction of an Abnormal Estimation

Although the proposed method works well for most parameters, some parameters tend to converge far away from meaningful values, and as a result, the entire estimation sometimes becomes meaningless.

The specular reflection weight w_{4p} at the edge pixels is particularly volatile, which causes the surface normals at the edge pixels to be incorrectly reconstructed and some light positions to be estimated far away from the other positions. This problem is caused by specular reflections, which means that Steps 2 and 3 are vulnerable. Thus, we added a procedure to avoid this problem. (i) If a specular reflection weight is more than 100 times the median of the weights for the other pixels, it is then corrected to the median value. If the value is negative, it is then corrected to 0. (ii) If the distance of a light from the object is greater than 100 times the median of the other distances, it is then corrected to the median value.

This procedure often improves our estimation process. However, since this procedure is performed without checking the cost function value, the estimation process sometimes collapses. Thus, a more sophisticated approach would be to formulate the reconstruction problem in a quadratic programming algorithm

with several linear constraints, such as certain parameters not being be less than 0, or to formulate using regularization terms to avoid an estimation that is so far away.

2.6 Extensibility

The proposed method can be extended to deal with other image formation models, such as multiple light sources and/or reflection models that are more complicated than the Torrance-Sparrow model. The main difficulty in implementation is the derivation of the Jacobian matrix J_{fp}.

In addition, we can consider interreflections and cast shadows. Even though deriving the Jacobian is the main difficulty, it is straightforward to calculate the residual vector $\boldsymbol{r}_{fp}$ using computer graphics algorithms with respect to these effects.

3 Experiments

3.1 Experimental Setup

In order to validate the proposed method, we estimated the shape and reflectance properties of several objects, as well as the light positions of several real images and numerically generated images. The real images were taken in a room, as shown in Fig. 2, where the only light source was a light bulb held by a human operator. Photographic images of several static objects were captured by a static camera while the light source was moving. We also used a set of images extracted from the Yale Face Database B [4].

The captured images were RGB color images with 8-bit resolution. We did not use a technique that is required for a high dynamic range acquisition.

As an evaluation criterion, we used the RMS error of the estimation, which is defined as

$$\sqrt{\frac{\sum_{fp} \boldsymbol{r}_{fp}^T \boldsymbol{r}_{fp}}{3M}}, \tag{18}$$

where M is the number of terms contained in the cost function, which is at most FP. The proposed method was almost completely validated for a simulation

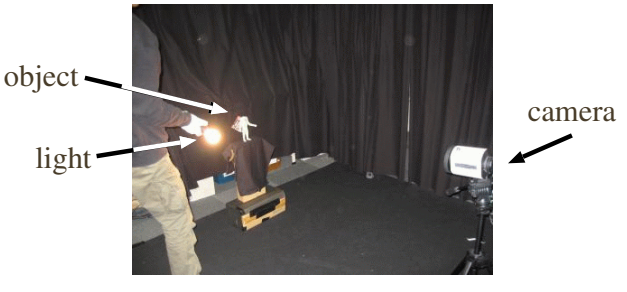


Fig. 2. Experimental setup in a darkened room

image set, where the RMS error approached 10^{-6}, which is an inevitable error because the input measurements were given in single-precision floating point variables.

For the real images, although we would like to compare the obtained shape with the true shape, we did not have ground truth data. Therefore, for now, we evaluated the shape by comparing the obtained result and the real object from various viewpoints, and partially validated the proposed method based on the RMS error, Eq. (18). The quantitative evaluation is left as an important future study. For the light positions, we could compare the results with the ground truth data, because the true light directions were available for one of our data sets and Yale Database B. We also present several videos of the reconstruction results as supplemental material.

3.2 Experiments on Real Images

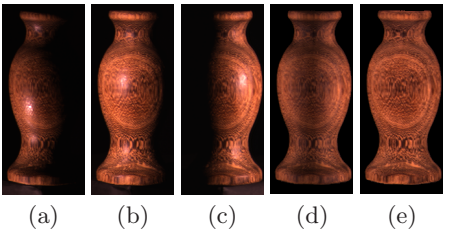
Wooden Figure: A total of 36 images were taken of this figure. Three of the images are shown in Figs. 3 (a)-(c). Each image was of size 128×296, and there were 25,480 foreground pixels.

The extrinsic parameters of the camera were not required for the proposed method, and the intrinsic parameters were simply constructed based on the image size and an approximation of the focal length, as follows:

$$\boldsymbol{P} = \begin{bmatrix} 1000 & 0 & 64 & 0 \\ 0 & 1000 & 148 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \tag{19}$$

Actually, we tested several focal lengths and chose the one that provided the best result. This procedure can be replaced with a camera calibration.

