

A Flexible Method for Gauge-Based Multiple Image Photometric Stereo

Suppressed to avoid author identification

Abstract

We describe a method for determining the surface normals (slopes) of a three-dimensional scene, from three or more digital photographs taken from the same viewpoint under different lighting conditions. We assume that the scene contains a light gauge, a test object with known shape and uniform color. The slope at a point P of the image is found by locating a point q of the gauge that displays similar reaction to the incident light in all images. The method can be used also with simulated images of gauges (virtual gauges) computed from a description of the relevant light fields. In particular, we show how to fit a simple light field model to a photo of an actual light gauge of known geometry. This procedure can be used to extract the significant photometric information from the gauge's photos, while discarding small-scale noise that arises from dents or stains of the real object.

1. Introduction

1.1. Variable-lighting photometric stereo

We describe a technique for *variable-lighting photometric stereo* (VLPS), whose aim is to recover the surface normals (slopes) of a three-dimensional scene from a collection of 2D images taken with different lighting conditions but with the same pose and viewpoint [12].

The main features of our method are (1) the use of light gauges to assess the illumination field; (2) a procedure for fitting a simple illumination model to images of real light gauges; and (3) efficient inversion of the shading function by a two-dimensional bucket grid.

This method can be applied to a variety of 3D reconstruction problems in medicine, geology, engineering, virtual reality, archeology, and many other areas. The normal computation procedure is fairly robust and can be applied to objects with Lambertian surfaces of arbitrary color and texture. As a byproduct, the normal computation procedure also yields the “true color” (per-channel reflectance coefficient) of the scene at every point of the image. The

method can be applied also to surfaces with arbitrary (non-Lambertian) finish, provided that the relevant object has the same uniform color and finish as the light gauge — that is, the same bi-directional reflectance function (BRDF) everywhere. The method works even in the presence of proper shadows (parts of the target surface which are turned away from the light) and with extended sources and other complex light fields, as long as the light field is uniform over the scene. (Moderately non-uniform light fields could be handled by including multiple light gauges in the scene and interpolating their information.)

On the other hand, the normal computation method cannot be directly applied to scenes with projected shadows; but can detect them, and can still be used at every pixel which is fully illuminated in at least three photos. The method does not handle complex optical paths, such as light incident at P that has been scattered, reflected, or refracted by other parts of the scene.

Figure 1 (a,b,c) show a typical set of input images. The round objects around the corners are light gauges.

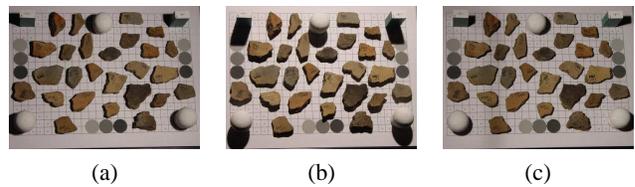


Figure 1. Three images of the same scene under different lighting conditions.

1.2. Related work

The principles of photometric stereo must have been known at least since the establishment of photography and the development of photometry, in the early 20th century. Early articles assumed very simple settings: a smooth object with uniform color and Lambertian finish, illuminated by a single light source, distant and point-like (i.e. by a uniform and unidirectional light field). Their emphasis was on methods that could extract depth information from a single monochromatic image [7, 8, 15]

The extraction of slope information from multiple images under different light fields, using light gauges, was pioneered by Woodham in 1980 [14] and has been extensively researched since then.

Woodham [16] also introduced the technique of the *look-up table* to search the best match between the brightness vector of the sphere (gauge object) and the brightness vector of the scene object. The table size of look-up table is exponential in number of images (the total size of the look-up table is 2^{bd}), which depends on the dimension d and the number of bits b into which brightness is quantized). When the number of registered images increase, we have many brightness vector of the gauge object and the dimension look-up table has a prohibitive cost.

Osamu Ikeda [9] extended the three-image method to color images, by using the HSV color space. Shadows and highlights are a major nuisance in photometric stereo, since pixel values carry no shape information wherever those features are present. Barsky and Petrou [1] described a method to detect such regions in a set of four or more images.

In 1999, Gionis, Indyk and Montwani [4] introduced a *locality sensitive hashing* (LSH) to search similar images in large databases. The basic idea of LSH is to hash high dimensional points so as to ensure that the probability of collision is much higher for points that are close to each other than for other are far apart.

