Recursive photometric stereo when multiple shadows and highlights are present

Vasileios Argyriou and Maria Petrou
Imperial College London
v.argyriou@imperial.ac.uk, maria.petrou@imperial.ac.uk

Abstract

We present a recursive algorithm for 3D surface reconstruction based on Photometric Stereo in the presence of highlights, and self and cast shadows. We assume that the surface reflectance outside the highlights can be approximated by the Lambertian model. The algorithm works with as few as three light sources, and it can be generalised for N without any difficulties. Furthermore, this reconstruction method is able to identify areas where the majority of the lighting directions result in unreliable pixel intensities, providing the capability to adjust a reconstruction algorithm and improve its performance avoiding the unreliable sources. We report results for both artificial and real images and compare them with the results of other state of the art photometric stereo algorithms.

1. Introduction

Photometric stereo (PS) methods recover shape and albedo of an object using multiple images under varying illumination conditions in which the viewpoint is fixed. In order to recover gradient information of a surface patch, photometric stereo assumes that neither highlights nor shadows are present in the images used. In fact under orthographic projection, the only light source direction illuminating a scene which can result in an entirely shadowless image is the one aligned with the viewing direction. Therefore, shadowing is nearly unavoidable in any image. Most reconstruction algorithms either ignore the problematic pixels [2] or explicitly exclude data with shadows and highlights.

Traditional photometric stereo assumes that each image in the photometric set is acquired under a single point light source [17]. Since at least three images, corresponding to three different lighting directions, are needed to recover the surface normal, it is a common practice to use more than three illumination directions, to increase the possibility of all surface points being illuminated by at least three sources [5, 14, 11, 7, 16]. Thus, the challenges are first to identify which locations are problematic (i.e. correspond to highlights or shadows), and then to exploit the implied constraints to the fullest extent to solve the PS problem using for each pixel the three most appropriate images.

The motivation of this work is the problem arising when performing photometric stereo in the presence of highlights and shadows when the majority of the lighting directions result in unreliable pixel intensities. In [6, 15] only the problem of highlights is analysed while in [4] solutions to the problem of 3D surface reconstruction in the presence of shadows only are suggested. Furthermore, both [2] and [4] are not able to provide solution in the case when half (or more) of the light sources result in shadowed or highlighted pixels. Additionally, these methods require at least four light sources and cannot provide indications in the case when two or more of the sources produce highlights or shadows: they are designed to cope with only one problematic illumination direction per pixel. So, in this paper, we propose a recursive method for reconstructing 3D surfaces by using N-source photometric stereo (where $N \geq 3$) in the presence of highlights and shadows.

2. Notation and preliminaries

The basic assumptions in standard photometric stereo theory are a) the light source is a point source at infinity, b) the object is much smaller than the viewing distance, c) the object is directly illuminated without the presence of cast- or self-shadows and d) the surface of the object is Lambertian.

Based on [2] let us consider a Lambertian surface illuminated in turn by three illumination sources with directions $\vec{L}^1$, $\vec{L}^2$ and $\vec{L}^3$. In this case, the intensities of the obtained pixels can be expressed as

$$I^k = \rho(\vec{L}^k \cdot \vec{N}) \text{ where } k = 1, 2, 3 \quad (1)$$

where ρ is the surface albedo, i.e. the ratio of the reflected to incident radiation, and $\vec{N}$ is the surface normal, expressed as $(-p_n, -q_n, 1)$.

We stack the intensities and the illumination vectors I^k to form the pixel intensity vector $\vec{I} = (I^1, I^2, I^3)$ and the illumination matrix $[L] = (\vec{L}^1, \vec{L}^2, \vec{L}^3)^T$. Then equation (1) could be rewritten in matrix form as

$$\vec{I} = \rho[L]\vec{N} \quad (2)$$

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

If the three illumination vectors $\vec{L}^k$ are not lying in the same plane (non-coplanar), then the photometric illumination matrix $[L]$ is non-singular and can be inverted, giving

$$\vec{M} \equiv [L]^{-1}\vec{I} = \rho\vec{N} \tag{3}$$

where $\vec{M} \equiv (m_1, m_2, m_3)^T$. The surface gradient components could be obtained from $p_n = -m_1/m_3$ and $q_n = -m_2/m_3$ and the surface albedo is recovered by calculating the length of vector $\vec{M}$, $\rho = \sqrt{m_1^2 + m_2^2 + m_3^2}$.

