A Survey of Photometric Stereo Techniques

Jens Ackermann
TU Darmstadt
jens.ackermann@gris.informatik.tu-darmstadt.de

Michael Goesele
TU Darmstadt
michael.goesele@gris.informatik.tu-darmstadt.de
# Contents

1 **Introduction** 151
   1.1 Scope of this survey 152
   1.2 Overview 153

2 **Background** 154
   2.1 Light-Matter-Interaction 154
   2.2 Lambertian photometric stereo 155
   2.3 Error sources 156
   2.4 Baseline quality 164
   2.5 Integration of surface gradients 167

3 **Foundations** 171

4 **Unknown lighting** 176
   4.1 Point light source 177
   4.2 General light source 179
   4.3 Other generalizations 181

5 **Unknown reflectance** 183
   5.1 Diffuse and specular components 184
   5.2 More complex BRDF models 186
   5.3 Invariants 189
Abstract

Reconstructing the shape of an object from images is an important problem in computer vision that has led to a variety of solution strategies. This survey covers photometric stereo, i.e., techniques that exploit the observed intensity variations caused by illumination changes to recover the orientation of the surface. In the most basic setting, a diffuse surface is illuminated from at least three directions and captured with a static camera. Under some conditions, this allows to recover per-pixel surface normals. Modern approaches generalize photometric stereo in various ways, e.g., relaxing constraints on lighting, surface reflectance and camera placement or creating different types of local surface estimates.

Starting with an introduction for readers unfamiliar with the subject, we discuss the foundations of this field of research. We then summarize important trends and developments that emerged in the last three decades. We put a focus on approaches with the potential to be applied in a broad range of scenarios. This implies, e.g., simple capture setups, relaxed model assumptions, and increased robustness requirements. The goal of this review is to provide an overview of the diverse concepts and ideas on the way towards more general techniques than traditional photometric stereo.
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>normal</td>
</tr>
<tr>
<td>$X$</td>
<td>a 3D point</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>set of 3D points (i.e. a patch or mesh)</td>
</tr>
<tr>
<td>$L$</td>
<td>radiance</td>
</tr>
<tr>
<td>$L_s$</td>
<td>incoming (source) radiance</td>
</tr>
<tr>
<td>$\omega$</td>
<td>direction vector</td>
</tr>
<tr>
<td>$D$</td>
<td>light matrix/vector</td>
</tr>
<tr>
<td>$\omega_{in/out}$</td>
<td>incoming/outgoing direction vector</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>solid angle</td>
</tr>
<tr>
<td>$d\omega$</td>
<td>differential solid angle</td>
</tr>
<tr>
<td>$\rho$</td>
<td>BRDF or diffuse constant</td>
</tr>
<tr>
<td>$c,C$</td>
<td>arbitrary constant (scalar or matrix)</td>
</tr>
<tr>
<td>$M$</td>
<td>number of images</td>
</tr>
<tr>
<td>$I$</td>
<td>image intensity</td>
</tr>
<tr>
<td>$f$</td>
<td>camera response</td>
</tr>
<tr>
<td>$\theta,\phi$</td>
<td>angles (typically zenith and azimuth)</td>
</tr>
<tr>
<td>$p,q$</td>
<td>coordinates of the gradient</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>standard deviation</td>
</tr>
<tr>
<td>$N$</td>
<td>normal distribution</td>
</tr>
<tr>
<td>$Z$</td>
<td>depth map/height field</td>
</tr>
<tr>
<td>$(u,v)$</td>
<td>2D coordinates</td>
</tr>
<tr>
<td>$N$</td>
<td>normal map/normal field (possibly in matrix notation)</td>
</tr>
<tr>
<td>$E$</td>
<td>error (matrix, measure, etc.)</td>
</tr>
</tbody>
</table>

| $R$ | reflectance map or albedo matrix |
| $P$ | number of pixels/patches |
| $S,T$ | factorization of the radiance matrix |
| $A$ | transformation matrix |
| $B$ | arbitrary matrix |
| $B^\dagger$ | pseudoinverse of $B$ |
| $\alpha,\mu,\nu$ | parameters |
| $D_R$ | reflection vector |
| $\varphi$ | angle |
| $\mathcal{D},\mathcal{K}$ | factorization of $L$ |
| $\Phi$ | lobe function |
| $\mathcal{M}$ | general model |
| $W$ | reciprocal pair matrix |
| $o$ | general observation |
| $K$ | error free radiance matrix |
| $H$ | indicator matrix |
| $v$ | viewing direction |
| $W$ | intensity profile matrix |
| $w$ | intensity profile |
| $m$ | mixing coefficients |
| $\mathcal{S}$ | the unit sphere in $\mathbb{R}^3$ |
| $\mathcal{M}$ | manifold |
| $h_i$ | EMOR basis function |
| $\xi$ | shadow function |
The shape of an object, its reflectance, and the incoming illumination define the image that the object forms in our eyes or on a camera sensor. Even for a uniformly colored, diffuse object, the shading typically changes depending on the local orientation of the surface and the properties of the incoming light. Highlights form on glossy or specular surfaces, providing additional cues about the surface shape and reflectance as well as the illumination. Humans are remarkably good at deducing such information from images even under very general conditions (Thompson et al. [2011]). In contrast, recovering one or several of these individual components is a much more difficult task for a computer system.

To solve such inverse problems, different techniques exist such as shape from defocus (Pentland [1987]), shape from texture (Blostein and Ahuja [1989]), or (multi-view) stereo (Seitz et al. [2006]). In this survey, we direct our attention to photometric approaches. These exploit the intensity variations caused by illumination changes. Typically, the input is a set of images of an object or scene under a varying set of lighting conditions, and the output is the 3D shape, e.g. encoded as a field of surface normals.
The goal of classic photometric stereo is to recover the surface orientation from a known combination of reflectance and lighting in multiple images. Starting from this, a rich body of literature was developed that generalizes the problem and its solution in various directions. Many of these approaches assume a controlled capture setup because well-behaved input data is crucial for inverse problems. These requirements constrain the application mostly to the research community and production settings. In this survey, we put a focus on approaches with a potential to broaden this applicability, e.g., through simple capture setups or relaxed requirements.

1.1 Scope of this survey

Shape and appearance reconstruction have connections to different fields in computer vision, computer graphics, optimization theory, statistics, optics, etc. We will only cover those areas that are most relevant for this survey. These are approaches that rely on varying illumination in multiple images to recover at least the surface orientation and possibly even reflectance and illumination.

That excludes for example the large area of shape from shading approaches that operate on single images (Zhang et al. [1999], Johnson and Adelson [2011], Oxholm and Nishino [2012], Han et al. [2013]) and the related works on intrinsic image decomposition (e.g., Barron and Malik [2012, 2013]). These are highly ill-posed problems and their solution requires strong regularization. Using multiple images provides more information and better constrains the result space. Some works (e.g., Magda et al. [2001], Koppal and Narasimhan [2007], Liao et al. [2007]) exploit the fall-off in radiance with the distance from a near point light or more general changes in the apparent source intensity (e.g., Davis et al. [2005]). Our focus will be on illumination variations caused by directional changes in the incident light. We also do not consider purely specular surfaces (Healey and Binford [1986], Bonfort and Sturm [2003], Tarini et al. [2005], Chen et al. [2006], Nehab et al. [2008], Weinmann et al. [2013]) or the specialized approaches for face (Debevec et al. [2000], Zhou et al. [2007], Ghosh et al. [2011]) and dy-
namic performance (Ahmed et al. [2008], Vlasic et al. [2009], Wu et al. [2012]) capturing. Similarly, we will mention only a few approaches that require highly complex capture setups because they have less potential to be adapted for unconstrained environments.

In many modern works ideas from different disciplines get combined, e.g. when silhouette constraints are merged with photometric cues in certain multi-view approaches. This makes a clear distinction to other fields difficult at times. We include such cases in our discussion if it suits the overall perspective.

1.2 Overview

We begin this survey with an introduction for readers unfamiliar with photometric stereo methods in Chapter 2. It serves as a tutorial and to promote awareness for the challenges present in practical applications of such techniques. Before proceeding to the main part, we briefly discuss some early works in Chapter 3 that laid the foundations for present-day research.

Finding an optimal categorization of photometric approaches to structure this survey is not trivial. We decide for a selection scheme based on common challenges among the algorithms such as unknown lighting in Chapter 4, complex reflectance in Chapter 5, or extremely uncontrolled conditions in Chapter 9. Within each category we aim at presenting relevant representatives instead of cluttering the exposition with lists of similar approaches.

Finally, we conclude the survey in Chapter 10 with a summary of the most important concepts and our views on future developments.
This chapter summarizes some theoretical and practical background information to make the survey more self-contained. It aims to be an introduction for readers unfamiliar with photometric reconstruction methods. In §2.1 and §2.2, we establish some notation that is used throughout this work and introduce a few formulas that will be referenced later. We then explain the challenges arising in real-world setups and discuss about the accuracy to be expected from photometric stereo in §2.3 and §2.4. The results obtained from such techniques are typically estimates of the surface gradient or orientation. If the actual 3D geometry, e.g. represented as a height map, is required, such a normal field has to be integrated. §2.5 presents common strategies towards that end.