We estimated 127,512 parameters, where the initial shape formed a flat plane, and the initial light positions formed another flat plane. The initial estimate for the weights is as shown in Fig. 3 (d), which is the average of the input images, including Figs. 3(a)-(c), and the resulting diffuse weight is shown in Fig. 3(e).



(a) (b) (c) (d) (e)

Fig. 3. Images of the wooden figure
(a)(b)(c) Examples of the image set,
(d) Mean of the input images, used as the initial weight
(e) Estimated diffuse weights

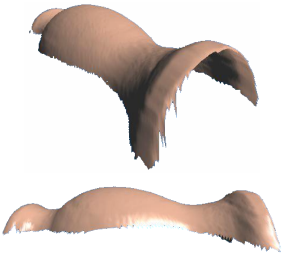


Fig. 4. Estimated shape of the wooden figure

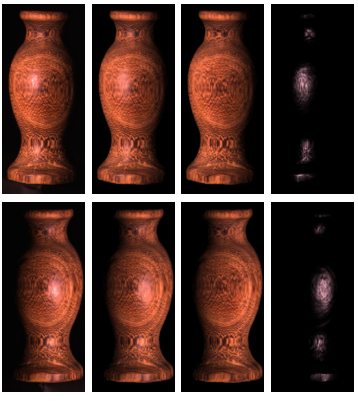


Fig. 5. Input and estimation result of the wooden figure (a) Input images, (b) Reconstructed images, (c) Diffuse components, (d) Specular components

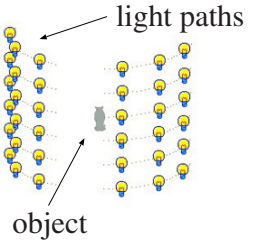


Fig. 6. The object and the circular paths of the light

Fig. 7. A toy model (a) Overall view and target region, (b) Estimated shape, (c) Input images, (d) Reconstructed images

The estimated shape is as shown in Fig. 4. We can confirm that a rotationally symmetric shape was successfully reconstructed without considerable noise.

Figure 5 shows, from left to right, the input images, the reconstructed images, and the reconstruction of the diffuse components and specular components, for two different images. The RMS error was approximately 5, which is 2% of the intensity range, and the error indicates that the model of Eq. (2) effectively approximated the input images. In addition, the diffuse and specular components were meaningfully separated.

During this experiment, the light bulb traveled along several controlled circular paths, as shown in Fig. 6, even though that information was not added to the optimization. By comparing the estimated positions and the controlled trajectory, the average error in the estimated direction was calculated to be 25

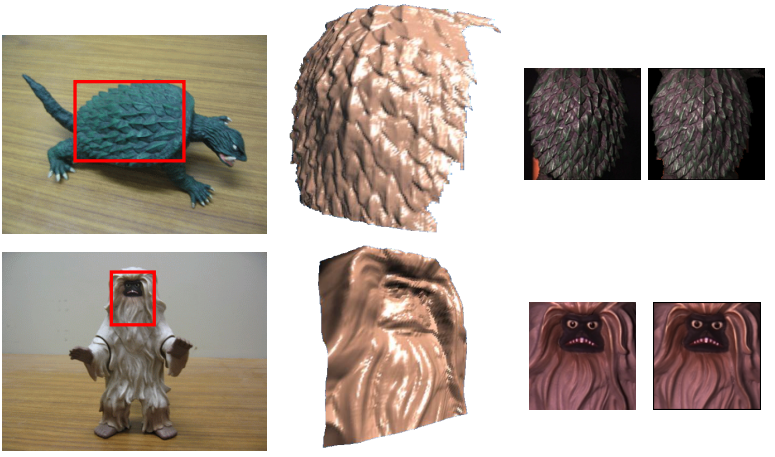


Fig. 8. The other toy models (a) Overall views and target regions, (b) Estimated shapes, (c) Examples of the input images, (d) Reconstructions

Fig. 9. Input images of a human face

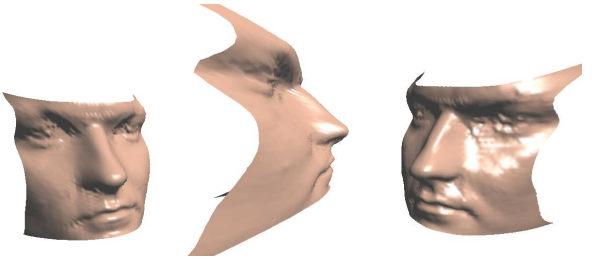
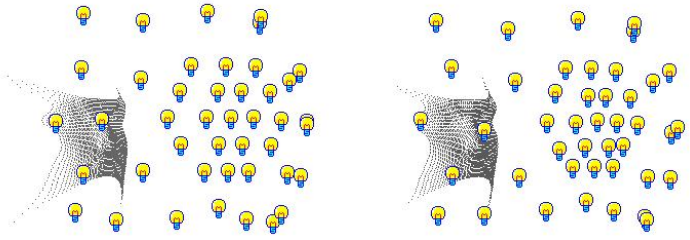


Fig. 10. Estimated shapes of the human face

degrees. This is a rather poor result, even though the reconstructed shape does not seem to be greatly distorted. Intrinsically, the estimation of the light positions is ill-conditioned, since the specular intensity is a consequence of multiple factors, including the positions of the light, the curvature of the object, and the roughness of the surface.

