Shape Reconstruction Using a Gauge-Based Photometric Stereo Method

Rafael Felipe V. Saracchini[1], Jorge Stolfi[1], Helena Cristina G. Leitão[2]

[1] Institute of Computing – University of Campinas (UNICAMP)
[2] Institute of Computing – Federal Fluminense University (UFF)

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. In order to validate the method, we presented tests using real objects.

1 Introduction

1.1 Variable-lighting photometric stereo

We describe a technique for *variable-lighting photometric stereo* (VLPS), which aims 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. 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.

The normal computation procedure also yields the albedo 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 attached 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 shows a typical set of input images.

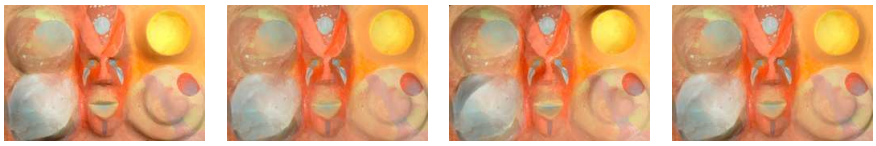


Fig. 1. Four images of a scene under different lighting conditions.

1.2 Related work

The principles of photometric stereo has been known 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). Those methods emphasized extraction of the depth information from a single monochromatic image [5].

The extraction of slope information from multiple images under different light fields, using light gauges, was pioneered by Woodham in 1980 [10] and has been extensively researched since then. Woodham [11] 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 (2bd), 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 vectors of the gauge object and the dimension look-up table has a prohibitive cost.

In 2005, Hertzmann and Seitz [4] 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). Zhong and Little [12] developed a method that extend a work of Hertzmann and Seitz [4]. To speed up the search, the authors adapted the LSH (*Locality Sensitive Hashing*) to perform searches for matches of brightness vectors in a high dimensional space. However, the above methods have a common shortcoming: they consider the set of all gauge signatures to be a generic cloud of points scattered in m-dimensional space, and therefore use general m-dimensional nearest-neighbor search algorithms, which are inherently expensive in space and/or time [6]. Our

search method [9] is exact and always yields the best matching entry in the table, and not merely a close approximation.

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 and Chandraker *et al.* [2] used a graph-cut approach to determine shadowed regions. Recently, we developed a non Euclidean metric to circumvent this problem [3].

2 Principles and notation

The main input data for our method is a list of $m \geq 3$ digital photos $S_1, .. S_m$ 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 a certain point of the scene is projected at 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 $\max\{0, \hat{u} \cdots \hat{n}\}$ 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.

In gauge-based VLPS, the relevant information about $\beta[p]$ and the flows Φ_i are obtained indirectly from a set $G_1, .. G_m$ 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. 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.

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 albedo* 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 albedo 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,

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

for some set of *shading functions* $L_1, .. L_m$. 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 albedo 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

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

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.

2.1 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 albedo map G^*. If the gauge has a simple

geometric shape and uniform color, that information can be computed analytically from a few parameters.

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

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

for all i and j. That is, the vectors

$$\begin{array}{c} \mathbf{S}[p] = (S_1[p], .. S_m[p]) = S^*[p](L_1[p], .. L_m[p]) \\ \text{and} \\ \mathbf{G}[q] = (G_1[q], .. G_m[q]) = G^*[q](L_1[q], .. L_m[q]) \end{array} \tag{4}$$

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

$$\mathbf{s}[p] = \frac{\mathbf{S}[p]}{|\mathbf{S}[p]|} \qquad \mathbf{g}[q] = \frac{\mathbf{G}[q]}{|\mathbf{G}[q]|} \tag{5}$$

and $|\cdot|$ is any norm of $\mathbb{R}^m$, e.g. the Euclidean norm $|\mathbf{X}| = \left(\sum_{i=1}^m X_i^2\right)^{1/2}$. 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 albedo map of the scene at p by

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

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

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

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 $\bar{\beta}$. 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.

2.2 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 [8, 7]. 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 the paper [9].

3 Light gauge processing

As explained in section 2.1, 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.

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 $\hat{s}$.

3.1 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.

3.2 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

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

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 $\hat{g}[q] \cdots \hat{w}_i \geq \sin \rho$. 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 $\hat{w}$. 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 $t = \langle \hat{w}_i | \hat{n} \rangle$ 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.

4 Experiments

4.1 Paper Models

In this test, we used a set of images of 3D paper models (a tetrahedron and a cone). All images tested had three balls of ball-mouses placed near the top center of the image and near the scene objects. The gray ball and the dust ball were test objects, and the white ball was the gauge. The original images were acquired with a Sony Cybershot DSC-W50 camera, in high-resolution 1632×1224 pixels, JPEG format). The test was performed with 23 images of the scene, taken with different illuminations (Figure 2).