2.1 Light-Matter-Interaction

When light interacts with matter on a macroscopic level, i.e. discounting wave-like effects, it can be absorbed, transmitted, or reflected. Although all of these effects occur simultaneously, we will here only consider the fraction of light that is reflected.
2.2 Lambertian photometric stereo

Let \( X \in \mathbb{R}^3 \) be a surface point with normal \( n \in S := \{ Y \in \mathbb{R}^3, \| Y \| = 1 \} \). The solid angle of an object with respect to a point \( X \) is the surface area of its projection onto a unit sphere around \( X \). Let \( L_s(\omega_{\text{in}}) \) be the incoming radiance from an infinitesimal solid angle \( d\omega_{\text{in}} \) centered around the direction \( \omega_{\text{in}} \in S \). The irradiance on a small patch around \( P \) coming from the whole hemisphere of directions \( \Omega \) is the integral

\[
\int_{\Omega} L_s(\omega_{\text{in}}) \langle n, \omega_{\text{in}} \rangle \, d\omega_{\text{in}}.
\]  

(2.1)

The total radiance \( L(\omega_{\text{out}}) \) leaving \( X \) in the direction \( \omega_{\text{out}} \) is then given as

\[
L(\omega_{\text{out}}) = \int_{\Omega} \rho(\omega_{\text{out}}, \omega_{\text{in}}) L_s(\omega_{\text{in}}) \langle n, \omega_{\text{in}} \rangle \, d\omega_{\text{in}},
\]

(2.2)

where the function \( \rho \) is the proportionality factor between the radiance leaving the surface point and the irradiance caused by incoming light from \( \omega_{\text{in}} \). Nicodemus et al. [1977] calls \( \rho \) the bidirectional reflectance-distribution function (BRDF). A detailed look into these radiometric topics is provided, e.g., in Chapter 4 of Forsyth and Ponce [2011].

2.2 Lambertian photometric stereo

Many photometric stereo techniques, cf. §4.1, consider the special case of Lambertian reflectance, i.e. \( \rho(\omega_{\text{out}}, \omega_{\text{in}}) = \rho/\pi \), and a single point light source \( L_s(\omega_{\text{in}}) = c \delta(\omega_{\text{in}} - \omega_s) \) from direction \( \omega_s \). In this case, the outgoing radiance is independent of the direction \( \omega_{\text{out}} \) and the integral in Equation [2.2] vanishes:

\[
L = \frac{\rho c}{\pi} \max\left( \langle n, \omega_{\text{in}} \rangle, 0 \right).
\]

(2.3)

The clamping of negative dot products will be implicitly assumed from here on.

Observing a Lambertian surface point \( M \) times under varying light directions \( \omega_{\text{in}} \), which we denote as direction vectors \( D_1, \ldots, D_M \in \mathbb{R}^3 \), and source strength \( c_1, \ldots, c_M \) yields radiance values

\[
L = (L_1, \ldots, L_M)^T = \frac{\rho}{\pi} \left( n^T D_1 c_1, \ldots, n^T D_M c_M \right)^T =: \frac{\rho}{\pi} D \cdot n,
\]

(2.4)

where the scaled row vectors \( c_i D_i \) form the \( M \times 3 \) light matrix \( D \).
Expressing the Lambertian image formation using matrix notation makes all tools from linear algebra available to study the photometric stereo problem. For $M \geq 3$ and three linearly independent directions, the linear system is overdetermined but has full rank. Thus, $D^T D$ is invertible and the pseudoinverse $(D^T D)^{-1} D^T$ exists. Multiplying Equation 2.4 from the left

$$(D^T D)^{-1} D^T L = \frac{\rho}{\pi} (D^T D)^{-1} D^T D \cdot n = \frac{\rho}{\pi} n$$

(2.5)

allows to compute the normal from $D$ and $L$:

$$n = \frac{(D^T D)^{-1} D^T L}{\| (D^T D)^{-1} D^T L \|}$$

(2.6)

### 2.3 Error sources

In a real setup, photometric stereo techniques have to deal with various sources of error. We believe that these need to be understood to achieve the goal of broader applicability and dedicate this section to the discussion of the most common ones. In addition, a couple of simple experiments will provide an intuition about the impact of some of the errors.

#### 2.3.1 Noise

All photometric techniques rely on radiance measurements. Like with any other measurement process, the observed values might be subject to noise. One can reduce its impact by capturing additional data, e.g. averaging multiple images under the same conditions. Whether its influence can be neglected, however, depends on the specific application and setting.

Sometimes, the noise levels for specialized equipment like cameras for industrial inspection or microscopy are provided in their specification sheet, cf. Point Grey Research, Inc. [2015]. For consumer cameras, however, such data is rarely available. Nevertheless, these cameras are regularly used in computer vision applications and will play a central role in bringing photometric methods to a broader audience. Quantifying sensor noise precisely requires highly controlled conditions,
2.3. Error sources

<table>
<thead>
<tr>
<th></th>
<th>planar</th>
<th>sphere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/20s</td>
<td>1/10s</td>
</tr>
<tr>
<td>mean (%)</td>
<td>27.8</td>
<td>56.7</td>
</tr>
<tr>
<td>deviation (%)</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 2.1: Statistics computed over a small patch in the green channel of 51 images from an 18 megapixel consumer sensor (Canon EOS 700D). The experiment for the planar target is repeated for three different exposure times and a patch of 16 × 16 pixel. For the sphere, the exposure time is 1/10s and the patch size is 10 × 10 pixel. Results are listed in percent of the full sensor range.

cf. European Machine Vision Association [2010]. Instead, we ask what the noise levels are in an experimental setup (in situ) with a consumer camera in a darkened room containing a high-quality, but not exceptional light source and a real object.

Noise levels depend very much on the sensor model, as shown by the data in Point Grey Research, Inc. [2015]. Table 2.1 gives an impression of the noise behavior for the Canon EOS 700D. For exposure times of 1/20s, 1/10s, 1/8s, we repeatedly captured 51 raw images of a planar reflectance standard under constant illumination. The left part of Table 2.1 shows the statistics computed in the green channel of a small patch in that image stack. We notice a clear dependence of the standard deviation on the mean intensity.

To exclude effects of exposure time changes, we keep it fixed at 1/10s and perform a second experiment with a diffuse sphere. Statistics are now computed over two patches corresponding to dark and bright regions on the sphere. The right part of Table 2.1 shows a similar relationship of mean and standard deviation. This indicates that it does not depend on the exposure time. Overall, we observe that in reasonably exposed images the noise level is below 1% of the full sensor range.

2.3.2 Shadows

The argumentation in §2.2 leading to the solution of the photometric stereo problem in Equation 2.6 implicitly assumes that the surface is
Background

not in shadow in any of the images. If something blocks the light, the observed radiance will vanish independent of the orientation of the surface. Thus, Equation 2.6 is not fulfilled and any solution based on that model will be incorrect. Similar restrictions also apply for more advanced lighting and reflection models than the Lambertian one.

As we will see in §6.1, shadows still convey some information about the scene that can be exploited. In general, however, it is important to either avoid shadows already during capture. One solution is, e.g., to place the light source close to the camera. This can lead, however, to ill-conditioned inverse problems if the lights become too similar. Another strategy is to detect and address shadows after capture, e.g. excluding those observations from further computations.

2.3.3 Interreflections

The exitant radiance at a point according to Equation 2.2 depends on the incoming radiance from all directions. The integral vanishes in Equation 2.3 for a single point light source. This is, however, only strictly valid for an isolated surface patch in empty space or convex objects in general. In all other cases, the incoming radiance at a point \( p \) will consist of the radiance exiting the light source and the radiance leaving any other surface point in the direction towards \( p \). These global illumination effects are usually not modeled in the kind of inverse problems studied here. Whether this leads to significant errors depends on the scene geometry and reflectance, e.g. a mirror reflecting light onto \( p \). In practice, at least the effects of interreflections from the environment can be reduced by covering non-target surfaces with black cloth. This is, however, not easily possible in uncontrolled scenarios.

2.3.4 Camera

As discussed above, noise is present in any measurement process. In the case of modern photometric techniques, the measurement device is a camera and as such has some additional sources of systematic error.
2.3. Error sources

Falloff

One kind of error is introduced by the optical system of the camera. So-called vignetting leads to reduced image intensities in peripheral regions compared to the center. One way to address this issue is by calibrating the falloff based on an evenly illuminated gray card. Another possibility is to reconstruct only in small regions—in terms of the image plane—where the falloff can be assumed constant.

Radiance to pixel conversion

The camera transforms continuous radiance $L$ into digital pixel values $I$. Most cameras use a sensor with an added color filter array. Thus, individual sensor elements capture contributions either of the red, green, or blue channel. These contributions have to be interpolated over the image plane to produce a full-sized, colored image with RGB information at each pixel. Effects introduced by this interpolation are relatively small compared to other errors (and can often be avoided by using raw image data).