In 2005, Hertzmann and Seitz [5, 6] presented a stereo photometric approach, with the use of gauge objects, which enables reconstructing surfaces with arbitrary BRDF (the method assumes orthography, distant lighting, no cast shadow, no inter-reflection, no subsurface scattering, and no transparency).

Also in 2005, Zhong and Little [17] developed a method that extend a work of Hertzmann and Seitz [5] (that simplified the traditional shape from shading experiment avoiding calibration). To speed up the search, the authors adapted the LSH to search for matches in high dimension brightness vector space. The computational cost is $O(N)$, i.e. the LSH hashing is constant per pixel, but there is a cost involved in setting up the hash table, but the authors envision inspection applications where the cost can be absorbed at the beginning an only inexpensive operations are done per object.

Recently, Basri, Jacobs and Kemelmacher [2] presented a model-fitting method for photometric stereo under generally lighting condition. The method analyzes the gauge images to compute approximate models of the shading functions using spherical harmonics of first or second order, and then uses these models to extract the slope information from the scene images.

2. Principles and notation

2.1. The scene images

The main input data for our method is a list of $m \geq 3$ digital photos S_1, \dots, S_m of some optically passive scene, the *scene images*, satisfying certain conditions. First, all m images must be taken with different lighting conditions, but with the same pose and viewpoint; and must have been geometrically corrected and aligned, so that any point P of the scene projects to the same pixel position p on all images. The photos must be effectively monochromatic, with a linear color scale (i.e. pixel values must be proportional to physical light intensity). To simplify the exposition, we also assume a simple orthogonal projection of the scene onto the images, without perspective distortions.

Under these assumptions, we may assume that the relative intensity $S_i[p]$ at a point p of photo S_i depends only on three attributes, associated to the portion P of the scene's surface that projects onto p :

- the surface's *intrinsic optical properties* $\beta[p]$ (that depend on its material and finish, such as emissivity, reflectance, polish, etc);
- its *normal direction* $\hat{s}[p]$ (the unit vector perpendicular to the surface and directed away from the underlying object); and
- the *incident light field* Φ_i (the intensity of light flowing in each direction towards the surface) used in each photo.

The intrinsic optical properties $\beta[p]$ can be modeled as a *bidirectional reflectance function*, or BRDF — a function $\beta[p](\hat{n}, \hat{u}, \hat{v})$ that gives the apparent brightness of the surface when oriented with normal \hat{n} , viewed from the direction v , and illuminated with unidirectional light of unit intensity flowing in the direction u . (Note that we consider the geometric light spread factor $\max\{0, \hat{u} \cdot \hat{n}\}$ included in the BRDF β).

The fundamental principle of *multiple image photometric stereo* is that we can in principle recover the surface normal $\hat{s}[p]$ at each image point p from analysis of the m pixel intensities $S_i[p]$, provided that we have sufficient knowledge of the BRDF β and the light fields Φ_i .

2.2. The gauge images

In gauge-based VLPS, the relevant information about $\beta[p]$ and the flows Φ_i are obtained indirectly from a set G_1, \dots, G_m of *gauge images* — photos of a reference *light gauge object* of known shape. The gauge object must have with the same BRDF as the scene's surface at p , except for a constant factor; and each photo G_i is taken with the same

camera position and under the same lighting conditions as the corresponding scene photo S_i .

It is often convenient to include the gauge object as part of the scene, as in figure 1. In that case, the gauge photos G_i are actually contained in the scene photos S_i .

Each scene or gauge photo is assumed to be a real-valued function of two real-valued *image coordinates*, defined in some *image domain* — a subset of \mathbb{R}^2 . The domains of S_i and G_i are denoted \mathcal{S} and \mathcal{G} , respectively.

2.3. Image formation model

Under the above assumptions, the BRDF $\beta[p]$ at every point p of the scene images can be factored into the product of some unknown scalar $S^*[p]$ — the scene’s *intrinsic lightness* at p — and a single unknown BRDF $\bar{\beta}$. Likewise, the gauge’s BRDF at a point q of the gauge images is assumed to be $G^*[q]\bar{\beta}$, for a *known* intrinsic lights G^* .

