

Combination Photometric Stereo Using Compactness of Albedo and Surface Normal in the Presence of Shadows and Specular Reflection

Naoto Ienaga¹(✉), Hideo Saito¹, Kouichi Tezuka², Yasumasa Iwamura²,
and Masayoshi Shimizu²

¹ Department of Information and Computer Science, Keio University, 3-14-1,
Hiyoshi, Kohoku-ku Yokohama-shi, Kanagawa 223-8522, Japan

{ienaga,saito}@hvrl.ics.keio.ac.jp

<http://www.hvrl.ics.keio.ac.jp>

² Fujitsu Laboratories Ltd., 4-1-1, Kamikodanaka, Nakahara-ku Kawasaki-shi,
Kanagawa 211-8588, Japan

{ktezuka,iwamura.yasumas,shimizu.masa}@jp.fujitsu.com

<http://jp.fujitsu.com/group/labs>

Abstract. We present a novel combination photometric stereo which can estimate surface normals precisely even for images including shadows and specular reflection. We can use photometric stereo if there are more than three input images. Therefore we can employ photometric stereo with ${}_nC_3$ combinations for n input images. We make 3D distribution of albedos and surface normals estimated from pixel intensities of ${}_nC_3$ pixel combinations. In the distribution, we define a novel value “compactness” to distinguish pixels which are included in neither shadows nor specular reflection from pixels which are included in shadows or specular reflection. Through experimental results, we demonstrate that the proposed method can estimate surface normals in the presence of shadows and specular reflection. Moreover the proposed method is superior to previous works in better accuracy.

Keywords: Photometric stereo · Shadow · Specular reflection · 3D shape reconstruction

1 Introduction

As found in the emergence of 3D printers, 3D shape reconstructions have recently drawn attention. Photometric stereo is the well-known effective method to obtain a 3D shape of a target object. 3D shapes are reconstructed from surface normals which are provided by photometric stereo. Photometric stereo can estimate the albedo (the ratio of incident to reflected light) and the surface normal of each pixel of some input images in which only the light direction changes. However, photometric stereo assumes that a reflection of surface of target object follows

Lambertian reflectance. Therefore photometric stereo does not work properly in the presence of shadows and (or) specular reflection in which the assumption of Lambertian reflectance does not be satisfied.

Photometric stereo needs at least 3 images to estimate albedos and surface normals. When we have n images, we can employ photometric stereo with ${}_nC_3$ combinations. We can compute a “triplet”, which has three values of p , q (a surface normal) and an albedo, from one of ${}_nC_3$ pixel combinations of a certain pixel of n input images. Then in the 3D space of p , q and an albedo, we can consider a distribution of ${}_nC_3$ triplets and we define a “compactness” of triplets. The compactness indicates the degree of concentration of triplets representing albedos and surface normals. In this paper, we propose to use the compactness so that we can remove pixels which do not obey Lambertian model.

This paper is organized as follows. In section 2, we make mention of related works. In section 3, combination photometric stereo using the compactness of albedos and surface normals is proposed. Experimental results are given in section 4 and section 5 is devoted to concluding remarks.

2 Related Works

Photometric stereo has been studied for a long time in consideration of the influence of shadows and specular reflection. Chung et al. [1] introduced an approach of estimation of parameters of Ward BRDF model by using cast shadows to overcome the influence of wide specular lobes. In [2], Hern et al. or in [3], Barsky et al. argued that methods in cases of getting 3 or 4 images. Therefore each pixel can be included in shadows or specular reflection in at most one image of n input images. However, considering a utilization of photometric stereo, it is hard to think about only less than 4 images are gained. It is rather more possible that though we can capture many images, each pixel is included in shadows or specular reflection areas in some images. For this reason, the proposed method permits shadows and specular reflection to extend plural images while the proposed method needs some input images.

Chandraker et al. [4] and Dulac et al. [5] proposed similar algorithms. Dulac et al. focus on only the shadow problem and presume the darkest pixel must be a shadow. First, they compute a surface normal from the darkest pixel, the brightest pixel and the third brightest pixel. Second, they compare observed pixel intensities with pixel intensities which are obtained from a back calculation using the surface normal obtained in the previous step. Third, they try to remove the darkest pixel if the difference is bigger than a pre-defined value. These steps are repeated until the difference will be smaller. For more details, please refer to [5].