Toy Models: Several vinyl models were also used to test the proposed method. For each model, thirty images were captured in a darkened room. The overall views of the objects are shown in Figs. 7 and 8, along with the estimated shapes. They also show (c) input images and (d) reconstructed images. We can confirm that the input images were well reconstructed by the proposed model, Eq. (2). We compared the reconstructed shapes with the real models from various viewpoints and confirmed that the shapes were successfully reconstructed.



(a) (b)

Fig. 11. Estimated light positions of the human face (a) True light positions, (b) Estimated light positions

Human Face: Images extracted from the Yale Face Database B [4] were used to validate the proposed method. Figure 9 shows examples of the input images of subject #7 in the database, and Fig. 10 shows the estimated shape from a set of 43 images within Subset 4. The light positions are documented in the database, and the average estimation error of the light direction was 9.5 degrees with a standard deviation of 4.2 degrees. We did not use known light directions for our optimization.

4 Conclusions

In the present paper, we described a method that can be used for the recovery of a shape, reflectance property, and light positions that does not need any special devices other than a camera and a light source in a darkened room. For a set of numerically generated images, the method recovers almost the exact parameters. For real images, the method recovers satisfactory shapes. The method is based on the Levenberg-Marquardt algorithm combined with the preconditioned conjugate gradient algorithm for handling a large-scale nonlinear optimization problem. We used a three-step algorithm (coarse model to fine model) to increase the stability of the process. In addition, we do not need precise calibration or initialization based on special devices such as range finders, spherical mirrors, or robotic arms.

The method is based on the Torrance-Sparrow model and the assumption that a single point light source will be used, which might limit the applicability of the method. Future research will include the replacement of the original models with more flexible models and the use of multiple cameras.

References

1. Georghiades, A.: Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In: ICCV 2003, pp. 816–823 (2003)
2. Georghiades, A.: Recovering 3-d shape and reflectance from a small number of photographs. In: Eurograpgics Symposium on Rendering, pp. 230–240 (2003)

3. Sato, Y., Ikeuchi, K.: Reflectance analysis for 3d computer graphics model generation. CVGIP 58(5), 437–451 (1996)
4. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. on PAMI 23(6), 643–660 (2001)
5. Lensch, H.P.A., Kautz, J., Goesele, M., Heidrich, W., Seidel, H.P.: Image-based reconstruction of spatial appearance and geometric detail. ACM Trans. on Graphics 22(3), 1–27 (2003)
6. Hertzmann, A., Seitz, S.M.: Example-based photometric stereo: Shape reconstruction with general varying brdfs. IEEE Trans. on PAMI 27(8), 1254–1264 (2005)
7. Mallick, S.P., Zickler, T.E., Kriegman, D.J., Belhumeur, P.N.: Beyond lambert: Reconstructing specular surfaces using color. In: CVPR 2005, vol. 2, pp. 619–626 (2005)
8. Sato, I., Okabe, T., Yu, Q., Sato, Y.: Shape reconstruction based on similarity in radiance changes under varying illumination. In: ICCV (2007)
9. Mercier, B., Meneveaux, A., Fournier, A.: A framework for automatically recovering object shape, reflectance and light sources from calibrated images. IJCV 73(1), 77–93 (2007)
10. Yu, Y., Xu, N., Ahuja, N.: Shape and view independent reflectance map from multiple views. IJCV 73(2), 123–138 (2007)
11. Goldman, D., Curless, B., Hertzmann, A.: Shape and spatially-varying brdfs from photometric stereo. In: ICCV 2005, pp. 230–240 (2005)
12. Belhumeur, P., Kriegman, D., Yuille, A.: The bas-relief ambiguity. IJCV 35(1), 33–44 (1999)
13. Boivin, S., Gagalowicz, A.: Image-based rendering of diffuse, specular and glossy surfaces from a single image. SIGGRAPH, 107–116 (2001)
14. Paterson, J., Claus, D., Fitzgibbon, A.: Brdf and geometry capture from extended inhomogeneous samples using flash photography. EUROGRAPHICS 24(3), 383–391 (2005)
15. Triggs, B., McLauchlan, P.F., Hartley, R., Fitzgibbon, A.W.: Bundle adjustment — a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, pp. 298–375. Springer, Heidelberg (2000)
16. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. 11, 431–441 (1963)
17. Shashua, A.: Geometry and photometry in 3d visual recognition, Ph. D. thesis, Dept. Brain and Cognitive Science, MIT (1992)