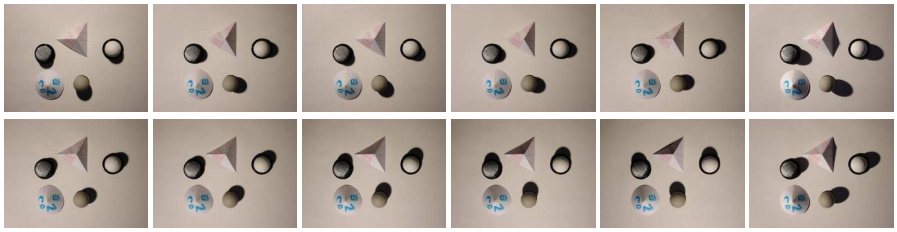


Fig. 2. Twelve of the 23 images of 3D paper models under different lighting conditions.

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 image 3.

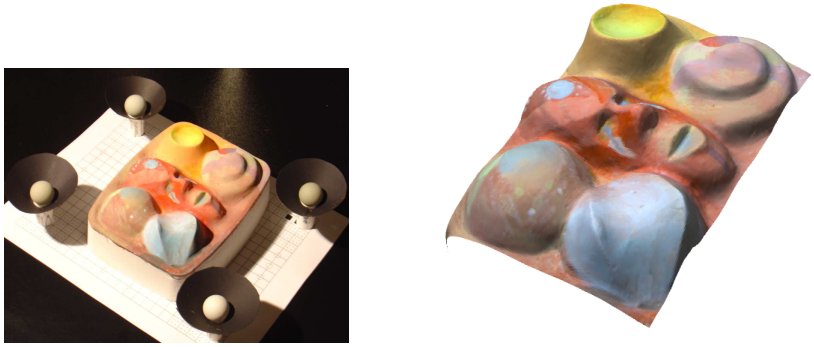
4.2 Plaster Sculpture

In this test, we used a scene consisting of a plaster sculpture painted with tempera colors (see figure 1). The light gauges were four gray rubber mouse-balls

Fig. 3. Three-dimensional views of the recovered height map.

— two of them unpainted, and two sprayed with matte white paint. The gauges were nested into shallow conical cups made of black paper, to block stray light from the scene (see figure 4(a)). We took 24 digital pictures of this scene (with different illuminations) with a 3 Mpixel consumer camera (Sony CyberShot DSC-V1). The digital photos were aligned, cropped from 2048×1536 to 672×1000 pixels, and then reduced to 168×250 in order to remove most of the JPEG compression noise.

Figure 4(b) shows a 3D view of the height map obtained by integrating the normal map, colorized with the albedo map.



(a) (b)

Fig. 4. A perspective photo of the plaster sculpture (a), and a 3D view of the recovered height map (b).

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. The main contributions of our method are (1) the use of simple 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. Our bucket-grid [9] scheme provides fast and accurate best-match table search. Our two-dimensional grid is faster and more space-efficient than approaches like LSH [12] or general nearest-neighbor data structures [4] due the disposition of the signature vectors on the m-dimensional space.

We are confident that our results will can be used for a software will become a helpful tool to solve a variety of 3D problems in medicine, geology, engineering, virtual reality, archaeology, and many other areas.

Acknowledgements

This project was partly supported by research grants from CNPq (304581/2004-6, 301016/19-5 and 550905/2007-3), FAPESP(2007/52015-0 and 2007/59509-9) and FAPERJ(JCNE-2007).

References

1. Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in presence of highlights and shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, October 2003.
2. Manmohan Krishna Chandraker, Sameer Agarwal, and David J. Kriegman. Shadowcuts: Photometric stereo with shadows. In *CVPR*, 2007.
3. Helena Cristina da Gama Leitão, Rafael Felipe V. Saracchini, and Jorge Stolfi. Matching photometric observation vectors with shadows and variable albedo. In *SIBGRAPI*, pages 179–186, 2008.
4. Aaron Hertzmann and Steven M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, August 2005.
5. B.K.P. Horn. Understanding image intensities. *Artificial Intell.*, 8(2):201–231, 1977.
6. Piotr Indyk and Rajeev Motwani. Approximate nearest neighbour: Towards removing the curse of dimensinality. In *Proceedings of the 13th Annual ACM Symposium on Theory of Computing (STOC'98)*, pages 604–613, 1998.
7. Joseph O'Rourke. *Computational Geometry in C*. Cambridge Univ.Press, 1994.
8. Franco P. Preparata and Michael I. Shamos. *Computational Geometry: An Introduction*. Springer, 1985.
9. Rafael F. V. Saracchini, Jorge Stolfi, and Helena C. G. Leitão. A bucket grid structure to speed up table lookup in gauge-based photometric stereo. In *XX Brazilian Symposium on Computer Graphics and Image Processing*, pages 221 – 230, 2007.
10. Robert J. Woodham. Photometric method for determining suface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
11. Robert 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.
12. Lin Zhong and James 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 104–111, 2005.