It follows from these assumptions that each scene or gauge photo can be factored into the product of two images: the *intrinsic lightness map*, S^* or G^* , and a factor that depends only on the lighting conditions and the local orientation of the surface. Specifically,

$$\begin{aligned} S_i[p] &= S^*[p] L_i(\hat{s}[p]) \\ G_i[q] &= G^*[q] L_i(\hat{g}[q]) \end{aligned} \quad (1)$$

for some set of *shading functions* L_1, \dots, L_m . Here \hat{s} and \hat{g} are the *normal maps* of the scene and gauge, respectively. That is, $\hat{s}[p]$ is the the unit-length vector perpendicular to the visible portion P of the scene’s surface which projects to point p of the image; and $\hat{g}[q]$ is similarly defined for the gauge object and the gauge images.

Note that, in this model, the intrinsic color maps G^* and S^* are distinct but the same for all i , whereas the shading functions L_i , that map surface normals to relative apparent intensities, are different for each i but are shared by the scene and gauge objects. The function L_i depends only on the light field Φ_i and the common BRDF $\bar{\beta}$, by the formula

$$L_i(\hat{n}) = \int_{\mathbb{S}^2} \Phi_i(\hat{u}) \bar{\beta}(\hat{n}, \hat{u}, \hat{v}) d\hat{u} \quad (2)$$

where \hat{v} is the (fixed) viewing direction in all photos.

This model obviously holds for a scene consisting of purely diffusive (Lambertian) surfaces of arbitrary and varying lightness. In that case, $S^*[p]$ is the surface’s reflectance at pixel p ; and $\bar{\beta}(\hat{n}, \hat{u}, \hat{v})$ reduces to the geometric spread factor, $\max\{0, \langle \hat{u} | \hat{n} \rangle\}$.

However, this model also fits some non-Lambertian surfaces, under more restrictive conditions. It applies, for example, when the scene and gauge are made of the same material, with the same intrinsic color and finish (e.g. molded pieces of the same plastic material), possibly with an obscuring layer of black “dust” over it. In that case, the S^* or G^* factor would be the transmittance of that dust layer.

2.4. Input gauge model

Besides the scene and gauge images, our method also needs to be given the gauge’s normal map \hat{g} and intrinsic lightness map G^* . If the gauge has a simple geometric shape and uniform color, that information can be computed analytically from a few parameters. Alternatively, these can be given as of two additional images, respectively one channel (G^*) and three channels ($\hat{g}[q]$).

2.5. The signature matching principle

If the model (1) holds, then, for any two pixels p, q such that $\hat{s}[p] = \hat{g}[q]$, we must have $L_i(\hat{s}[p]) = L_i(\hat{g}[q])$ for all i , and thus

$$\frac{S_i[p]}{G_i[q]} = \frac{S_j[p]}{G_j[q]} = \frac{S^*[p]}{G^*[q]} \quad (3)$$

for all i and j . That is, the vectors

$$\begin{aligned} \mathbf{S}[p] &= (S_1[p], \dots, S_m[p]) = S^*[p](L_1[p], \dots, L_m[p]) \\ &\text{and} \\ \mathbf{G}[q] &= (G_1[q], \dots, G_m[q]) = G^*[q](L_1[q], \dots, L_m[q]) \end{aligned} \quad (4)$$

must be multiples of each other, by the ratio $S^*[p]/G^*[q]$. Therefore, we could in principle determine the normal $\hat{s}[p]$ at a point of the image, by looking for a point q in \mathcal{G} such that $\mathbf{G}[q]$ is a multiple of the vector $\mathbf{S}[p]$. Assuming that neither vector is zero, this is equivalent to matching the *lighting signatures* $\mathbf{s}[p] = \mathbf{g}[q]$, where

$$\mathbf{s}[p] = \frac{\mathbf{S}[p]}{|\mathbf{S}[p]|} \quad \mathbf{g}[q] = \frac{\mathbf{G}[q]}{|\mathbf{G}[q]|} \quad (5)$$

and $|\cdot|$ is any norm of \mathbb{R}^m , e.g. the Euclidean norm $|\mathbf{X}| = (\sum_{i=1}^m X_i^2)^{1/2}$. That is, we locate a point q on the gauge images that reacts in the same way as point p of the scene to changes in the light field, except for a fixed constant factor $\alpha[p]$.