Discounting any color space conversions or image processing such as sharpening filters, a simple model of the transformation from radiance to pixel intensity is given by the camera response curve $f(L) := I$. If raw values are available, the response is usually a clamped, linear function

$$f(L) = \min(\max(\alpha L - I_{\text{dark}}, 0), I_{\text{sat}})$$

for a certain scaling $\alpha$ that depends on the sensor and the quantization resolution (ranging from 10 to 14 bit for typical consumer cameras). $I_{\text{dark}}$ accounts for the nonzero output level that is present even if no light actually reaches the camera, e.g. with the lens cap attached.

If raw values are not available, the response of the camera is non-linear and has to be calibrated. Lots of methods for this task exist, e.g. based on calibration targets (Grossberg and Nayar [2004]), stacks of differently exposed images (Debevec and Malik [1997], Robertson et al. [1999, 2003], Kim et al. [2008]), color edges in a single image (Lin et al. [2004]), or probabilistic color de-rendering (Xiong et al. [2012]).
Ignoring non-linear responses is almost never a viable option since it can lead to drastically different results. We illustrate this with a short example for a simplified setting. Let \( n = (0, 0, 1) \) and assume that the light forms an angle \( \theta \in [0, \pi/2] \) with the z-axis in two images and shines from above in the third:

\[
\begin{bmatrix}
\sin \theta & 0 & \cos \theta \\
0 & \sin \theta & \cos \theta \\
0 & 0 & 1
\end{bmatrix}, \quad \begin{bmatrix}
\frac{1}{\sin \theta} & 0 & -\frac{\cos \theta}{\sin \theta} \\
0 & \frac{1}{\sin \theta} & -\frac{\cos \theta}{\sin \theta} \\
0 & 0 & 1
\end{bmatrix}.
\] (2.8)

The ensuing radiance values are

\[
L = (L_1, L_2, L_3) = (\cos \theta, \cos \theta, 1).
\] (2.9)

Applying photometric stereo as in §2.2 on the non-linear pixel intensities \( I = f(L) \) yields a normal \( \tilde{n} \) which deviates from the true normal \( n \) by an angle of

\[
\cos^{-1}(\langle \tilde{n}, n \rangle) = \cos^{-1}\left( \frac{D^{-1}f(L), n}{\|D^{-1}f(L)\|} \right) = \cos^{-1}\left( \frac{f(1)}{\|D^{-1}f(L)\|} \right).
\] (2.10)

Figure 2.1 shows the angular error for varying \( \theta \) and \( \gamma \) in the commonly assumed—though often oversimplifying—gamma curve response model \( f(x) = x^\gamma \). The systematic deviations from the true normal illustrate the importance of applying the correct inverse response prior to any photometric reconstruction.

**Camera motion**

Accurate correspondences between pixels in all images are important for photometric stereo. If no image alignment is applied as preprocessing, the use of a solid tripod and remote trigger are essential. The impact of this error depends on the scene content, e.g. it would not matter for a homogenous plane.

In multi-view settings such as discussed in Chapter 8 the camera is moved on purpose and its position is often assumed to be known. If this knowledge is inaccurate, e.g. due to bad pose estimation, the projection of a 3D point will occur at the wrong pixel, thus breaking correspondence. Combining pixels from multiple images will then mix information from different surface points leading to erroneous results.
2.3. Error sources

\[ \theta \]

\[ D_1 \]

\[ D_2 \]

\[ n \]

\[ \gamma = 1.2 \]

\[ \gamma = 1.8 \]

\[ \gamma = 2.2 \]

Figure 2.1: Impact of the non-linear response on photometric stereo. Left: Example setup with two light directions at an angle \( \theta \) with the normal and the third shining from above (\( D_3 = -n \)). Right: The angular deviation of the reconstructed normal if the intensities are not linearized properly.

2.3.5 Light source

For those algorithms relying on calibrated light sources, their position and intensity become another set of observables that are again subject to noise. Many ways to measure these have been proposed. The approaches differ in accuracy, capture setup, lighting model, additional constraints, and in the intended application, \( e.g. \), renderings in augmented reality, shape reconstruction, or image-based relighting. For photometric stereo applications, it is often assumed that the light is a simple point light source and that its location is known very precisely.

Some works exploit cast shadows \cite{Panagopoulos2011}, sample the complete incoming light field \cite{Sato1999, Kanbara2004}, or estimate the light source from stationary images \cite{Winnemoeller2005}. Approaches that minimize an intensity error compare actual images of a scene with known geometry and reflectance against renderings obtained from the current light estimate \cite{Hara2005, Weber2001, Xu2008}. These are, however, unsuited in contexts where the shape is to be reconstructed in a later step.

We narrow down the discussion to the estimation of point light sources based on target objects of known, usually spherical, geometry.
If space allows to place the light far away from the scene, techniques to recover its overall direction have been proposed by Dosselmann and Yang [2013], Wang and Samaras [2002], Wong et al. [2008], or Zhou and Kambhamettu [2002]. If the light cannot be placed sufficiently far away, however, its direction is not constant for every scene point, and the irradiance falls off with the square of the distance. Masselus et al. [2002] and Powell et al. [2001] demonstrate that some of the ideas used for recovering infinitely distant illumination can be readily used to estimate positions of near point light sources if applied to multiple spheres. The framework of Zhou and Kambhamettu [2004] is even able to triangulate an area light source by placing a specular sphere at different locations in the scene.

Here, we assume a single point light source and perform a small experiment to assess the influence of the camera on reconstructions obtained from a mirroring sphere. We also try to give a rough idea of the errors that one can expect in such a standard lab setting. For this experiment, we keep the light fixed at approximately 4 m distance and move a reflective sphere to 20 different positions spread about 25 cm around the center of the scene. Each sphere leads to a highlight that is observed by the camera (in our case a Canon 700D with a 110 mm lens). Given the radius and the position of the spheres, we can compute the light direction for each of the highlights by shooting a ray towards the sphere and reflecting it around the normal at the intersection point. Taking the perspective projection of the lens into account, the rays through the highlight pixels will fan out slightly. If we instead make the simplifying assumption of a purely orthographic camera—an assumption applied by many photometric stereo methods—these rays will all be parallel.

To compare both models, we create a synthetic dataset that allows us to study them without noise or inaccuracies in the highlight detections or sphere positions. For a point light source and a perspective camera, we analytically determine the highlight pixels and the center of the sphere in the image. We adjusted the position of the light source (sitting behind, above, and to the left of the camera) to match the highlights of the real experiment as closely as possible. We then reconstruct
2.3. Error sources

Figure 2.2: Accuracy of light direction estimation from a mirrored sphere. Each marker is the projection of one of the 20 sphere positions onto the image plane (x- and y-axes are in pixels). The gray value encodes the angular deviation from the synthetic ground truth. A cross indicates the optical axis of the camera. The varying distance of the sphere positions from the camera is not shown in this plot. Light directions reconstructed based on a simplifying orthographic camera model show a strong spatial dependency, both in a simulated, noise-free (left) and a real experiment (middle). For a perspective reconstruction, the errors between experiment and synthetic ground truth are all below $1.2^\circ$ (right).

Figure 2.3: Visualizing the direction of the error for the same experiments as in Figure 2.2. For each sphere, the difference of the stereographic projections of the reconstructed light direction and the ground truth are color mapped according to the reference on the right.

the light direction for each sphere assuming an orthographic model. Figure 2.2 shows the angular deviation from the directions computed from the true perspective model. The direction of the error is visualized in Figure 2.3. As expected the error increases towards the edges of the field of view where the orthographic model deviates more strongly from a perspective one.

Next, we evaluate the gap between theory and practice by comparing the synthetic light directions with those reconstructed from the real experiment. For the perspective reconstruction, the deviation is below $1.2^\circ$ in all cases and does not show strong spatial variations in Figure 2.2. This indicates that effects from an inaccurate calibration or noisy highlight detections cause only minor errors. The same holds for
the orthographic reconstruction which behaves similar to the simulated case.

Nevertheless, the performance could be improved by taking erroneous detections into account. Aoto et al. [2012] and Ackermann et al. [2013] allow for such inaccuracies by considering the reprojection error in the triangulation of near light sources.

2.4 Baseline quality

Given the large number of potential error sources discussed in §2.3, one might ask whether photometric techniques have any chance of success. Common quality measures for the overall result of such methods are the angular normal error, the height differences of the integrated surface, or the intensity error in rendered images. Here, we use the angular error and conduct a controlled experiment in order to provide an intuition of the quality level that can be achieved. The controlled setting thereby establishes a baseline for more complex algorithms that will be discussed later.

Experiment

The target object is a sphere made of Spectralon, a highly Lamber
tian material. We use a sharply defined light source with a very stable radiance output that we place several meters away to approximate a distant point light. To estimate the light direction, we place two mirror spheres in the scene. Since the extent of the whole setup is small, the respective light vectors are very similar and we just average them. The experiment is performed with a consumer camera (Canon EOS 700D) equipped with a 135 mm telephoto lens. We took great care to avoid the error sources listed in §2.3 as much as possible, e.g. covering all surfaces with cloth, using raw images, averaging ten pictures for each light position to reduce noise, etc.