We assume that shadows and highlights can be treated as disturbances of the Lambertian law. We also assume that both camera and light sources are far away from the surface, and the viewing and illumination directions are the same for every point on the surface. In case shadows and highlights are present, the surface recovery will be affected, with the body colour appearing different and the normal leaning more towards the light source which produced the highlight or away from the source which produced the shadow.

Based on the fact that any n vectors, where $n > 3$, in a 3D space are linearly dependent, there must be a linear equation expressing the relationship between the n illumination direction vectors [2]:

$$a_1\vec{L}^1 + a_2\vec{L}^2 + a_3\vec{L}^3 + \ldots + a_n\vec{L}^n = 0 \tag{4}$$

If we multiply both sides of this equation with the surface albedo ρ and take the dot product of both sides with the normal of the surface patch, we obtain:

$$a_1I^1 + a_2I^2 + a_3I^3 + \ldots + a_nI^n = 0 \tag{5}$$

In other words, linear dependence of the illumination vectors leads to the same linear equation for the corresponding pixel intensities, if the Lambertian assumption holds. We may rewrite equation (5) in vector form:

$$\vec{a} \cdot \vec{I} = 0 \tag{6}$$

where $\vec{a} = (a_1, a_2, \ldots, a_n)^T$. Vector $\vec{a}$ can be directly calculated from the illumination matrix as an eigenvector of matrix $[LL^T]$ which corresponds to the zero eigenvalue [2]. Unreliable pixels, containing either shadows or highlights do not satisfy equation (5) and therefore this method allows one to rule out the majority of the pixels that are not purely Lambertian.

3. Recursive Photometric Stereo

Let us assume that photometric stereo has been performed using all the available images I^k based on equation (1) and that the unreliable pixels have been determined according to equation (5). Prior to integration of the obtained surface normals, the information provided by the non-Lambertian pixel map is used to correct the erroneous normals. Interpolation or robust fitting [13], is applied to replace the unreliable normals, preserving the altitude variation inside these areas during the integration stage. The result of this step is shown in figure 1, and prevents erroneous estimation of shadows due to inaccurate surface reconstruction.

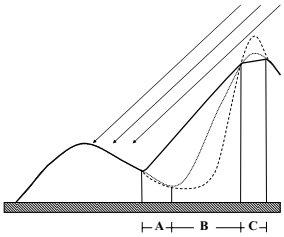


Figure 1. Original (dotted line), erroneously reconstructed (dashed line) and interpolated (solid line) surfaces, where A indicates cast shadows, B self shadows and C highlights.

In the next stage we consider the problem of reconstructing the surface from the modified normals. In this case, to compute the shape of the surface, we need to obtain the depth map. This suggests representing the surface as $(x, y, f(x, y))$, so the normal as a function of (x, y) is

$$\vec{N}(x,y) = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}} \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right)^T \tag{7}$$

To recover the depth map, we need to determine $f(x, y)$ from measured values of the unit normal. There are a number of ways in which a surface may be recovered from a field of surface normals [3, 8, 9, 12, 18, 10]. There are local and global methods based on trigonometry and the minimisation of error functionals, respectively and the most suitable could be selected for this part of the process.

Assume that the measured value of the unit normal at some point (x, y) is $(a(x, y), b(x, y), c(x, y))$. Then

$$\frac{\partial f}{\partial x} = \frac{a(x,y)}{c(x,y)} \quad \frac{\partial f}{\partial y} = \frac{b(x,y)}{c(x,y)} \tag{8}$$

At this stage we may perform another check on our data set. Because

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \tag{9}$$

we expect

$$A(x,y) \equiv \frac{\partial\left(\frac{a(x,y)}{c(x,y)}\right)}{\partial y} - \frac{\partial\left(\frac{b(x,y)}{c(x,y)}\right)}{\partial x} \tag{10}$$

to be small (close to zero) at each point.