Having located the matching gauge point q , we can recover the normal map and intrinsic lightness map of the scene at p by

$$\begin{aligned} \hat{s}[p] &= \hat{g}[q] \\ S^*[p] &= \frac{|\mathbf{S}[p]|}{|\mathbf{G}[q]|} G^*[q] \end{aligned} \quad (6)$$

2.6. Feasibility conditions

Formally, the result of the procedure consists of two functions $\hat{\sigma}$ and σ_* , from the scene image domain \mathcal{S} to the set of all normals (that is, the unit sphere \mathbb{S}^2) and to the real, respectively, where

$$\hat{\sigma}[p] = \hat{g}(\mathbf{g}^{-1}(\mathbf{s}[p])) \quad \text{and} \quad \sigma_*[p] = G^*(\mathbf{g}^{-1}(\mathbf{s}[p])) \quad (7)$$

for all $p \in \mathcal{S}$.

Thus, in order for the method to be feasible, the following conditions must be satisfied:

- (C1) the range of \mathbf{g} must include the set of all vectors $\mathbf{s}[p]$ that occur in the scene images; and
- (C2) the function \mathbf{g} must be invertible.

Condition (C1) means that, for any visible point P of the scene, the gauge must have at least one visible point Q with the same normal as P . Condition (C2) is more complicated, since it depends on the BRDF β . Basically, it says that any change in the surface’s normal \hat{n} must imply a change in the normalized gauge image intensities $\mathbf{g}[p]$ — that is, a change in the vector $\mathbf{G}[p]$ which is not simply a scaling. By counting degrees of freedom, it is obvious that for every normal direction \hat{n} there must be at least three lighting setups Φ_i which illuminate surfaces with that orientation. This condition is usually sufficient for Lambertian surfaces, if those three light fields are dominated by compact light sources in well-separated and non-coplanar directions.

2.7. The light table

Computationally, the hardest part of the method is inverting the \mathbf{g} function, that is, finding the point $q \in \mathcal{G}$ such that $\mathbf{g}[q] = \mathbf{s}[p]$.

The light gauge normal map \hat{g} and the gauge photos G_i can be preprocessed to produce a *light signature table*, a set T of triplets $(\hat{g}_k, \mathbf{g}_k, \alpha_k)$, each consisting of the light signature \mathbf{g}_k , the surface normal \hat{g}_k , and the normalizing factor α_k for some pixel q_k in the gauge photos. Note that the pixel q_k itself is irrelevant and need not be stored. Thus the normal computation reduces to finding the normal \hat{g}_k in this table whose associated light signature \mathbf{g}_k best matches the scene’s signature $\mathbf{s}[p]$. However, due to the somewhat irregular nature of the function $\hat{g} \circ \mathbf{g}^{-1}$, the computation is still quite expensive, since the table T must have thousands of entries in order to produce a reasonably accurate normal map.

To speed up this search, we use a *bucket grid*, a data structure that has been quite effective in many geometric search problems [11, 10]. We use an original formulation that is both simpler and more efficient than those reported in the literature. The key idea is that the set of all normalized signatures \mathbf{g}_k spans a two-dimensional manifold in m -dimensional space, which can be projected onto a plane with moderate geometric distortion. Therefore, the table can be efficiently hashed into a two-dimensional bucket grid structure, and the search can be confined to a few buckets. The details are given in a companion paper [13].

3. Light gauge processing

3.1. Spherical light Gauges

As explained in section 2.4, the normal computation method requires knowledge of the light gauge’s intrinsic lightness map G^* and normal map \hat{g} , precisely matching the photos G_i . For Lambertian scenes, we use smooth Lambertian spheres of uniform color. The spheres should be small enough to be placed near the target objects without significantly disturbing the light field. The spherical shape provides a fairly uniform sampling of the light field, and simplifies considerably the computation of the gauge’s surface normals.

3.2. Virtual light gauges

In practice, the light gauges deviate from the ideal spherical shape, due to manufacturing defects, dents, scratches, etc.. Even a small defect, covering a couple of pixels, may introduce large errors in the signature-to-normal table. A small dent or bump may create an arbitrarily large error between the actual surface normal and the given normal map \hat{g} . A small stain, especially one that changes the surface finish, will change the normalization α , thus introducing a bogus complementary stain at every part of the scene that has that normal orientation. Moreover, if the stain has a different BRDF than the scene, it may also change the normalized signature \mathbf{g} by a large amount. Either kind of defect will introduce grossly incorrect normal-signature pairs $(\hat{g}_k, \mathbf{g}_k)$ into the table T , which may produce large errors over large areas of the output normal map \hat{s} .

