DIFFUSE-SPECULAR SEPARATION OF MULTI-VIEW IMAGES UNDER VARYING ILLUMINATION

Kouki Takechi *Takahiro Okabe*

Department of Artificial Intelligence
Kyushu Institute of Technology

ABSTRACT

Separating diffuse and specular reflection components is important for preprocessing of various computer vision techniques such as photometric stereo. In this paper, we address diffuse-specular separation for photometric stereo based on light fields. Specifically, we reveal the low-rank structure of the multi-view images under varying light source directions, and then formulate the diffuse-specular separation as a low-rank approximation of the 3rd order tensor. Through a number of experiments using real images, we show that our proposed method, which integrates the complement clues based on varying light source directions and varying viewing directions, works better than existing techniques.

Index Terms— photometric stereo, reflectance separation, light field, higher-order SVD

1. INTRODUCTION

Photometric stereo is a technique for estimating surface normals from the shading observed on an object surface under varying illumination. In contrast to multi-view stereo for estimating the depths of possibly sparse corresponding points, photometric stereo can estimate pixel-wise surface normal, *i.e.* densely recover the shape of an object.

In general, the reflected light observed on an object surface consists of a diffuse reflection component and a specular reflection component. The conventional photometric stereo [1] assumes the Lambert model and known light sources, and then specular reflection components are outliers to be removed. On the other hand, in the context of uncalibrated photometric stereo [2] assuming the Lambert model and unknown light sources, specular reflection components are an important clue for uniquely recovering the shape of an object [4, 5] by resolving the Generalized Bas-Relief (GBR) ambiguity [3]. Therefore, separating diffuse and specular reflection components of the images of an object taken under varying light source directions is an important topic to be addressed for photometric stereo.

In this paper, we address diffuse-specular separation for photometric stereo based on light fields. Since light field cameras [6] can capture multi-view images in a single shot, we study diffuse-specular separation of multi-view images under varying illumination, and propose a robust method by integrating a clue based on varying light source directions and a clue based on varying viewing directions.

Specifically, we represent the multi-view images of an object under varying light source directions as the 3rd order tensor whose axes are the position on the object surface, the light source direction, and the viewing direction. Then, we reveal the low-rank structure of the tensor on the basis of the facts that the brightness of a diffuse reflection component is independent of viewing directions and that it is given by the inner product between a light source direction and a normal, *i.e.* bilinear with respect to them. Based on the above theoretical insight, our proposed method separates diffuse and specular reflection components via low-rank tensor approximation using higher-order Singular Value Decomposition (SVD) [7]. In addition, SVD with missing data [8] is incorporated into our method in order to take outliers such as specular reflection components and shadows into consideration.

To demonstrate the effectiveness of our proposed method, we conducted a number of experiments using real images. We show that our method, which integrates the complement clues based on varying light source directions and varying viewing directions, performs better than existing techniques.

2. RELATED WORK

Existing techniques for diffuse-specular separation can be classified into the following four approaches.

Varying viewing directions: The reflected light observed on an object surface consists of a viewpoint-invariant diffuse reflection component and a viewpoint-dependent specular reflection component. Then, we can consider the darkest intensity at a point on the surface seen from varying viewing directions as the diffuse reflection component [9, 10]. In contrast to this approach, our proposed method makes use of varying light source directions as well as varying viewing directions.

Varying light source directions: An image of a Lambertian object under an arbitrary directional light source is represented by a linear combination of three basis images of the object [11]. Then, the approach based on varying light source

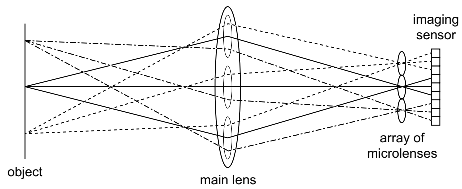

Fig. 1. A microlens-based light field camera: the rays emitted from a point on an object surface are focused on a microlens.

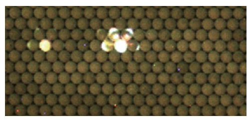

Fig. 2. The closeup of a raw light field image.