Assuming that the partial derivatives pass the above sanity test, we can reconstruct the surface up to some constant error in depth. The partial derivatives give the change in surface

height with a small step in either the x or the y direction. This means that we can get the surface by summing these changes in height along some path. In particular, we have

$$f(x,y) = \oint_C \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot \vec{dl} + c \qquad (11)$$

where C is a curve starting at some fixed point and ending at (x,y), $\vec{dl}$ is the infinitesimal element along the curve and c is a constant of integration, which represents the unknown height of the surface at the starting point. The recovered surface should not depend on the choice of the path followed for performing the integral.

For example, we may reconstruct the surface at (u,v) by starting at $(0,0)$, summing the y-derivatives along the line $x = 0$ to point $(0,v)$, and then summing the x-derivatives along the line $y = v$ to point (u,v):

$$f(u,v) = \int_0^v \frac{\partial f}{\partial y}(0,y)dy + \int_0^u \frac{\partial f}{\partial x}(x,v)dx + c \qquad (12)$$

Since any other set of paths would work as well, it is best to use many different paths and average the results so that we reduce the error in the estimates of the derivatives. This integration algorithm did not yield satisfactory results: while it seemed to do an average job on sharp edges, smooth surfaces were noisy. In order to improve further the reconstruction we combine it with a multigrid 2D integration algorithm, which iteratively solves a global minimization problem, and is less sensitive to the propagation of local errors.

Now, the challenge is to identify which pixels in the image correspond to surface patches that are in shadow with respect to each light source. A problem analogous to shadow detection is determining light source visibility at every pixel. In fact, the two problems are equivalent for a scene illuminated by a single point source. Therefore, an approach used in computer graphics was selected to identify the shadowed (non-visible) areas of the reconstructed surface. According to this method, a Z-buffer is created and during rasterisation the depth/Z value of each pixel is checked against an existing depth value. If the current pixel is behind the pixel in the Z-buffer, the pixel is rejected, otherwise its depth value replaces the one in the Z-buffer. The above procedure is repeated for all illumination directions and a shadow-map is obtained for each light source. Let S_i be the shadow-map obtained by the i-th light source. Then the highlight-map, in set theory notation, is given by

$$H = E - (S_1 \vee S_2 \vee \ldots \vee S_n) \qquad (13)$$

where E is the non-Lambertian pixel map.

In order to identify which of the n light sources produce highlights at each pixel location, the sources producing shadows are rejected and the remaining n' sources are classified based on their intensity values. A threshold is used to reject the highlighting sources and the remaining ones are used to estimate the surface normal at that pixel location.

The whole procedure is repeated using the new normal vectors at the highlighted areas. Again new shadow maps are estimated, since the new reconstructed surface may produce different self- or cast-shadows. For each pixel location the intensities produced by the remaining reliable illumination sources are used to estimate the surface normals. This algorithm (figure 2) utilises only reliable pixels to apply photometric stereo and can apply different strategies in case fewer than three lights are reliable (e.g. taking into consideration neighbouring non-affected by shadows or highlights pixels).

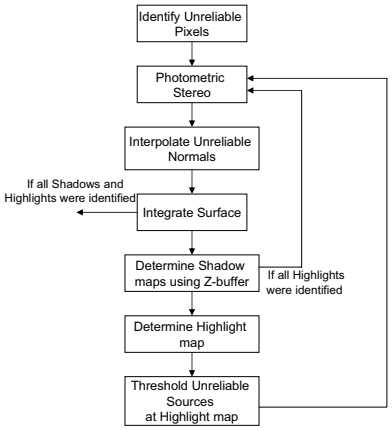


Figure 2. Flow chart of the proposed recursive photometric stereo algorithm.

4. Experiments and Results

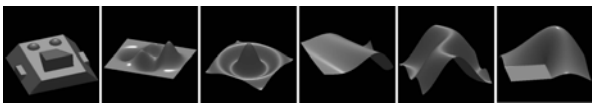
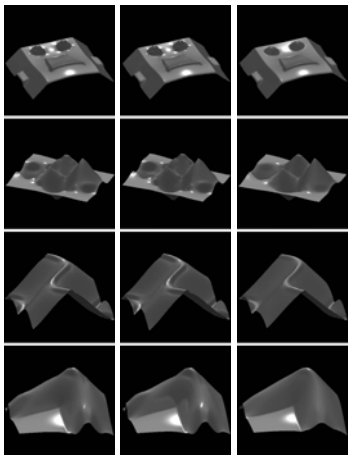


Figure 3. Simulated data used to evaluate the photometric stereo algorithms. All surfaces are of size 128×128.