Figure 2.4 contains the results for all pixels that were never in shadow for any of the six light sources. The error in Figure 2.5 is spatially varying because the real world surface has fine ridges, porous structure, and several pronounced dents that deviate from a true
2.4. Baseline quality

Figure 2.4: Experimental results for all pixels that were never in shadow for any of the six light sources. \textit{Left:} The estimated light source directions in spherical camera coordinates. \textit{Right:} Color coded normals of the reconstruction and an ideal sphere (inset).

Figure 2.5: Comparison with an ideal sphere. \textit{Left:} The spatial error distribution (mean: 2.5°, variance: 1.3°) shows fine ridges and dents where the target object deviates from an ideal sphere. \textit{Right:} A cumulative histogram of the angular errors corresponding to the reconstruction on the left.

sphere. In addition, we observe a low-frequency bias. The errors are higher on the left and right side than on the top or bottom. One possible source for this systematic error is the deviation of the Spectralon BRDF from a true Lambertian reflector as pointed out by Bhandari et al. [2011]. Another lies in the configuration of light sources which yields a non-uniform error distribution. The effect of the latter will be confirmed through synthetic experiments below, \textit{cf.} Figure 2.6.

Overall, the histogram in Figure 2.5 shows that the deviation from ground truth is about 2.5° in general. These findings are in line with other experiments on photometric stereo with known light sources as
Background

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
<th>Comparison</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ray et al. 1983</td>
<td>C</td>
<td>sphere</td>
<td>&lt; 5°</td>
</tr>
<tr>
<td>Tagare and deFigueiredo 1991</td>
<td>C</td>
<td>sphere</td>
<td>4.8°</td>
</tr>
<tr>
<td>Wu and Tan 2013</td>
<td>C+U</td>
<td>sphere</td>
<td>4°</td>
</tr>
<tr>
<td>Shi et al. 2010</td>
<td>U</td>
<td>calibrated PS</td>
<td>6−7°</td>
</tr>
<tr>
<td>Favaro and Papadhimitri 2012</td>
<td>U</td>
<td>calibrated PS</td>
<td>5−12°</td>
</tr>
<tr>
<td>Papadhimitri and Favaro 2013</td>
<td>U</td>
<td>calibrated PS</td>
<td>2−3°</td>
</tr>
<tr>
<td>Abrams et al. 2012</td>
<td>S</td>
<td>Google Earth</td>
<td>20°</td>
</tr>
<tr>
<td>Tunwattanapong et al. 2013</td>
<td>S</td>
<td>sphere</td>
<td>&gt; 5°</td>
</tr>
</tbody>
</table>

Table 2.2: Evaluation results in a selection of works on calibrated (C) and uncalibrated (U) photometric stereo. Also listed are two examples of specialized techniques (S) with more complex models. The comparisons are either performed with another, presumably more accurate, technique as reference or based on objects with known geometry such as a sphere or simple models from Google Earth.

listed in Table 2.2. We conclude that in a controlled setting calibrated photometric stereo achieves an accuracy of about 3° − 5°. The table also indicates that uncalibrated techniques differ from calibrated ones by several degrees. If compared to ground truth geometry, however, they achieve similar results as shown by Wu and Tan 2013. Obviously, the less assumptions about the target, e.g. regarding reflectance, or the capture setup are made, the bigger these errors may become as shown by two exemplary techniques in the bottom rows.

Error Analysis

The evaluations above and in Table 2.2 are based on empirical data. We can also take an ab initio point of view to study the effect of the measurement errors. Based on the assumption of known error magnitudes in the image intensity and light measurements, Ray et al. 1983 present a theoretical error analysis for the Lambertian case. They assume a setup with three light sources on a circle around the optical axis. Then, the photometric stereo problem can be solved in closed form, cf. Equation 2.6. We encode this in a function $h$:

$$n = h(I_1, I_2, I_3, \theta, \phi_1, \phi_2, \phi_3)$$

(2.11)
where $\phi_i$ and $\theta$ are the azimuth and zenith angles of the light sources. A first order Taylor expansion then yields an approximation of the error in gradient space $n = (-p, -q, 1)$ given the deviations $dI_1, dI_2, dI_3, d\theta, d\phi_1, d\phi_2, d\phi_3$ of the input data:

$$dp = \sum_i \frac{\partial h_1}{\partial I_1} dI_i + \sum_i \frac{\partial h_1}{\partial \phi_1} d\phi_i + \frac{\partial h_1}{\partial \theta} d\theta, \quad dq = \sum_i \frac{\partial h_2}{\partial I_1} dI_i + \ldots$$ (2.12)

For more than three light sources, a solution $h$ involves the pseudo inverse and computing the derivative is non-trivial. We therefore use a different strategy to analyze settings such as in the empirical experiment above. Assuming a ground truth normal $n_{GT}$ and light directions $\hat{\Theta} = (\theta_1, \ldots, \theta_6), \hat{\Phi} = (\phi_1, \ldots, \phi_6)$, we model the observables with normal distributions

$$I \sim \mathcal{N}(D(\hat{\Theta}, \hat{\Phi}) \cdot n_{GT}, \sigma_I), \quad \Theta \sim \mathcal{N}(\hat{\Theta}, \sigma_\theta), \quad \Phi \sim \mathcal{N}(\hat{\Phi}, \sigma_\phi)$$ (2.13)

where $D(\hat{\Theta}, \hat{\Phi})$ is the light matrix. Then, we draw random samples $\tilde{\Theta}, \tilde{\Phi}, \tilde{I}$ from each distribution and compute the resulting normal as

$$n = n(\theta_n, \phi_n) = \frac{D(\tilde{\Theta}, \tilde{\Phi})^\dagger \tilde{I}}{\|D(\tilde{\Theta}, \tilde{\Phi})^\dagger \tilde{I}\|}$$ (2.14)

where $D^\dagger$ denotes the pseudoinverse.

Repeating this sampling 10000 times lets us estimate the error for any given normal $n_{GT}$ for a fixed light configuration $\hat{\Theta}, \hat{\Phi}$. We use the same set of normals and light directions as in the real experiment and set $\sigma_I = 0.8 \%, \sigma_\theta = 0.7^\circ, \sigma_\phi = 1.5^\circ$ to reflect the findings in §2.3. The results in Figure 2.6 show that deviations vary spatially over the sphere and are not isotropic. The accuracy decreases in the left and bottom regions because they are furthest away from the light directions. We also observe from Figure 2.6 that, given the uncertainties in the input data, we cannot expect to obtain much better results than a $1^\circ$ error. This puts the error of $2.5^\circ$ in the actual experiment into perspective and concludes our attempt to establish a baseline of quality.

### 2.5 Integration of surface gradients

Many photometric techniques recover the gradient of a surface, \textit{i.e.} its orientation. On the other hand, it is often desirable to reconstruct the
Figure 2.6: Results from simulated input errors through random sampling. Left: At each pixel, corresponding to the experiment in Figure 2.5, the angular error is averaged over 10000 random experiments. The nine pixels in red are selected for a closer examination. Center: For each of the exemplary normals marked in the left image, we fit a Gaussian to the stereographic projection of the reconstructed normals over 10000 random experiments. The ellipses correspond to the $1\sigma, 2\sigma, 3\sigma$ intervals and show the spatially varying spread induced by the light configuration (blue) used in the experiment. Right: A histogram of all angular errors from the plot to the very left.

actual surface, and not just its derivative. Here, we summarize some key ideas for surface integration to complete the introductory chapter.

If the object is represented as a height field $(u, v, Z(u, v))$ over the image plane, its gradient field is defined as

$$\nabla Z : \mathbb{R}^2 \to \mathbb{R}^2, (u, v) \mapsto \left( \frac{\partial Z(u, v)}{\partial u}, \frac{\partial Z(u, v)}{\partial v} \right) = \begin{pmatrix} Z_u(u, v) \\ Z_v(u, v) \end{pmatrix}. \quad (2.15)$$

What we obtain from photometric stereo is a vector field

$$g : \mathbb{R}^2 \to \mathbb{R}^2, (u, v) \mapsto \begin{pmatrix} p(u, v) \\ q(u, v) \end{pmatrix}. \quad (2.16)$$

or, interpreted as surface normals,

$$N : \mathbb{R}^2 \to \mathbb{R}^3, (u, v) \mapsto \frac{1}{\sqrt{1 + p^2 + q^2}} (p(u, v), q(u, v), -1). \quad (2.17)$$

2.5.1 Path integrals

One strategy to recover the surface is to integrate the gradient field. Assume the photometric stereo reconstruction is perfect, \textit{i.e.} $p = Z_u$,
2.5. Integration of surface gradients

$q = Z_v$, and the gradient is differentiable. Then, the definition

$$
\tilde{Z}(u, v) := \int_{\gamma} g(w) \cdot dw = \int_{\alpha}^{b} g(\gamma(t)) \cdot \gamma'(t) \, dt \tag{2.18}
$$

is independent of the choice of $\gamma$ (which is an arbitrary path from $(0, 0)$ to $(u, v)$). Conversely, it holds that $\nabla \tilde{Z} = g$. Thus, the path integral in Equation 2.18 is one way to obtain a solution. It is, however, not unique and might differ from the true solution by an additive constant because $\nabla (\tilde{Z} + c) = \nabla \tilde{Z} = g$. Examples of this approach can be found in the works by Wu and Li [1988] or Klette and Schluens [1996].