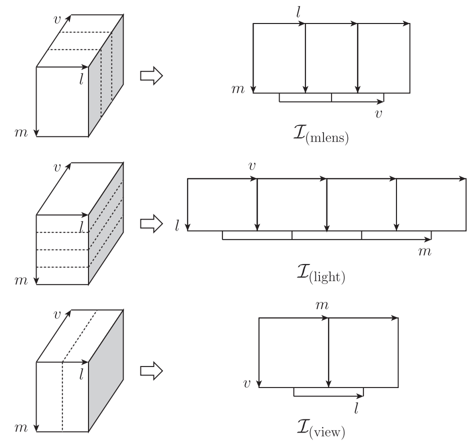

Fig. 3. Unfolding the 3rd order tensor with respect to each axis: $M = 4$, $L = 2$, and $V = 3$ for display purpose.

directions is often used for photometric stereo, and a number of techniques such as low-rank matrix recovery [11, 12] have been proposed. In contrast to this approach, our method makes use of not only varying light source directions but also varying viewing directions.

Polarizations: When we observe the reflected light from an object surface illuminated by polarized light, specular reflection components are polarized whereas diffuse reflection components are unpolarized [13]. Then, the difference of the polarization states can be used for diffuse-specular separation. The polarization-based approach requires a set of images taken by placing linear polarizing filters in front of a light source and a camera, and rotating one of them.

Colors: The color-based approach makes use of the difference between the colors of a specular reflection component and a diffuse reflection component. According to the dichromatic reflection model [14], the former is the same as the light source color whereas the latter depends on the spectral reflectance of a surface. This approach does not work well for objects with low-saturation reflectance and for narrow-band light sources such as LED, because the color of a specular reflection component is almost the same as the color of a diffuse one for them.

3. PROPOSED METHOD

3.1. Raw Light Field Image

In this study, we assume microlens-based light field cameras in which an array of microlenses is placed in front of an imaging sensor. As shown in Fig. 1, our proposed method assumes that a camera focuses on an object of interest [1], *i.e.* the rays emitted from a point on the object surface are focused on a microlens. Therefore, the pixels just below a microlens capture the intensities of the same point on the object surface seen from varying viewing directions. We can observe the silhouette of an array of microlenses in the closeup of a raw light field image in Fig. 2.

3.2. Structure of 3rd Order Tensor

Let us consider a set of images of a static object taken by a static light field camera but under varying light source directions. We denote the pixel value of the v-th ($v = 1, 2, 3, ..., V$) pixel, *i.e.* observed from the v-th viewing direction, in the m-th ($m = 1, 2, 3, ..., M$) miclolens, *i.e.* observed at the m-th position on the object surface, under the l-th ($l = 1, 2, 3, ..., L$) light source direction by i_{mlv}[2]. Our proposed method represents those pixel values as the 3rd order tensor $\mathcal{I}$ whose axes are the position on the object surface m, the light source direction l, and the viewing direction v.

As shown in Fig. 3, we unfold the 3rd order tensor $\mathcal{I}$ with respect to each axis, and then consider three matrices $\mathcal{I}_{(\mathrm{mlens})}$, $\mathcal{I}_{(\mathrm{light})}$, and $\mathcal{I}_{(\mathrm{view})}$ whose sizes are $M \times (LV)$, $L \times (VM)$, and $V \times (ML)$ respectively. Hereafter, we show that those matrices are low-rank, when the all pixel values are described by the Lambert model.

First, we consider the matrix $\mathcal{I}_{(\mathrm{mlens})}$ at the top of Fig. 3. This matrix consists of V blocks with the size of $M \times L$. According to the Lambert model, the pixel values of the l-th column in the first block are given by

$$\begin{pmatrix} \rho_1 \boldsymbol{n}_1^\top \\ \vdots \\ \rho_M \boldsymbol{n}_M^\top \end{pmatrix} \boldsymbol{s}_l. \tag{1}$$

[1]We consider the target objects and/or imaging conditions such that parallaxes are negligible, *e.g.* near-planar objects.

[2]We assume that M, L, and V are significantly larger than 3.