Table 1. The mean angular error (MAE) of the surface normals for the simulated data. In bold the best result for each surface.

MAE	Artf1	Artf2	Artf3	Artf4	Artf5	Artf6
C & J	10.6	14.0	20.5	15.8	23.5	16.7
B & P	10.4	13.6	19.5	14.5	22.8	15.7
Rec. PS	**10.1**	**10.6**	**15.6**	**12.7**	**16.6**	**13.1**

In case of artificial data, ground truth is known a priori. Therefore, in order to compare the performance of the proposed recursive algorithm with the methods proposed by



(a) (b) (c)

Figure 4. The reconstructed surfaces for the first artificial surface using (a) Coleman's and Jain's, (b) Barsky's and Petrou's and (c) the proposed recursive method. Note the multiple highlights in (a) and (b) which indicate that the flat surfaces were wrongly reconstructed especially at the regions where curvature changes significantly or areas containing peaks.

Figure 5. Input images, one for each illumination direction.

(a) (b)

Figure 6. (a) Unreliable pixel map in grey scale, according to equation (5). The brighter the pixel, the higher the error of the estimated surface normal. (b) Unreliable binary pixel map obtained using a threshold, $T = 2\sigma$ on the intensity of the grey scale map.

Table 2. Average intensity of the estimated highlights for all the tested faces. In bold the best result for each surface.

	bej	bln	fav	mut	pet	rob	srb
C & J	189	79	84	56	110	66	96
B & P	**238**	193	160	117	152	134	198
Rec. PS	225	**198**	**165**	**136**	**154**	**150**	**214**

Coleman and Jain [6], and Barsky and Petrou [2], the MAE measure suggested by Barron, *et al* [1] is used:

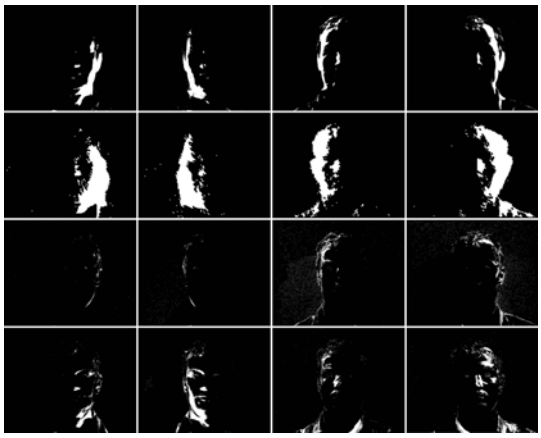
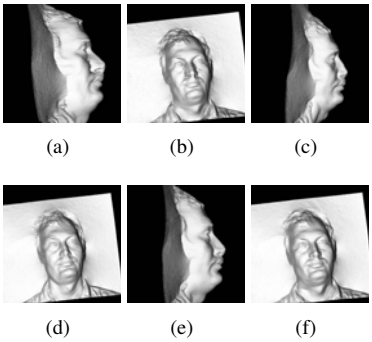


Figure 7. From top to bottom: (1^{st} row) Areas containing self-shadows for each light source obtained using the normal vectors. (2^{nd} row) Areas containing cast-shadows for each light source obtained using ray tracing. (3^{rd} row) Shadows obtained using the criterion proposed by Coleman and Jain. (4^{th} row) Shadows obtained using the criterion proposed by Barsky and Petrou.



(a) (b) (c)

(d) (e) (f)

Figure 8. The reconstructed surfaces for the 'bln' image set using (a)-(b) Coleman's and Jain's, (c)-(d) Barsky's and Petrou's (e)-(f) the proposed recursive method.

$$\Psi_{AE} = \cos^{-1}\left[\frac{1}{\sqrt{1+x_r^2+y_r^2+z_r^2}}\begin{pmatrix} x_r \\ y_r \\ z_r \\ 1 \end{pmatrix}^T \frac{1}{\sqrt{1+x_e^2+y_e^2+z_e^2}}\begin{pmatrix} x_e \\ y_e \\ z_e \\ 1 \end{pmatrix}\right] \quad (14)$$

where $(x_r, y_r, z_r)^T$ and $(x_e, y_e, z_e)^T$ are the real and the estimated surface normals, respectively. To form an estimate for the whole image, the mean value can be calculated using

$$\Psi_{MAE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\Psi_{AE}(i,j) \quad (15)$$