Miyazaki et al. also tackled this challenge using graph cut in [6] and using a median value in [7]. In [7], they compute ${}_nC_3$ surface normals for all pixels like the proposed method. Medians of sets including surface normals of 4-connected pixels and ${}_nC_3$ surface normals and averages of surface normals of 4-connected pixels are used for computing conclusive surface normals. Furthermore they remove reflections of a transparent display case because the problem

they addressed in [7] is a situation that an object is placed in a transparent display case like museums.

3 Proposed Method

The proposed method is performed for each pixel by starting the estimation of ${}_nC_3$ triplets that represent albedos and surface normals. Next, we compute the compactness to select pixels which are not included in shadows and specular reflection.

3.1 Combinations for Triplets

First of all, we summarize the basic principle of photometric stereo. We assume \mathbf{V} represents a n dimensions vector including pixel intensities of a certain pixel of n input images, ρ is an albedo, s is a light intensity, \mathbf{L} is a known matrix of light directions of input images and \mathbf{n} is a surface normal. If a reflection of surface of target object follows Lambertian reflectance, \mathbf{V} can be expressed as follow:

$$\mathbf{V} = \rho s \mathbf{L} \mathbf{n} \quad (1)$$

We assume $s = 1$. Hence:

$$\mathbf{L}^{-1} \mathbf{V} = \rho \mathbf{n} \quad (2)$$

\mathbf{n} is a unit vector. Therefore the length of the left side of the equation (2) is ρ :

$$\rho = \|\mathbf{L}^{-1} \mathbf{V}\| \quad (3)$$

Then \mathbf{n} is as follow:

$$\mathbf{n} = \frac{\mathbf{L}^{-1} \mathbf{V}}{\|\mathbf{L}^{-1} \mathbf{V}\|} = (n_x, n_y, n_z)^T \quad (4)$$

We use a x direction vector $\mathbf{r}_x = (1, 0, p)^T$ parallel to an object surface and a y direction vector $\mathbf{r}_y = (0, 1, q)^T$ to remove the redundancy of surface normals. Since p and q represent slopes of a surface in the x and the y directions respectively, they are called gradients of a surface. A surface normal can be computed by taking the cross-product of these two vectors. The relationship between these two vectors and the equation (4) is as follow.

$$\mathbf{n} = \mathbf{r}_x \times \mathbf{r}_y = \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix} = \begin{pmatrix} n_x/n_z \\ n_y/n_z \\ 1 \end{pmatrix} \quad (5)$$

We compute p , q and ρ of ${}_nC_3$ pixel combinations of each pixel of input images according to the equation (3) and (5). We define these $p, Cq, C\rho$ as a ‘‘triplet’’.

3.2 Compactness

Fig. 1 shows the distribution of ${}_n C_3$ triplets in the 3D space of p , q and ρ for a pixel of Fig. 7 (shown in Sec. 4.2). Circles (hereinafter referred to as “correct triplets”) are computed from pixels which are not included in shadows and specular reflection (hereinafter referred to as “correct pixels”). Meanwhile triangles (hereinafter referred to as “wrong triplets”) are computed from pixels which are included in shadows (pixels included in shadows or specular reflection, hereinafter referred to as “wrong pixels”).

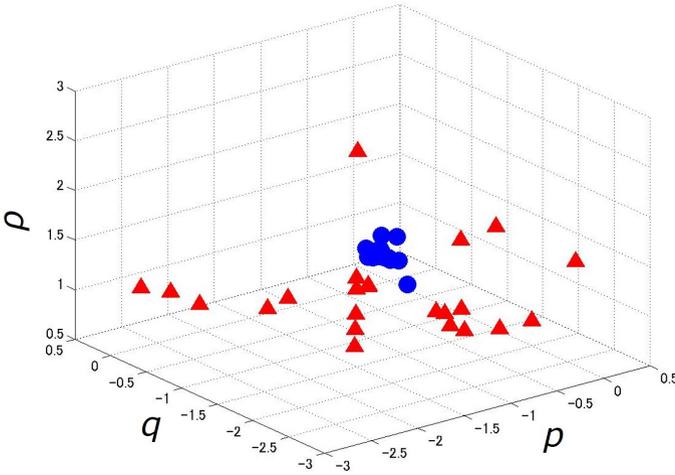


Fig. 1. Distribution of triplets of a certain pixel position. Since there are eight input images, there are 56 (${}_8 C_3$) triplets of the pixel. Triplets are manually classified whether it is correct or not. All correct triplets are almost concentrating at a particular position, while most of wrong triplets are widely distributed.