Here, ρ_m and $\boldsymbol{n}_m$ are the albedo and normal at the position on the object surface corresponding to the m-th microlens, and $\boldsymbol{s}_l$ is the light source vector whose norm and direction are the intensity and direction of the l-th light source respectively. Since the degree of freedom of the light source vector is 3 in general [3], the rank of the first block is 3. This means that an image of a Lambertian object under an arbitrary directional light source is represented by a linear combination of three basis images of the object [11]. In addition, since the brightness of a diffuse reflection component is independent of viewing directions, the other blocks are the same as the first block, and therefore the rank of the matrix $\mathcal{I}_{(\text{mlens})}$ is also 3.

Second, let us consider the matrix $\mathcal{I}_{(\text{light})}$ at the middle of Fig. 3. This matrix consists of M blocks with the size of $L \times V$. Since the brightness of a diffuse reflection component is viewpoint-invariant, the rank of each block is 1. According to the Lambert model, the pixel values of an arbitrary column in the m-th block are given by

$$\begin{pmatrix} \boldsymbol{s}_1^\top \\ \vdots \\ \boldsymbol{s}_L^\top \end{pmatrix} \rho_m \boldsymbol{n}_m. \tag{2}$$

Since the degree of freedom of the vector $\rho_m \boldsymbol{n}_m$ is 3 in general [4], the rank of the matrix $\mathcal{I}_{(\text{light})}$ is also 3.

Third, we consider the matrix $\mathcal{I}_{(\text{view})}$ at the bottom of Fig. 3. It is obvious that the rank of the matrix $\mathcal{I}_{(\text{view})}$ is 1, since the brightness of a diffuse reflection component does not depend on viewing directions. Hence, the structure of the 3rd order tensor $\mathcal{I}$ is revealed; the matrices given by unfolding the tensor with respect to each axis are low-rank.

3.3. Low-Rank Tensor Approximation

Based on the low-rank structure of the 3rd order tensor, our method separates diffuse and specular reflection components via low-rank tensor approximation using higher-order SVD [7]. First, we decompose the matrix $\mathcal{I}_{(\text{mlens})}$ as

$$\mathcal{I}_{(\text{mlens})} = U_{(\text{mlens})} \Sigma_{(\text{mlens})} V_{(\text{mlens})}^\top \tag{3}$$

by using SVD. Since the rank of the matrix $\mathcal{I}_{(\text{mlens})}$ is 3, we obtain the $M \times 3$ matrix $\hat{U}_{(\text{mlens})}$ by keeping the eigenvectors in the matrix $U_{(\text{mlens})}$ corresponding to the 3 largest eigenvalues. In a similar manner, we obtain the $L \times 3$ matrix $\hat{U}_{(\text{light})}$ and the $V \times 1$ matrix $\hat{U}_{(\text{view})}$.

By using those matrices, we compute the core tensor $\hat{\mathcal{Z}}$ as

$$\hat{\mathcal{Z}} = \mathcal{I} \times_1 \hat{U}_{(\text{mlens})}^\top \times_2 \hat{U}_{(\text{light})}^\top \times_3 \hat{U}_{(\text{view})}^\top \tag{4}$$

from the tensor $\mathcal{I}$. Here, $\times_n$ stands for the n-th mode product. We compute the low-rank approximation of the tensor $\mathcal{I}$ as

$$\hat{\mathcal{I}} = \hat{\mathcal{Z}} \times_1 \hat{U}_{(\text{mlens})} \times_2 \hat{U}_{(\text{light})} \times_3 \hat{U}_{(\text{view})}, \tag{5}$$

[3] We assume that the light source vectors are not degenerate.

[4] The exception includes degenerate shapes such as a plane and a cylinder.

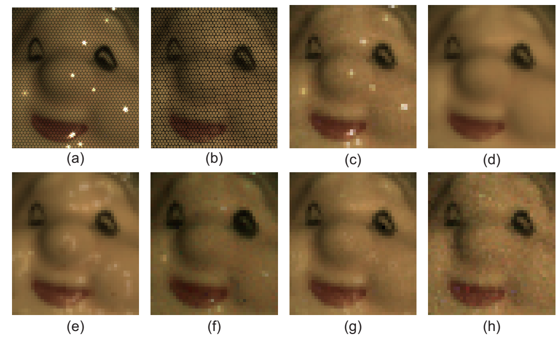

Fig. 4. The experimental results (ceramic dwarf): (a) an input raw image, (b) our result raw image, (c) an input image, (d) our result image, (e) our result image without outlier removal, and the result images of existing techniques: (f) varying viewing directions, (g) varying light source directions, and (h) polarizations.

and consider it as diffuse reflection components.