2.5.2 Variational formulation

In practice, the reconstruction is not perfect: the recovered gradients might differ from the true surface derivatives or they might not form an integrable vector field. It is therefore more common, cf. Durou et al. [2009], to define the solution in a variational framework and minimize

$$
E(\tilde{Z}) = \iint (\tilde{Z}_u(u, v) - p(u, v))^2 + (\tilde{Z}_v(u, u) - q(u, v))^2 \, d(u, v) \tag{2.19}
$$

or similar error measures. Moreover, such a formulation can be incorporated directly into the reconstruction algorithm as shown by several shape from shading approaches, e.g. Horn and Brooks [1986]. To avoid outliers and discontinuities, it can be beneficial to multiply the gradient terms with spatially varying weights as for example demonstrated by Agrawal et al. [2006]. For practical computations it is often required to discretize the problem, e.g. in the spatial domain. Xie et al. [2014] show how this can be achieved by casting the integration as a mesh deformation task.

2.5.3 Additional constraints

Another approach to obtain absolute depth from gradients is to add additional information to guide the integration. This usually comes in the form of known depth values $\hat{Z}$ at sparse points or in the whole image area obtained through other means, e.g. laser scanning (Horovitz and Kiryati [2004]), active stereo methods (Nehab et al. [2005]), or depth
Background

Background cameras (Haque et al. 2014). A possible extension of Equation 2.19 is, for example,

$$E(\tilde{Z}) = \int \int (\tilde{Z}_u - p)^2 + (\tilde{Z}_v - q)^2 + (\tilde{Z} - Z)^2 \, d(u, v).$$  \hspace{1cm} (2.20)

Combining normals and depth in such a way is a key part in several of the multi-view photometric stereo methods discussed in Chapter 8.
At lot of the techniques used today still rely in part on methods and concepts developed 30 years ago. It is instructive to briefly look at some of these early works and give credit to the ideas presented.

One of the first applications of photometry was in exploring the surface of the Moon. Rindfleisch [1965] derives and solves a differential equation for the distance of a surface point from the image plane that depends on the angles of incident and emitted light. His derivation is based on paths in the image plane, and the final integration depends on at least one depth value along a path to be known. The paths are chosen as straight lines that meet in the point given by the intersection of the image plane with the ray emanating from the camera center in the direction of the Sun. Rindfleisch assumes a certain reflection function for the surface of the Moon that essentially depends on two angles only. Furthermore, the camera geometry must be known and the discussion relies heavily on the Sun as known point light source.

Inspired by that work, Horn [1970] formulates the analytical shape from shading problem for arbitrary, but known, isotropic reflectance and known light sources. Again, this amounts to a first order partial differential equation. He transforms this into five ordinary differential

Equations that are solved along characteristic strips grown from an initial curve, see Figure 3.1. The theoretical formulation and the treatment of several special cases that simplify the occurring equations are truly pioneering work. Interestingly, Horn also mentions limitations of his imaging equipment. This aspect has become even more important today as computer vision is employed in consumer hardware with unknown characteristics.

In his PhD thesis—supervised by Horn—Woodham [1977] combines reflectance, illumination, and viewing geometry into a single function, a so-called reflectance map, that relates surface orientation directly to image intensities

$$I(u, v) = R(n_{u,v}).$$

(3.1)

In general, such a relationship can only be established if each object point receives the same incident illumination, has the same reflectance, and is observed from the same direction. This amounts to a distant light source, an untextured object, and an orthographic camera. The problem of inverting Equation 3.1 in a single image is similar to the work by Horn [1970], but instead of formulating a set of differential
equations, Woodham defines additional constraints that can guide the inversion process. More important than the single image case is his extension of the photometric shape recovery problem to multiple images under varying illumination. He coins the term photometric stereo for a scenario of two or more independent equations

\[ I_1(u, v) = R_1(n_{u,v}) \]
\[ \vdots \]
\[ I_M(u, v) = R_M(n_{u,v}) \]

and explores this concept in the much cited paper [Woodham 1980]. Since a unit normal \( n_{u,v} \) can be described by two angles, Equation 3.2 is an over-determined system of \( M \) non-linear equations in two unknowns. Figure 3.1 shows how the intersection of the curves arising from each equation defines a solution in gradient space. In the case of a distant point light source \( c \) shining from direction \( \omega_i \) and Lambertian reflectance \( \rho/\pi \), we obtain

\[ R_i(n_{u,v}) = \frac{\rho}{\pi} c \langle n_{u,v}, \omega_i \rangle. \]  

Inserting into Equation 3.2 yields the formulation in Equation 2.4 which is the basis of many subsequent works in this area.

Horn et al. [1978] pick up the concept of a reflectance map and enhance the previously mentioned methods to cope with spatially varying albedo in the Lambertian case by means of ratio images

\[ I_1 = \rho(u, v)R_1(n_{u,v}) \]
\[ I_2 = \rho(u, v)R_2(n_{u,v}) \]

\[ \rightarrow I_1 : I_2 = \frac{R_1(n_{u,v})}{R_2(n_{u,v})} =: R_{12}(n_{u,v}). \]  

The authors also give an overview about the possibilities of shape reconstruction from shading in one, two, or three images and compare these with stereo based methods. Several properties of the latter are complementary to those of photometric methods and provide the motivation for more recent, hybrid approaches.

One of the biggest hindrances for the broad application of photometric stereo is, however, its necessity for carefully controlled lighting conditions. Silver [1980] notes this problem in his master’s thesis and
develops an approach that works under varying but otherwise unknown light sources. It is based on a look-up scheme that relies on a reference object with the same reflectance as the target. This allows matching of intensity sequences between target and reference. He transfers the normals from the known reference based on the insight that two surface points with the same normal reflect the same amount of light. Several modern approaches draw inspiration from this orientation consistency as we will describe in §7.3.

These techniques rely on a known reflectance map either through calibration or a reference object. Coleman and Jain [1982] present a technique that works for unknown, non-Lambertian objects without this restriction. They propose to use four known light sources even though three would suffice in the Lambertian case. This allows them to compute a solution for each of the $\binom{4}{3} = 4$ combinations of three light sources. If the surface point is Lambertian, these are very similar and will all lead to the same albedo estimate. If the point exhibits a specularity under one of the sources, these estimates differ and the normal with the lowest albedo is accepted. Figure 3.2 illustrates this idea. The procedure relies on the assumption that a surface point behaves almost diffusely under most illuminations and that specularity only arises for a few constellations. Barsky and Petrou [2003] extend this approach using four known light sources with different color, which simplifies

Figure 3.2: Four normals are computed from each combination of light triplets. Left: For a diffuse point all normals coincide. Right: If the point exhibits specularity under one of the light sources (dashed) the deviation among normals is higher. (Reprinted from E. North Coleman and Ramesh Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. Computer Graphics and Image Processing, 18:309–328, 1982. Copyright 1982, with permission from Elsevier.).
detection of highlights if the light source color is sufficiently different from the Lambertian surface color. If light source and surface color are similar, they propose an alternative method to identify a specular highlight or shadowed light source based on the direction of the erroneous initial surface normal estimate.

One of the early approaches to consider varying viewpoint in addition to lighting conditions is presented by [Hartt and Carlotto 1989]. Their formulation is cast in a probabilistic framework and allows the inclusion of smoothness priors and a model of image noise. Its core is a comparison of observed intensities $I$ with renderings of hypothesized height fields $Z$. This is already quite similar to many modern approaches, as we will see in the following chapters.
The first works on photometric methods, presented in the previous chapter, assumed that the light source is a known quantity. This is a realistic assumption in controlled settings such as an industrial factory or scientific laboratory. Lots of techniques have been developed to provide this kind of input data and we briefly listed some of them in §2.3. Ultimately, we are interested in uncontrolled settings where this information is not accessible.

Performing an additional preprocessing step to estimate light source positions is cumbersome and introduces another source of errors. More importantly, equipment such as precise calibration spheres or positioning devices might not be available in a given acquisition setup. We will therefore look at uncalibrated photometric stereo techniques in §4.1 and §4.2 which recover the shape and lighting at one go.

These approaches are categorized according to the lighting model they employ. They all assume, however, an ideal camera model that might not be adequate for some real scenarios. §4.3 presents generalizations that lessen the calibration effort and allow for a broader applicability.
4.1 Point light source

One of the first methods to cope with the problem of unknown lighting was presented by Hayakawa [1994] and gave rise to other so called uncalibrated photometric stereo approaches. He arranges the radiance at $P$ pixels in $M$ images into a matrix

$$L = \begin{pmatrix} L_{1,1} & \cdots & L_{1,P} \\ \vdots & \ddots & \vdots \\ L_{M,1} & \cdots & L_{M,P} \end{pmatrix}. \quad (4.1)$$

Assuming Lambertian reflectance, we know from Equation 2.4 that

$$L = C \cdot D \cdot N \cdot R =: T \cdot S. \quad (4.2)$$

where the diagonal matrix $C \in \mathbb{R}^{M \times M}$ contains the source radiance, $D \in \mathbb{R}^{M \times 3}$ represents the light directions, $N \in \mathbb{R}^{3 \times P}$ is the stack of all normals, and $R \in \mathbb{R}^{P \times P}$ contains the reflection coefficients on its diagonal.