The purpose of the proposed method is to select correct pixels, which are not included in shadows and specular reflection. Therefore we aim to locate correct triplets because if we can do it, what we have to do besides is only about checking pixels constituting correct triplets. Pixels constituting correct triplets are correct pixels (this is our definition). However, we need a value that indicates correctness of triplets to find correct triplets.

Correct triplets are concentrating at a particular position as shown in Fig. 1 because correct triplets have similar values. Correct triplets should have same values if all pixel intensities of input images perfectly obey Lambertian model without any error. This property is useful to find correct triplets. Then we count other triplets around each triplet to utilize the property. The more triplets have others, the more it is correct.

By using this property, we propose the definition of compactness of triplet i as follow:

$$Compactness(i) = \sum_{\substack{j=1 \\ j \neq i}}^{nC_3} U(i, j) \tag{6}$$

$$U(i, j) = \begin{cases} 1 & \text{if } \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2} < th_{dpq}, |\rho_i - \rho_j| < th_{d\rho} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

We compute the compactness for each triplet i ($i = 1, \dots, nC_3$). Correct triplets have a large compactness because they tend to concentrate. th_{dpq} and $th_{d\rho}$ are thresholds of surface normals and albedos. The equation (7) means that $U(i, j)$ is 1 only when triplet j is within the thresholds of triplet i . As a result, the compactness means the number of other triplets around each triplet.

Finally we vote to judge which pixels are suspected to be shadows or specular reflection instead of utilizing the surface normal of the triplet which has the maximum compactness. Because when we use constantly only three pixels though there are n pixels than when we use pixels which are not included in shadows and specular reflection as many as possible, we can calculate surface normals more accurately. We vote with only triplet(s) which have the maximum compactness. Concretely, we count pixels constituting all triplets within th_{spq} and $th_{s\rho}$ of the triplets which have the maximum compactness. For example, in the case of Fig. 2, the pixel 1 will get 2 votes, the pixel 2 will get 3 votes, the pixel 3 will not get any votes etc.

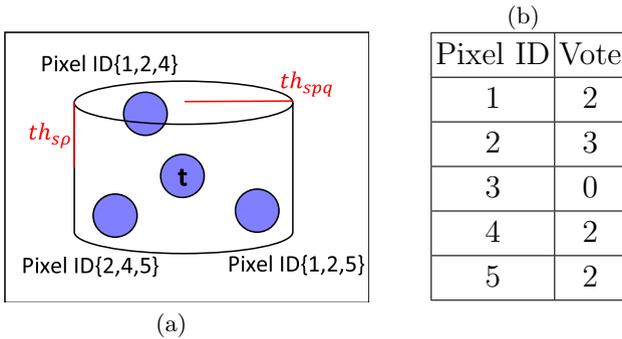


Fig. 2. Example of voting. The cylinder represents the volume within th_{spq} and $th_{s\rho}$ around the target triplet denoted by “t”. There are three other triplets within the cylinder volume. The right table shows that the vote of the pixels constituting those triplets.

The equations (8) and (9) must be satisfied. Because th_{dpq} and $th_{d\rho}$ should be small in order to determine the triplets which have the maximum compactness

while th_{spq} and $th_{s\rho}$ should be some higher value so that more correct triplets can be extracted. Triplets which have the maximum compactness may not be one and even if so, we vote using the same table. Since correct triplets should have the maximum compactness and there should be many other correct triplets around them, votes of correct pixels should be more than votes of wrong pixels. Then we choose pixels which have votes more than *average – standard deviation* for calculation of conclusive surface normals.

$$th_{spq} > th_{dpq} \quad (8)$$

$$th_{s\rho} > th_{d\rho} \quad (9)$$

As stated above, we have to set initially 4 thresholds th_{dpq} , $th_{d\rho}$, th_{spq} , $th_{s\rho}$ and in fact another threshold th_f . As a simple way of th_{dpq} and $th_{d\rho}$ determination, we use the following algorithm: When all triplets have less compactness than th_f with initial th_{dpq} and $th_{d\rho}$, we add the minimum distance between triplet x and triplet y outside of th_{dpq} and $th_{d\rho}$ of triplet x to initial th_{dpq} and $th_{d\rho}$ alternately. Then re-calculate the compactness. Thanks to this algorithm, setting of th_{dpq} and $th_{d\rho}$ is very easy because we just set them to small values for choosing triplets which have the maximum compactness. However, we are immune from setting th_{dpq} and $th_{d\rho}$ to too small value due to the aforesaid algorithm.