In general, actual images consist of not only diffuse reflection components but also outliers such as specular reflection components and shadows. Therefore, the low-rank matrices $\hat{U}_{(\text{mlens})}$, $\hat{U}_{(\text{light})}$, and $\hat{U}_{(\text{view})}$ computed via SVD are contaminated by those outliers. Accordingly, our proposed method makes use of SVD with missing data [8] instead of the conventional SVD for computing the low-rank matrices without contaminations due to those outliers[5].

4. EXPERIMENTS

4.1. Experimental Results

To confirm the effectiveness of our proposed method, we conducted a number of experiments using real images. We used 10 input images ($L = 10$) for each object taken by a static light field camera, LYTRO ILLUM, but under varying light source directions. The center coordinates of the microlenses are calibrated by using the existing library [15].

In Fig. 4, we show (a) an input raw image of a ceramic dwarf and (b) our result raw image [6] under a certain light source direction. We can see that the highlights due to specular reflection components are observed at saturated pixels and their neighbors in (a) the input image, but they are successfully removed from (b) our result image.

Fig. 4 (c) and (d) are the images seen from the central viewing direction synthesized from the raw images (a) and (b) respectively. Those images also show that our proposed method works well. In addition, (d) our result image is very

[5] Actually, we iteratively update $\hat{U}_{(\text{mlens})}$, $\hat{U}_{(\text{light})}$, $\hat{U}_{(\text{view})}$, and $\hat{\mathcal{Z}}$ by taking outliers into consideration.

[6] We handle the pixels within the radius of 6 pixels from the centers of microlenses, and black out other pixels.

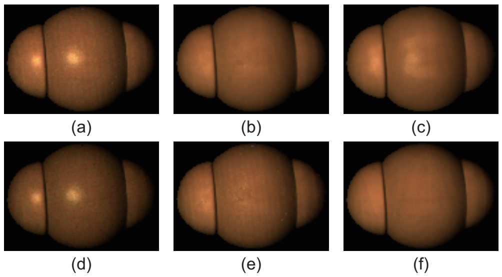

Fig. 5. The experimental results (wood bread): (a) an input image, (b) our result image, (c) our result image without outlier removal, and the result images of existing techniques: (d) varying viewing directions, (e) varying light source directions, and (f) polarizations.

smooth, because noises, *e.g.* zero-mean Gaussian noises, are canceled out by applying SVD to the pixel values observed at a point on an object surface from varying viewing directions. Those results qualitatively show that our method works well.

4.2. Comparison with Existing Techniques

We compared the performance of our proposed method with those of four closely related methods. In Fig. 4, we show the result images of (e) our method without outlier removal, (f) the method based on varying viewing directions, (g) the method based on varying light source directions, and (h) the polarization-based method.

First, Fig. 4 (e) shows that our method without outlier removal cannot remove specular reflection components, in particular, the result image is contaminated by the highlights observed under the all light source directions. This result shows the effectiveness of outlier removal, *i.e.* the use of SVD with missing data instead of the conventional SVD.

Second, in Fig. 4 (f), we can see that the method based on varying viewing directions cannot remove specular reflection components, when they are observed from the all viewing directions. In addition, the method is sensitive to noises and then the result image is rough, because it considers the darkest intensity as the diffuse reflection component.

Third, Fig. 4 (g) shows that the method based on varying light source directions can remove most specular reflection components, although some of them are still remaining. The method is rather sensitive to noises and then the result image is a little rough, because a set of images seen from a single viewing direction instead of multiple ones are used.

Fourth, in Fig. 4 (h), we can see that the polarization-based method [7] can remove specular reflection components well. Unfortunately, however, the method is sensitive to noises and then the result image is rough, because a set of images seen from a single viewing direction and under a single light source direction are used.