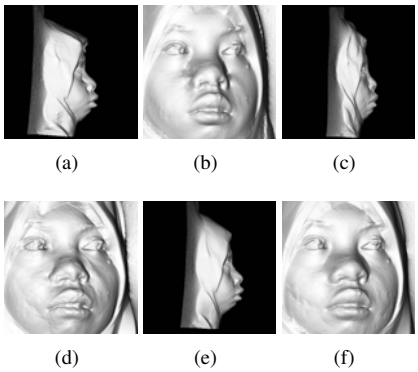


Figure 9. The reconstructed surfaces for the 'bej' image set using (a)-(b) Coleman's and Jain's, (c)-(d) Barsky's and Petrou's (e)-(f) the proposed recursive method.

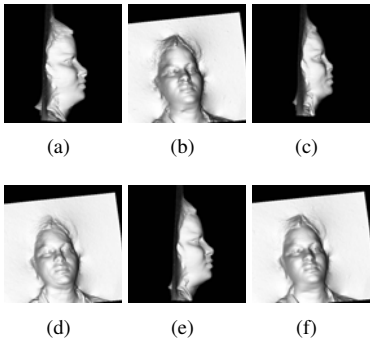


Figure 10. The reconstructed surfaces for the 'srb' image set using (a)-(b) Coleman's and Jain's, (c)-(d) Barsky's and Petrou's (e)-(f) the proposed recursive method.

For the case of real data there is no ground truth, so we cannot evaluate the performance of the algorithm in the same way. Instead, the average intensity of the estimated highlights was used as an indicator of the performance.

4.1. Experiments with simulated data

Experiments were performed with simulated data, where six surfaces were artificially created (see figure 3). The surfaces were selected on the basis of containing highlights, self- and cast-shadows, in order to provide a comprehensive evaluation data set. The results for the simulated surfaces are shown in Table 1. The proposed method outperforms both other methods resulting in the lowest error. Figure 4 shows the reconstructed surfaces, where it can be observed that the proposed algorithm provides more accurate and smoother results especially at the areas with self- and cast-shadows.

4.2. Experiments with real data

We applied the aforementioned algorithm in the reconstruction of 7 human faces using photometric data captured

Figure 11. The reconstructed surfaces for some image sets using from top to bottom Coleman's and Jain's, Barsky's and Petrou's and the proposed recursive method.

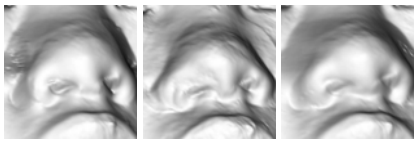


Figure 12. The reconstructed zoomed surfaces for the 'bej' image set using (a) Coleman's and Jain's, (b) Barsky's and Petrou's (c) the proposed recursive method.

Figure 13. Areas where less than three reliable pixels are available for the first image set.

with four illumination directions. Figure 5 presents the input images for one of the faces, one for each illumination direction. The lights were placed on the vertices of a square $2m \times 2m$ and the distance between the camera and the observed person was $2m$. In such an arrangement the difference of the tilt angles between neighbouring illumination directions is 90^o degrees and the slant angle is constant equal to 35^o degrees for all light sources. The person is assumed to be still during the acquisition stage, eliminating the registration problem.

In this case, ground truth is not available and therefore the average intensity of the estimated highlights was used as an indicator of the performance, (see Table 2). The higher values indicate a more accurate approximation of highlights, since in general highlights tend to be bright pixels. But this

is not always true, because there might be highlights that are not brighter than some non-highlighted areas. Figure 6 shows the unreliable pixels obtained according to equation (5) while figures 7(1^{st} row) and 7(2^{nd} row) present the self- and cast-shadows obtained by the proposed algorithm, for the first of the faces. The shadow maps obtained by the other two methods are shown in figures 7(3^{rd} row) and 7(4^{th} row). The estimated shadow maps by Coleman and Jain are not accurate at the regions under the jaw and the ears. The results obtained by Barsky's and Petrou's are more precise but not solid as in the case of the proposed method, since the latter is not pixel based and takes into account the reconstructed 3D shape.