3.3 Recovering 3D Shape

After computing surface normals by using pixels selected in the previous step, we integrate surface normals and recover a 3D shape. In this process, we use Xu et al. [8]’s method.

4 Experiments and Results

4.1 Effect on Shadows and Specular Reflection

In this experiment, we show the effect of the proposed method on shadows and specular reflection, comparing conventional photometric stereo [9]. Conventional photometric stereo uses whole input images although they contain shadows and specular reflection. Fig. 3 shows one of input images produced by POV-Ray [10] and the obj file¹ downloaded from [11]. Fig. 4 shows the 3D shapes obtained from conventional photometric stereo (left) and the proposed method (right).

Fig. 5 shows the effect on specular reflection. With conventional photometric stereo (left), we can observe the projection in the center of the object because of specular reflection as shown in Fig. 3, while the projection does not appear with the proposed method (right).

The effect on shadows is shown in Fig. 6. Shadows make the form of conventional photometric stereo (left) almost a trapezoid unnaturally. While the proposed method (right) is more round and has a natural shape.

¹ Josea, “Clay vase garden pottery,” (http://artist-3d.com/free_3d_models/dnm/model_disp.php?uid=3792)



Fig. 3. One of eight input images contain shadows and specular reflection

Fig. 4. 3D shapes. Left: obtained from conventional photometric stereo [9]. Right: proposed method.

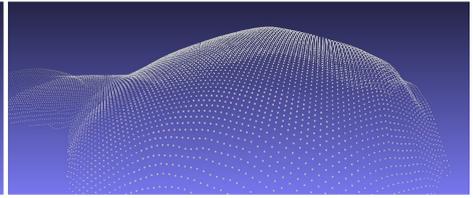
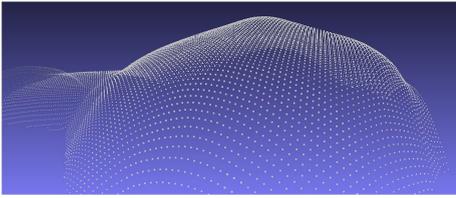


Fig. 5. Effect on specular reflection. 3D points are observed from the bottom left. Left: recovered shape by conventional photometric stereo [9]. Right: proposed method.

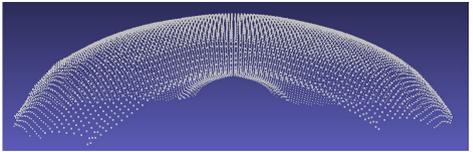
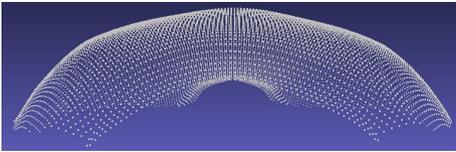


Fig. 6. Effect on shadows. 3D points are observed from the bottom. Left: recovered shape by conventional photometric stereo [9]. Right: proposed method.



Fig. 7. Three of eight input images do not contain specular reflection but shadows.

4.2 Comparing with Related Works Using Actual Images

This experiment has been carried out using Xiong et al. [12]’s data² shown in Fig. 7. We compare the proposed method with conventional photometric stereo [9], Miyazaki et al. [7] and Dulac et al. [5]. Seeing Fig. 7, input images do not contain much specular reflection but include many shadows. We chose these images to compare with Dulac’s method which focuses only shadows.

We compare angles defined by the obtained surface normal of each method and the ground truth. In Fig. 8, black pixels indicate more than 15 degrees and white pixels indicate 0 degrees (equal to the ground truth). Top left is conventional photometric stereo. The most of the regions with shadows are black. Top right and bottom left are Miyazaki’s and Dulac’s respectively. The black regions are less than those of conventional photometric stereo. However, their methods could not eliminate the black regions enough. Bottom right is the proposed method in which there are the fewest black regions. We also evaluate the proposed method quantitatively. Table 1 also shows that the proposed method is more effective than the other methods.