For given $L$, the goal is to find the matrix $S$. However, multiple candidates $\hat{T}, \hat{S}$ might fulfill

$$L = \hat{T} \cdot \hat{S}. \quad (4.3)$$

In fact, any invertible $3 \times 3$ matrix $A$ defines a candidate pair $\hat{T} := T \cdot A, \hat{S} := A^{-1} \cdot S$. Such a candidate pair can be obtained from $L$ using singular value decomposition. Additional constraints are necessary to find the actual $S, T$.

Hayakawa proposes to use six or more pixels with the same or known albedo, i.e.

$$\rho_1^2 = \|S_1\|^2 = \langle S_1, S_1 \rangle, \quad \ldots, \quad \rho_6^2 = \|S_6\|^2 = \langle S_6, S_6 \rangle \quad (4.4)$$

where $S_i$ are the corresponding columns of $S$. Choosing the same columns in a candidate matrix $\hat{S}$ yields

$$\rho_i^2 = \|S_i\|^2 = \langle S_i, S_i \rangle = \langle A\hat{S}_i, A\hat{S}_i \rangle = \langle \hat{S}_i, A^T A\hat{S}_i \rangle. \quad (4.5)$$

These equations constrain the entries of the symmetric matrix $B = A^T A$, which has six degrees of freedom. Once this system of equations
Unknown lighting

is solved, $A$ can be recovered—up to an unknown rotation—from $B$ using singular value decomposition. Multiplying the candidate with this transformation yields the final result $S = A \cdot \hat{S}$.

Techniques like the one just presented reconstruct a normal field $n(u, v)$ without considering an underlying surface. Instead, [Belhumeur et al., 1999] formulate the problem in terms of a height field, i.e. a graph $(u, v, Z(u, v))$, with scaled normals

$$n(u, v) = (\partial_u Z, \partial_v Z, -1). \quad (4.6)$$

To define a surface, $Z$ must satisfy the integrability constraint

$$\partial_u \partial_v Z = \partial_v \partial_u Z, \quad (4.7)$$

which had already been used for shape from shading, e.g. by [Horn and Brooks, 1986] and [Frankot and Chellappa, 1988]. This provides constraints on any normal field that belongs to a surface. Not all transformed candidates $A \cdot \hat{S}$ are able to fulfill those. [Belhumeur et al., 1999] show that for an integrable normal field $S$, the set of matrices $A$ that preserve this property is equivalent to

$$\begin{pmatrix} 1 & 0 & -\mu/\lambda \\ 0 & 1 & -\nu/\lambda \\ 0 & 0 & -1/\lambda \end{pmatrix} \quad (4.8)$$

for parameters $\mu, \nu \in \mathbb{R}$ and $\lambda > 0$. In practice, that means that an integrable normal field can be recovered by photometric methods only up to such a generalized bas relief transformation if no additional information is available. Figure 4.1 illustrates this finding with two differing constellations of light and surface that nevertheless yield the same image.

The bas relief transform not only applies to normals but also transforms the diffuse albedo. [Alldrin et al., 2007] exploit the fact that many objects are composed of a small set of albedo values whose histogram gets broadened by a bas relief transform. They define an energy based on the entropy of the albedo distribution. Minimizing this energy yields the parameters of the correct transform.

[Pavaro and Papadimitri, 2012] look at the function $f(u, v) := \langle n(u, v), \omega_s \rangle$ defined over the image domain. They discover that a maximum of this function constrains the parameters $\mu, \nu$ of the allowable
4.2. General light source

The works discussed above assume a distant point light source. That is an idealized setting. It is often violated, e.g. by ambient light that is hard to prevent in a real setup. Fortunately, the ideas introduced for point lights can easily be generalized to more complicated illumination.

Yuille and Snow [1997] use a similar matrix decomposition approach as Hayakawa but extend the shading model by a constant ambient term.
They enforce the integrability constraint to resolve the ambiguity in $A$ up to a generalized bas relief transform. The final solution is then defined by additionally assuming a light source with constant radiance.

Handling arbitrary unknown illumination is usually addressed by a decomposition of incoming radiance. If the incoming illumination $L_s$ can be decomposed into individual components

$$L_s = \sum_{j=0}^{r} \lambda_j L_{s,j},$$

(4.10)

this transfers to the outgoing radiance. Assuming Lambertian reflectance in Equation 2.2, we obtain

$$L = \int_\Omega \frac{\rho}{\pi} \sum_j \lambda_j L_{s,j}(\omega_{in}) \langle n, \omega_{in} \rangle \, d\omega_{in}$$

(4.11)

$$= \sum_j \lambda_j \int_\Omega \frac{\rho}{\pi} L_{s,j}(\omega_{in}) \langle n, \omega_{in} \rangle \, d\omega_{in}$$

(4.12)

where $L_j$ is the radiance observed if the scene was illuminated just by $L_{s,j}$. Under these assumptions, the matrix in Equation 4.1 can be decomposed as

$$L = \begin{pmatrix} \lambda_{1,1} & \ldots & \lambda_{r,1} \\ \vdots & \ddots & \vdots \\ \lambda_{1,M} & \ldots & \lambda_{r,M} \end{pmatrix} \cdot \begin{pmatrix} L_1(n_1) & \ldots & L_1(n_P) \\ \vdots & \ddots & \vdots \\ L_r(n_1) & \ldots & L_r(n_P) \end{pmatrix}.$$
the scaled normal at each pixel. Higher order decompositions reduce the approximation error.

In a second paper, [Basri and Jacobs, 2001b] show how to exploit these insights for photometric stereo reconstructions under arbitrary, unknown illumination. They use singular value decomposition to obtain a candidate factorization \( L \approx \tilde{\Lambda} \tilde{L}_s \) as the best rank \( r \) approximation of \( L \). Similar to the discussion about the work by [Hayakawa, 1994], this factorization is only unique up to an \( r \times r \) linear transformation. [Basri and Jacobs, 2001b] use a normalization constraint to reduce this ambiguity to a Lorentz transformation in the case \( r = 4 \). Enforcing integrability again leads to a unique solution.

### 4.3 Other generalizations

The works discussed so far assumed an orthographic camera model like the majority of photometric stereo approaches. [Papadhimitri and Favaro, 2013] show how to incorporate a perspective camera model in a photometric reconstruction. More importantly, they find that a perspective formulation of the uncalibrated photometric stereo problem, when enforcing the integrability constraint, does not suffer from the bas relief ambiguity. Their results indicate also that an incorrect focal length or principal point can lead to strongly biased normals.

If in addition to a perspective camera the assumption of a directional light source does not hold, e.g. in endoscopic setups, the photometric stereo problem becomes even more general. Even if the position of the light source is known, its direction at each pixel is not because it depends on the depth of the observed point. Furthermore, light attenuation due to the \( 1/r^2 \) fall-off and angle dependent emittance of the light source lead to non-linearities in the equations. [Collins and Bartoli, 2012] solve this problem in a multi step optimization procedure that involves not only the normals as unknowns but also the depth at each pixel. A more rigorous formulation is derived by [Mecca et al., 2014] who come up with a set of partial differential equations to encode the constraints on the depth mapping. They present two numerical schemes that recover the depth without separating it from its gradient, i.e. the
normals. Integrating the differential equations requires the depth at one or more pixels to be known from which suitable integration paths are propagated outward.

None of these approaches considers the radiometric calibration of the camera. It is either assumed that calibration can be performed in a preprocessing step or is circumvented by using special cameras with linear response. Shi et al. [2010] propose an uncalibrated photometric stereo technique that also recovers the camera response curve. They exploit that the ratio between color channels for a single pixel is constant in a sequence of linear images but curved if a non-linear response is present. Measuring the non-linearity of the profiles in the RGB space allows them to define a minimization problem on the coefficients of a polynomial response model. For photometric stereo, they automatically select pixels with equal albedo and different normals which allows them to remove the bas relief ambiguity as in the approach by Hayakawa [1994].

Another aspect that is often overlooked are non-convex shapes. For example, a point in a deep crevice will not be visible from the entire hemisphere above its local tangent plane. In that case, some entries in the light matrix might be missing which we briefly discuss in Chapter 6. Such effects are a lot harder to detect and handle in the presence of extended light sources. Then, the integration in Equation 4.11 would not be performed over all possible light directions and thus yield an attenuated radiance. If the light distribution can be split into a directional light source and a constant ambient term

$$L_s = L_{s,\text{dir}} + L_{s,\text{con}} \quad (4.14)$$

this attenuation is called ambient occlusion. Hauagge et al. [2013] estimate it from a set of images by studying a statistic over the pixel intensity distribution. They derive how this statistic relates to ambient occlusion and show its effectiveness in practice. Their approach could be a preprocessing step for photometric stereo techniques.
The appearance of a given object depends on its reflectance and the illumination. The approaches in Chapter 4 were focused on the case where the latter is unknown. They assume, however, that the surface reflects light according to the Lambertian model. We will now study the complementary setting in which the illumination is known or can be controlled and the reflectance is unknown.