3.3. Modeling the shading function

This difficulty can be overcome by observing that the BRDFs of typical materials, especially the Lambertian ones, broadly spread the light flowing in any direction u over the hemisphere of all directions that make an acute angle with the normal \hat{s} . It follows that the shading function L_i of Lambertian and near-lambertian surfaces is a fairly smooth function, even when the light flow is concentrated in a few directions.

Thus, the solution to the problem of imperfect gauges is to filter the “noisy” shading function L_i that is obtained by pairing the presumed normal map $\hat{g}[q]$ of the gauge with the noisy photo $G_i[q]$. The smoothing generally removes spurious entries due to small defects, leaving only the useful part of the shading data.

In any case, once we have determined the smoothed shading functions L_i , we can combine them with the normal map \hat{g} of any suitable *virtual gauge*, and produce the artificial gauge photos by the composition $G_i \leftarrow L_i \circ \hat{g}$

for input to the matching procedure. The obvious choice for a virtual gauge is a hemispherical surface, since it evenly samples all normal directions that are visible from the camera. In fact, the input images \hat{g} and G_1, \dots, G_m are simply a visually convenient way of entering the shading functions L_1, \dots, L_m to the method.

3.4. Fitting a simple light model

When computing the smoothed shading functions L_i , we may further improve the result by incorporating any information we have about the light flow Φ_i used in the scene S_i . For example, if we know that Φ_i was dominated by direct light from a single point-like source at an unknown location, we can restrict the smoothed shading function L_i to a function of that form.

In our tests, we assumed a slightly more complex model where each light field Φ_i was dominated by an isotropic and uniform ambient field of unknown intensity A_i , and a distant source of unknown intensity and direction \hat{w}_i , which, when seen from the scene, was contained in a cone with known angular radius ρ . Assuming a Lambertian BRDF, the shading function $L_i(\hat{n})$ generated by such a source coincides with that of a slightly dimmer point source in direction \hat{w}_i , for those surface orientations that are fully illuminated by the extended source — that is, whenever the angle between \hat{n} and \hat{w}_i is less than $\pi/2 - \rho$.

For those normal directions, the shading function (including the ambient term) is simply

$$\begin{aligned} L_i(\hat{n}) &= A_i + W_i(\langle \hat{w}_i | \hat{n} \rangle) \\ &= A_i + W_i \hat{w}_i.x \hat{n}.x + W_i \hat{w}_i.y \hat{n}.y + W_i \hat{w}_i.z \hat{n}.z \end{aligned} \quad (8)$$

Note that this is an affine (“linear”) function of the normal vector’s coordinates $\hat{n}.x, \hat{n}.y, \hat{n}.z$, with unknown coefficients $A_i, W_i w_i.x, W_i w_i.y, W_i w_i.z$.

Therefore, we compute the unknown parameters from the gauge photo G_i and the gauge normal map \hat{g} , by the following iterative procedure. Starting with a guess for the direction \hat{w}_i , we identify the subset \mathcal{G}' of all pixels q where $\hat{g}[q] \cdot \hat{w}_i \geq \sin \rho$. We then compute the coefficients $A_i, W_i w_i.x, W_i w_i.y, W_i w_i.z$ of formula (8), by a straightforward least-squares fitting of $G_i[q]$ over the pixels in \mathcal{G}' . From the fitted coefficients we extract an improved estimate for the direction \hat{w}_i .

This procedure is iterated until the set \mathcal{G}' has stabilized, and the fitted function is then extended to the whole sphere of normal directions \hat{n} . If $t = \hat{n} \cdot \hat{w}_i$ is greater than $\sin \rho$, $L_i(\hat{n})$ is defined by equation (8) above. If t is less than $-\sin \rho$, $L_i(\hat{n})$ is just A_i . If t lies between these two values, $L_i(\hat{n})$ is defined to be the unique quadratic polynomial in t that interpolates between the two parts with C_1 continuity. (The background grid is actually flat; the distortions in the reconstructed terrain seem to be due to local inhomogeneities in the light field.)

Figure 2(a) shows a photo of an actual light gauge, with several imperfections (including noticeable deviations from the ideal spherical shape). Figure 2(b) shows the artificial photo of a virtual light gauge, perfectly spherical, rendered with the simple shading function that was determined by the above procedure.