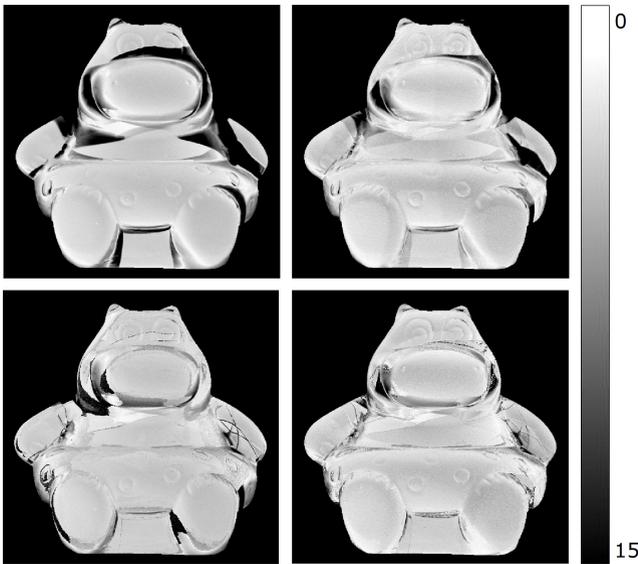


Fig. 8. Gray scale images. Angles 0-15 and pixel intensities 255-0 correspond. Top left: conventional photometric stereo [9]. Top right: Miyazaki et al. [7]. Bottom left: Dulac et al. [5]. Bottom right: proposed method.

Fig. 9 shows the 3D shapes of conventional photometric stereo (left) and the proposed method (right). Conventional photometric stereo couldn’t compute

² The information to download is here, (<http://vision.seas.harvard.edu/qsf/Data.html>)

surface normals accurately because of shadows. As a result, the hollow of the eyes and the nose swells out. Then we can conclude that 3D shapes tend to swell out in the process of integrating inaccurate surface normals.

Table 1. Root mean squared error (here, error means degrees).

Conventional photometric stereo [9]	Miyazaki et al. [7]	Dulac et al. [5]	proposed method
7.76	4.97	4.93	3.93

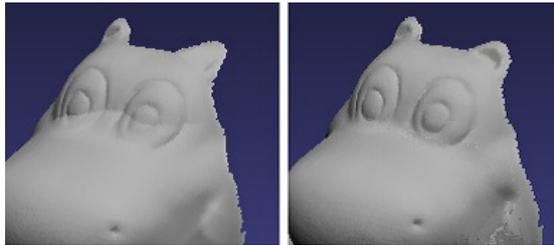


Fig. 9. 3D shapes. Left: conventional photometric stereo [9]. Right: proposed method.

4.3 Demonstration of Practical Use

In the end, we show experimental results using actual images we captured to demonstrate a practical use. We have developed the small device which supplies

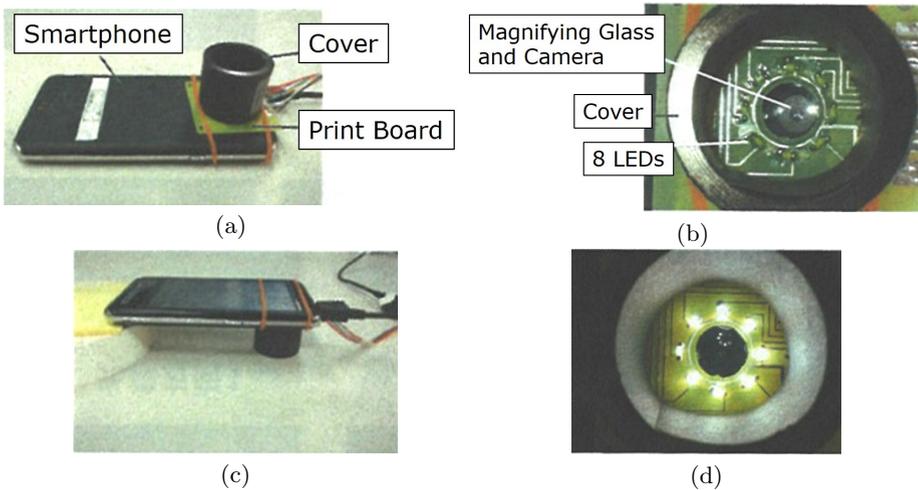


Fig. 10. Device we have developed: (a) (b) The details of the device. The LEDs turn on in turn and the camera captures eight images; (c) The state at the time of capturing; (d) The figure when all the LEDs turn on.



Fig. 11. Two of eight input images.