In figures 8– 11 results of the reconstructed faces obtained from the three algorithms are shown. Observing the results it can be inferred that the proposed algorithm outperforms the others especially at the areas containing cast-shadows (e.g. nose, shoulders, under the jaw). Figure 12 shows one surface zoomed, where the erroneous reconstructed areas close to the lips and the left side of the nose can be observed for the methods proposed by Coleman and Jain, and Barsky and Petrou. The main advantage of the proposed algorithm is observed at the shadowed areas, since the other methods provide sparse shadow maps resulting in inaccurate reconstruction (e.g. cavities), at smooth and flat regions. In order to reconstruct the surfaces at the areas where fewer than three reliable pixels are available, (see figure 13), we use the brightest pixels to complement the available reliable pixel values. The algorithm suggested by Coleman and Jain has lower performance in areas with shadows, while the method proposed by Barsky and Petrou is less accurate only when cast-shadows are present.

5. Conclusions

In this paper, we propose a recursive algorithm for 3D image reconstruction based on the well-known photometric stereo method, which is dealing with highlights and both self- and cast-shadows allowing one to obtain more reliable estimates of surface parameters. The proposed method does not have any restrictions regarding the maximum number of light sources, and can identify areas where the majority of the lighting directions result in unreliable pixel intensities. Also, indication is provided when the number of the reliable pixel intensities is less than three, giving the possibility to apply different techniques at these areas (e.g. shape from shading).

Experiments with both artificial and real images were performed in order to evaluate the performance of the proposed algorithm. The mean angular error was used in the artificial scenario. From the real images it can observed that the method works well in areas with self- and cast-shadows.

References

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994. 4

[2] S. Barsky and M. Petrou. The 4-source photometric stereo technique for 3-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. Patt. Anal. Machine Intell.*, 25(10):1239–1252, 2003. 1, 2, 4

[3] A. Bors, E. Hancock, and R. Wilson. Terrain analysis using radar shape-from-shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):974–992, August 2003. 2

[4] M. Chandraker, S. Agarwal, and D. Kriegman. Shadowcuts: Photometric stereo with shadows. *CVPR*, June 2007. 1

[5] J. Chang, K. Lee, and S. Lee. Multiview normal field integration using level set methods. *Computer Vision and Pattern Recognition*, 2007. 1

[6] E. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Vision, Graphics and Image Processing*, 18:309–328, 1982. 1, 4

[7] G. Finlayson, M. Drew, and C. Lu. Intrinsic images by entropy minimisation. *In Proc. ECCV*, pages 582–595, 2004. 1

[8] R. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading. *IEEE Trans. Pattern Anal. Machine Intell.*, 10(4):439–451, July 1988. 2

[9] B. Horn. Height and gradient from shading. *Int. J. Comput. Vision*, 5(2):37–75, 1989. 2

[10] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on PAMI*, 29(4):640–649, April 2007. 2

[11] P. Kovesi. Shapelets correlated with surface normals produce surfaces. *In Proc. Int. Conf. Computer Vision*, pages 994–1001, 2005. 1

[12] A. Robles-Kelly and E. Hancock. A graph-spectral approach to shapefrom-shading. *IEEE Trans. Image Process.*, 13(7):912–926, 2004. 2

[13] W. Smith and E. Hancock. Estimating cast shadows using sfs and class-based surface completion. *Proceedings of the 18th ICPR*, 4:86–90, 2006. 2

[14] W. Smith and E. Hancock. Recovering facial shape using a statistical model of surface normal direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1914–1930, December 2006. 1

[15] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):449–454, April 1996. 1

[16] J. Sun, M. Smith, L. Smith, S. Midha, and J. Bamber. Object surface recovery using a multi-light photometric stereo technique for non-lambertian surfaces subject to shadows and specularities. *Image and Vision Computing*, 25(7):1050–1057, July 2007. 1

[17] R. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intesit. *SPIE*, 155:136–143, 1978. 1

[18] Z. Wu and L. Li. A line-integration based method for depth recovery from surface normals. *Comput. Vision Graph. Image Process.*, 43:53–66, 1988. 2