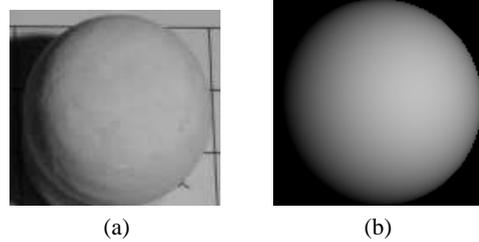


Figure 2. An actual gauge (a) and the virtual gauge (b)

4. Experiments

In order to validate the method, we used a set of images of ceramic fragments. All images tested had three gauges, placed near the top center of the image and near the bottom corners. The original images were acquired with a Sony Mavica CD-300 camera, in high-resolution mode (2048×1536 pixels, JPEG format). They were converted to PPM and reduced to 512×384 size so as to remove most of the camera noise and JPEG quantization artifacts.

The tests were performed with three images of each scene, taken with different illuminations. The normal maps computed by our method were converted to slope maps and integrated to produce height fields $Z(x, y)$. These height fields were then rendered from different viewpoints with a 3D terrain visualization program, to produce the images 3, 4, and 5 (The background grid is actually flat; the distortions in the reconstructed terrain seem to be due to local inhomogeneities in the light field.). All these images will be available at the project’s WWW site [3].

5. Conclusions

We have obtained fairly reliable results from multiple-image photometric stereo, by using photos of known light gauge objects to measure the actual light field, and numerical fitting to smooth out the shading functions derived from those photos.

We believe that the method can be extended and improved in many aspects, for instance by the use of more



Figure 3. A three-dimensional view of a reconstructed height field.

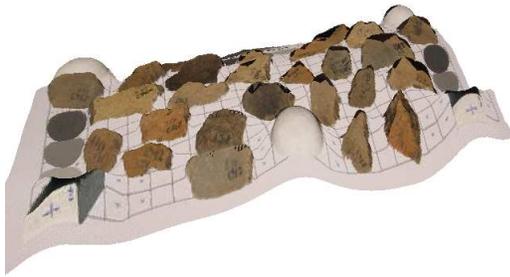


Figure 4. Another of three-dimensional view of the same field.

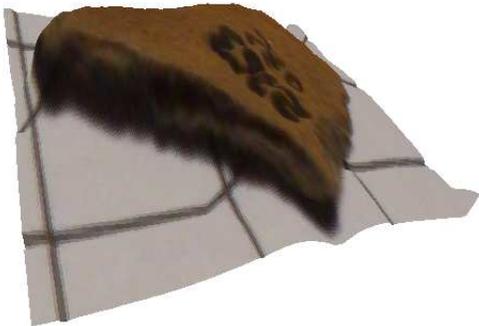


Figure 5. Another reconstructed height field.

Acknowledgments

[Suppressed to avoid author identification at the request of SIBGRAPI.]

References

- [1] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in presence of high-lights and shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, Oct. 2003.
- [2] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.
- [3] H. C. da Gama Leitão. Test images for the RAAB project. Electronic document at <http://www.ic.uff.br/~raab/figuras.html>, last accessed ??/??/???, 2003.
- [4] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. *The VLDB Journal*, pages 518–529, 1999.
- [5] A. Hertzmann and S. M. Seitz. Shape and materials by example: A photometric stereo approach. In *Proceedings IEEE CVPR 2003*, volume 1, pages 533–540, June 2003.
- [6] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, Aug. 2005.
- [7] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8(2):201–231, 1977.
- [8] B. K. P. Horn. Obtaining shape from shading information. In B. K. P. Horn and M. J. Brooks, editors, *Shape from Shading*, pages 121–171. MIT Press, Cambridge, Mass., 1989.
- [9] O. Ikeda. Shape reconstruction for color objects using segmentation and photometric stereo. In *Proceedings of ICIP 2004*, pages 1365–1368, 2004.
- [10] J. O’Rourke. *Computational Geometry in C*. Cambridge University Press, 1994.
- [11] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer, 1985.
- [12] A. suppressed to avoid identification]. [title suppressed]. 2005.
- [13] A. suppressed to avoid identification]. [title suppressed]. 2007.
- [14] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [15] R. J. Woodham. Determining surface curvature with photometric stereo. In *Proceedings of the 1989 IEEE International Conference on Robotics and Automation*, volume 1, pages 36–42, May 1989.
- [16] R. J. Woodham. Gradient and curvature from the photometric stereo method, including local confidence estimation. *Journal of the Optical Society of America, Series A*, 11(11):3050–3068, 1994.
- [17] L. Zhong and J. J. Little. Photometric stereo via locality sensitive high-dimension hashing. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV’05)*, pages ??-??+7, 2005.