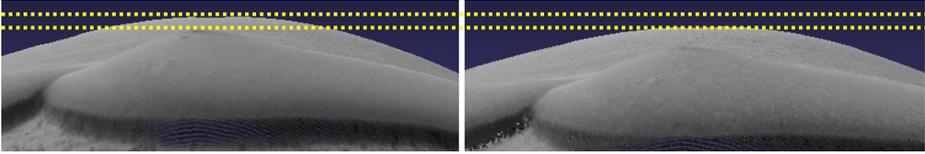


Fig. 12. 3D shapes. Left: obtained from conventional photometric stereo [9]. Right: proposed method. The yellow dotted lines are parallel to the images. The upper line and the lower line are as high as the center of the 3D shape of conventional photometric stereo and the bulge of the edge of the 3D shape of conventional photometric stereo. However, two lines are higher than those of proposed method.

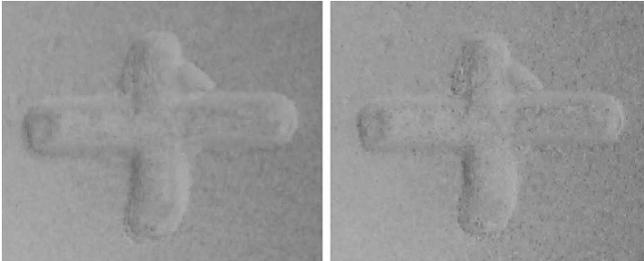


Fig. 13. Chinese character “Ten” of the 3D shapes. Left: obtained from conventional photometric stereo [9]. Right: proposed method.

eight images varying in the light positions and can be attached to smartphones (Fig. 10). Fig. 11 shows the silicon replica of the Japanese 50 yen coin captured by our device. The images include specular reflection and some shadows.

Fig. 12 are the figures of the 3D shapes which looked at from the bottom. The left is conventional photometric stereo [9] and the right is the proposed method. The left is sweller than the right like Fig. 9. It also seems that Chinese character “Ten” of conventional photometric stereo spreads more largely because of the swelling (Fig. 13). Therefore we can say our surface normals are more accurate than those of conventional photometric stereo from the conclusion of Fig. 9.

5 Conclusion

We have proposed an approach for the removal of the impact of shadows and specular reflection in photometric stereo. We utilize ${}_nC_3$ combinations of each pixel of n input images and define the novel value “compactness” based on the considerations of triplets which are comprised of pixels which are included in neither shadows nor specular reflection. Experimental results demonstrated that the proposed technique removes the impact of shadows and specular reflection, leads to the better results than previous works and shows the possibility of a practical use.

References

1. Chung, H., Jia, J.: Efficient photometric stereo on glossy surface with wide specular lobes. In: Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), June 2008
2. Hernández, C., Vogiatzis, G., Cipolla, R.: Overcoming shadows in 3-source photometric stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence* **33**(2), 419–426 (2011)
3. Barsky, S., Petrou, M.: The 4-source photometric stereo technique for three-dimensional surface in the presence of highlights and shadows. *IEEE Trans. Pattern Analysis and Machine Intelligence* **25**(10), 1239–1252 (2003)
4. Chandraker, M., Agarwal, S., Kriegman, D.: Shadowcuts: photometric stereo with shadows. In: Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), June 2007
5. Dulac, A., Martedi, S., Saito, H., Tezuka, K., Shimizu, M.: Combination method: photometric stereo with shadows. In: Proc. 2014 Irish Machine Vision and Image Processing (IMVIP2014), pp. 137–142, August 2014
6. Miyazaki, D., Ikeuchi, K.: Photometric stereo using graph cut and m-estimation for a virtual tumulus in the presence of highlights and shadows. In: Proc. Workshop on Applications of Computer Vision in Archaeology (ACVA2010), pp. 70–77, June 2010
7. Miyazaki, D., Hara, K., Ikeuchi, K.: Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision* **86**(2–3), 229–242 (2010)
8. Xu, J., Pachauri, D.: Project 3: photometric stereo, January 17, 2015. <http://pages.cs.wisc.edu/~jiaxu/projects/Stereo/>
9. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. *Optical Engineering* **19**(1), 139–144 (1980)
10. POV-Ray, January 21, 2015. <http://www.povray.org/>
11. ARTIST-3D.COM, January 21, 2015. <http://artist-3d.com/>
12. Xiong, Y., Chakrabarti, A., Basri, R., Gortler, S.J., Jacobs, D.W., Zickler, T.: From shading to local shape. *IEEE Trans. Pattern Analysis and Machine Intelligence* **37**(1), 67–79 (2015)
13. Horn, B.K.P.: *Robot Vision*. The MIT Press, Massachusetts (1986)