[7]The result image is darker than those of the other methods due to the use of polarization filters. Then, the result image is scaled so that its average pixel value is the same as that of our method.

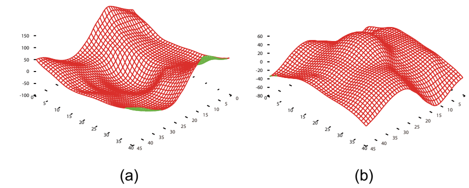

Fig. 6. The shapes of the dwarf (a) with and (b) without the GBR ambiguity.

Fig. 5 shows that the experimental results for a wood bread are similar to those for the ceramic dwarf. Here, the noises in (f) the polarization-based method are removed by averaging 64 images so that it is considered as the ground truth. The root-mean-square errors of pixel values in 8 bit images are 3.67, 7.49, 9.24, and 4.91 for (b) our method, (c) our method without outlier removal, (d) varying viewing directions, and (e) varying light source directions respectively. This result quantitatively shows that our proposed method performs better than those existing techniques.

4.3. Application to Uncalibrated Photometric Stereo

We can use the result of our proposed method for uncalibrated photometric stereo. Fig. 6 (a) shows the 3D shape of the ceramic dwarf recovered from the separated diffuse reflection components up to the GBR ambiguity by imposing the integrability constraint [16]. The separated specular reflection components are further useful for resolving the GBR ambiguity [5] as shown in Fig. 6 (b), where the convex/concave ambiguity is resolved manually. Note that the depth clue for resolving the GBR ambiguity is not available, when parallaxes are negligible.

5. CONCLUSION AND FUTURE WORK

In this paper, we addressed diffuse-specular separation for photometric stereo based on light fields. Specifically, we revealed the low-rank structure of the multi-view images under varying light source directions, and then formulated the diffuse-specular separation as a low-rank approximation of the 3rd order tensor. Through a number of experiments using real images, we confirmed that our proposed method, which integrates the complement clues based on varying light source directions and varying viewing directions, works better than the existing techniques. The extension to scenes with non-negligible parallaxes is one of the future directions of this study.

Acknowledgments This work was supported by JSPS KAKENHI Grant Numbers JP16H01676 and JP17H01766.

6. REFERENCES

[1] R. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[2] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *JOSA A*, vol. 11, no. 11, pp. 3079–3089, 1994.

[3] P. Belhumeur, D. Kriegman, and A. Yuille, "The bas-relief ambiguity," in *Proc. IEEE CVPR1997*, 1997, pp. 1060–1066.

[4] A. Georghiades, "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo," in *Proc. IEEE ICCV2003*, 2003, pp. 816–823.

[5] O. Drbohlav and M. Chantler, "Can two specular pixels calibrate photometric stereo?," in *Proc. IEEE ICCV2005*, 2005, pp. II–1850–1857.

[6] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Tech. Rep., Stanford CTSR 2005-02, 2005.

[7] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: tensorfaces," in *Proc. ECCV2002 (LNCS2350)*, 2002, pp. 447–460.

[8] H.-Y. Shum, K. Ikeuchi, and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling," *IEEE Trans. PAMI*, vol. 17, no. 9, pp. 854–867, 1995.

[9] K. Nishino, Z. Zhang, and K. Ikeuchi, "Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis," in *Proc. IEEE ICCV2001*, 2001, pp. 599–606.

[10] M. Tao, J.-C. Su, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras," *IEEE Trans. PAMI*, vol. 38, no. 6, pp. 1155–1169, 2016.

[11] A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image," *IJCV*, vol. 21, no. 1/2, pp. 99–122, 1997.

[12] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *Proc. ACCV2010 (LNCS6494)*, 2010, pp. 703–717.

[13] L. Wolff and T. Boult, "Constraining object features using a polarization reflectance model," *IEEE Trans. PAMI*, vol. 13, no. 7, pp. 635–657, 1991.

[14] S. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.

[15] Y. Bok, H.-G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," in *Proc. ECCV2014 (LNCS8694)*, 2014, pp. 47–61.

[16] A. Yuille and D. Snow, "Shape and albedo from multiple images using integrability," in *Proc. IEEE CVPR1997*, 1997, pp. 158–164.