Most real surfaces have reflectance properties that are neither purely Lambertian nor perfect mirrors. The linear equations in §2.2 no longer hold and other reflection models have to be employed. This makes a direct inversion of the underlying formulation in Equation 2.2 more challenging as we will see in §5.2.

This chapter presents examples for the most common high-level strategies for shape reconstruction in the presence of unknown reflectance. Diffuse-specular separation in §5.1 and parametric modeling in §5.2 both have the disadvantage of model assumptions that might not be fulfilled by an arbitrary real-world BRDF. They have, however, been shown to work for a broad range of materials in practice. A benefit of parametric BRDF models is that they can be used to create novel impressions of the scene and that they fit into artists’ editing pipelines.
The approaches based on general invariants such as symmetries in §5.3 only recover shape but are more general because they do not assume a fixed reflectance model. The downside is that they typically require special constellations of light source and camera during capturing.

5.1 Diffuse and specular components

A commonly followed path is to split the radiance into a diffuse and specular contribution

$$L = \hat{\alpha}_{\text{diff}} \cdot L_{\text{diff}} + \hat{\alpha}_{\text{spec}} \cdot L_{\text{spec}}. \quad (5.1)$$

Nayar et al. [1989] assume that the specular term consists of a single spike and approximate it with a delta function. That corresponds to a BRDF

$$\rho(\omega_{\text{out}}, \omega_{\text{in}}) = \hat{\alpha}_{\text{diff}} + \hat{\alpha}_{\text{spec}} \cdot \delta(D_R - \omega_{\text{in}}) \langle n, \omega_{\text{in}} \rangle \quad (5.2)$$

in Equation 2.2 where $D_R = -\omega_{\text{out}} + 2 \langle \omega_{\text{out}}, n \rangle n$ is the reflection vector. They only study the two-dimensional case and reconstruct surface orientation $\varphi_n$ within the plane formed by $\omega_{\text{in}}$ and $\omega_{\text{out}}$. Using an extended light source $L_s(\varphi, \varphi_s)$ centered around direction $\varphi_s$ yields

$$L = \alpha_{\text{diff}} \cdot \cos(\varphi_s - \varphi_n) + \alpha_{\text{spec}} \cdot L_s(2\varphi_n, \varphi_s), \quad (5.3)$$

where all angles are with respect to the viewing direction $D_{\text{out}}$.

Nayar samples this function $L(\varphi_s, i)$ for multiple known light source directions as shown in Figure 5.1. If the step size equals the extent of the light source, at most two samples $i, i+1$ are affected by the specular component because it is a delta peak. From these, $\alpha_{\text{spec}}$ and an estimate $\varphi_{n, \text{spec}}$ of the normal can be obtained. The remaining samples determine $\alpha_{\text{diff}}$ and $\varphi_{n, \text{diff}}$. They perform this computation for all $i$ and select the result where $\varphi_{n, \text{diff}}$ and $\varphi_{n, \text{spec}}$ are most similar.

Another way to deal with non-diffuse surfaces is to exploit polarization. As observed by Mersch [1984], specular reflections can often be removed using cross polarization. This requires placing polarization filters in front of both light sources and cameras and adjusting their polarization angle to be perpendicular (or crossed). In many cases, the
5.1. Diffuse and specular components

Figure 5.1: The image intensity at a surface point with normal $\phi_n$ is a function of the source direction $\phi_s$. If the sampling interval matches the extent of the light source, only the samples closest to the reflection vector $2\phi_n$ are affected by the specular component. If the sampling interval matches the extent of the light source, only the samples closest to the reflection vector $2\phi_n$ are affected by the specular component.

Koshikawa [1979], Nayar et al. [1993] use this principle to estimate the specular term from multiple images obtained by rotating a polarization filter in front of the camera. Removing its contribution yields an image as if the object was diffuse.

Sato and Ikeuchi [1994b] extend the work of Nayar et al. [1989] to a simultaneous analysis of all three color channels in RGB space. Using a similar reflection model, they consider multiple samples in Equation 5.3 and arrive at the matrix expression

$$L = D \cdot K$$

This model corresponds to the dichromatic reflection model introduced by Shafer [1984] who notes that the specular vector $(\alpha_{\text{spec}, R}, \alpha_{\text{spec}, G}, \alpha_{\text{spec}, B})$ usually has the same color as the light source. Sato and Ikeuchi estimate this vector from several pixels of different color. Then, the diffuse component $(\alpha_{\text{diff}, R}, \alpha_{\text{diff}, G}, \alpha_{\text{diff}, B})$ is obtained from similar arguments as exploited by Nayar et al. [1989]—most samples capture only the diffuse contribution. Finally, they recover the matrix $D$ from Equation 5.4 by inverting the color matrix $K$. To get the
actual orientation $\varphi_n$, however, knowledge of the light source directions $\varphi_{s,i}$ is required.

Mallick et al. [2005] also assume a dichromatic reflection model but propose to rotate the original RGB color space into an SUV color space so that the S channel of the color space corresponds to the color $(c_R, c_G, c_B)^T = C$ of the light source:

$$L_{SUV} = A \cdot L$$

with rotation $A \in SO(3)$ and $A \cdot C = [1, 0, 0]^T$. The two remaining channels U and V depend only on the diffuse reflectance component and can be analyzed using standard Lambertian photometric stereo methods. This allows the method to be applied to arbitrary colored or textured surfaces as long as the color of the light source $C$ is known or can be estimated. Purely white surfaces can, however, not be recovered since in this case the U and V channels contain no signal.

### 5.2 More complex BRDF models

Instead of approximating specular reflectance with a delta peak, more sophisticated BRDF models allow for broader lobes and smeared out highlights. Tagare and deFigueiredo [1991] give theoretical insights on photometric stereo under such $m$-lobed reflectance maps

$$R_i(n_{u,v}) = \sum_{j=1}^{m-1} \rho_j \Phi_j(p_{i,j}^T n_{u,v}) + \rho_m,$$

$$\Phi_j : \mathbb{R}^+ \to \mathbb{R}^+ \text{ positive and monotonically increasing},$$

$$p_{i,j} \in \mathcal{S} \text{ directions in the plane spanned by } \omega_{in,i}, \omega_{out},$$

$$p_{i,j}^T n_{u,v} \text{ scalar product with the normal},$$

that encompass fore- and backscattering cases. An example configuration is shown in Figure 5.2. The authors confirm experimentally that a simple Lambertian assumption for the reconstruction of non-Lambertian objects leads to high errors and thus motivate the use of more sophisticated models.

While Tagare and deFigueiredo explicitly invert the photometric stereo equation similar to §2.2, another approach is to employ non-
5.2. More complex BRDF models

Figure 5.2: Rough surfaces exhibit a mix of different reflection phenomena. They can be modeled as a weighted sum of individual contributions such as diffuse and specular reflection or backscattering lobes. (©1991 IEEE. Reprinted, with permission, from Hemant D. Tagare and Rui J. P. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 1991).

Linear optimization techniques to minimize the image intensity error

$$\arg \min_{N, \rho} \| I - \mathcal{M}(N, \rho) \|,$$

(5.10)

where $\mathcal{M}$ can be any model of light transport in the scene. However, such an optimization is not straightforward due to the possibly large amount of parameters and their non-linear relationship.

Goldman et al. [2005] alternatingly estimate the reflectance properties $\rho$ of a parametric BRDF model by Ward [1992] and the surface orientation in the form of a normal field $N$. Varying the reflectance parameters $\rho_{u,v}$ at each surface point would lead to lots of unknowns, require a huge amount of images, and affect the robustness negatively. Instead, they recover only a small set of basis BRDFs $\rho_j$ and per-pixel mixing coefficients $\gamma_{u,v,j}$ similar to Lensch et al. [2003]:

$$\rho_{u,v} = \sum_j \gamma_{u,v,j} \rho_j.$$

(5.11)

The optimization still requires a good initialization, which they obtain using Lambertian photometric stereo.

Aittala et al. [2013] also estimate spatially varying reflectance using a parametric BRDF model—based on sums of Gaussians—and non-linear optimization. In contrast to Goldman et al. [2005], they derive one set of parameters at each pixel. Their technique relies on an extended light source, e.g. a LCD display, that projects a series of illu-
mination patterns onto the target. It is primarily a BRDF acquisition setup but also recovers surface normals for almost planar objects.

An even more challenging case of unknown reflectance is considered by Dong et al. [2014a]. They study materials, such as skin or soap, that exhibit subsurface scattering. Then, reflectance at a single point can no longer be described by a four dimensional BRDF and applying photometric stereo leads to blurred results. Instead, they model outgoing radiance as a convolution with a bidirectional subsurface scattering distribution function (BSSRDF). A non-linear optimization that involves a blind deconvolution step recovers normals that are much closer to the ground truth.

The complexity of the employed BRDF models in these approaches leads to non-linear formulations that are often hard to optimize. Shi et al. [2012b] address this by introducing a simplified reflectance model based on a biquadratic factorization of isotropic BRDFs. They show that it approximates the low-frequency components of other parametric models and measured BRDFs. If the light source positions are known, surface normals and BRDF parameters can be reconstructed by simply alternating two linear least squares optimizations.

Ikehata and Aizawa [2014] also transform the problem of recovering normals and BRDF into a representation that is better suited for optimization. They base their derivations on the multi-lobed reflectance model introduced by Chandraker et al. [2011]

\[ \rho_{u,v} = \sum_{j=1}^{m} \Phi_j \left( p_j^T n_{u,v} \right) \] (5.12)

where \( \Phi_j \) are nonlinear functions and \( p_j \) are directions along which these functions are concentrated. The authors prove that if \( p_j \) lies in the plane spanned by the viewing and lighting direction, the whole image formation can be represented by a bivariate function \( f \)

\[ L_i(u,v) = h_i \left( \langle n_{u,v}, \omega_{in,i} \rangle, \langle \omega_{in,i}, \omega_{out,i} \rangle \right) \] (5.13)

with certain monotonicity constraints. These ensure the existence of an unknown function \( g \) such that

\[ \langle n_{u,v}, \omega_{in,i} \rangle = g \left( \langle \omega_{in,i}, \omega_{out,i} \rangle, L_i(u,v) \right) \] (5.14)
and thus separate the unknown normal from the nonlinear function. Approximating $g$ using Bernstein basis polynomials then leads to a constrained linear problem.

Note that the above BRDF models only approximate the real reflectance behavior and may break down in specific cases (typical examples are anisotropic reflectance or retro-reflectance). One thus needs to always check whether these models and techniques are applicable to the specific reconstruction scenario at hand.

### 5.3 Invariants

A different approach to handle complex BRDFs is to exploit general invariants, such as symmetries and physical properties, in the image formation model or capture setup. Such techniques are usually independent of any explicit parametric reflectance model. This is an advantage in terms of generality but can be a disadvantage if editing of scene appearance is desired.

[Zickler et al. 2002] use Helmholtz reciprocity, which states that a BRDF is invariant if the incoming and outgoing directions are swapped: $\rho(\omega_{\text{in}}, \omega_{\text{out}}) = \rho(\omega_{\text{out}}, \omega_{\text{in}})$. Their setup consists of a camera and a light source which can be exchanged to acquire a reciprocal pair. The intensity—ignoring irradiance fall-off due to the light source distance—of a surface point in both images of a pair yields

$$
\begin{align*}
I_r &= \rho(\omega_{\text{in}}, \omega_{\text{out}}) \cdot \langle n, \omega_{\text{in}} \rangle \\
I_l &= \rho(\omega_{\text{out}}, \omega_{\text{in}}) \cdot \langle n, \omega_{\text{out}} \rangle \\
\end{align*}
\Rightarrow \quad 0 &= \langle n, I_r \omega_{\text{out}} - I_l \omega_{\text{in}} \rangle. 
$$

(5.15)

For multiple pairs, the vectors $w$ can be stacked into a matrix $W$ that fulfills $0 = n^T W$ and rank $W = 2$. Note that $\omega_{\text{in}}, \omega_{\text{out}}$ are assumed to be in a coordinate system local to the surface point $X$ and depend on its position in relation to the camera/light. Using the rank of $W(X)$ as an indicator of surface depth, this method recovers the 3D position $X$. Once the correct depth is found, the normal is given by the kernel of $W(X)$.

Capturing reciprocal pairs requires a careful calibration and camera/light placement. [Zickler 2006] shows how such a calibration can
be obtained from features in the reciprocal images themselves. It remains, however, necessary to swap the light and camera for each pair. Delaunoy et al. [2010] use the same concept as Zickler et al. [2002] but formulate an energy over a global 3D model instead of separate depth maps.

In many cases, other kinds of symmetries can be exploited besides Helmholtz reciprocity. One example is bilateral symmetry for isotropic BRDFs, i.e., symmetry across the plane of incidence

$$\rho(\theta_{\text{in}}, \phi_{\text{in}}, \theta_{\text{out}}, \phi_{\text{out}}) = \rho(\theta_{\text{in}}, \phi_{\text{in}}, \theta_{\text{out}}, \phi_{\text{in}} + (\phi_{\text{out}} - \phi_{\text{in}}))$$

(5.16)

$$= \rho(\theta_{\text{in}}, \phi_{\text{in}}, \theta_{\text{out}}, \phi_{\text{in}} - (\phi_{\text{out}} - \phi_{\text{in}})),$$

(5.17)

which was introduced in Marschner’s Ph.D. thesis [Marschner 1998]. Alldrin and Kriegman [2007] use the concept of bilateral symmetry based on so-called isotropic light pairs. This requires a circle of light positions—parametrized by the angle $\varphi$—centered about the optical axis of the camera. For a single pixel, the radiance distribution $L(\varphi)$ given by these light positions exhibits a symmetry around a certain angle $\varphi_g$ depicted in Figure 5.3. This angle and the viewing direction define a plane that contains the normal. The authors show that this information is sufficient to recover iso-depth contours of the surface. Further constraints are necessary to obtain the unique normal or the absolute depth.
5.3. Invariants

Figure 5.4: Two common parametrizations of a BRDF. **Top:** Incoming $\omega_{in}$ and outgoing $\omega_{out}$ directions are encoded as azimuth in the tangential plane and zenith angle with respect to the normal. **Bottom:** The incoming direction is encoded by difference angles $(\theta_d, \phi_d)$ with respect to the halfway vector $h = (\omega_{out} - \omega_{in}) / \|\omega_{out} - \omega_{in}\|$ which is itself represented as zenith and azimuth angles $(\theta_h, \phi_h)$. $\omega_{out}$ is implicitly defined by reflecting $\omega_{in}$ around $h$.

Using this technique as initialization, [Alldrin et al. 2008] present a system that recovers not only surface orientation but also BRDFs. This makes it possible to render novel views of captured objects under different illumination. They make the additional assumption that the reflectance is well represented by a bi-variate function. Based on the halfangle representation introduced by [Rusinkiewicz 1998], cf. Figure 5.4, [Romeiro and Zickler 2010b] show that this holds if the BRDF does not change too much for rotations of the incoming and outgoing
Unknown reflectance

directions around the halfway vector
\[
\rho(\theta_h, \phi_h, \theta_d, \phi_d) \approx \rho(\theta_h, \phi_h, \theta_d, \tilde{\phi}_d).
\] (5.18)

The technique by Alldrin et al. then relies on a set of basis materials which are represented as data samples and thus do not depend on a parametric model. The normals and material weights at each pixel and the basis BRDFs for the full object are found through an alternating optimization.

As mentioned above, bilateral symmetry alone determines only the azimuth angle of the normals. Shi et al. [2012a] give an example of an additional constraint to also recover the zenith angle. It is based on the assumption of a BRDF as in Equation 5.12 with a dominant lobe concentrated around the halfway vector \( \Phi(h^T n) = \tilde{\Phi}(\theta_h) \). In general, the monotonicity \( \theta_{h_1} > \theta_{h_2} \Rightarrow \tilde{\Phi}(\theta_{h_1}) < \tilde{\Phi}(\theta_{h_2}) \) is broken for an incorrect normal estimate. The approach requires light directions to be known and distributed equally on the hemisphere to detect such a violation. Holroyd et al. [2008] also investigate symmetries similar to Alldrin and Kriegman [2007], but additionally consider tangents. This makes it possible to study anisotropic materials. They define a measure of symmetry for a pair \((n, t)\) of normal and tangent under three possible reflection operations. Minimizing that measure at each pixel yields a local coordinate frame which can, for example, be used to fit a parametric BRDF model. Their technique, however, requires thousands of images with known light position.

The methods discussed so far required the light source positions to be known explicitly. Chandraker et al. [2011] again use light sources that move on a circle around the optical axis but do not require their actual position. They exploit image derivatives both in the spatial and in the temporal domain, \textit{i.e.} neighboring pixels and successive light directions, to define a ratio that is constant over time and independent of the BRDF. From this invariant, it is possible to recover the direction of the gradient and, similar to the idea of Alldrin and Kriegman [2007], the iso-depth contours. The advantage of this technique is that it works for unknown light source positions as long as they fulfil the circular motion assumption. A disadvantage in practice is that the reconstruction involves higher-order derivatives of image intensities, which tend to be
5.3. **Invariants**


unstable due to noise. Furthermore, additional input data is needed to completely determine depth and normals.

As in Chapter 4, all techniques discussed here, except the one by Aldrin and Kriegman [2007], rely on pixel values that are linearly related to scene radiance. Higo et al. [2010] exploit three properties present in many BRDFs which even hold for observations under an unknown camera response function. From known light source directions $\omega_i$, they define a monotonicity constraint

$$o_i > o_j \iff \langle \omega_i - \omega_j, n \rangle > 0,$$

(5.19)

a visibility constraint

$$\langle \omega_i, n \rangle > 0,$$

(5.20)

and an isotropy constraint

$$o_1 \simeq \ldots \simeq o_k \implies \text{Var}_i(\langle n, \omega_i \rangle) \text{ minimal.}$$

(5.21)

Figure 5.5 visualizes this abstract formulation. While each of the constraints is relatively weak, e.g. visibility is valid for all normals in a half-space, combining them in a consensus manner over lots of light directions yields very good results.
No matter whether an approach is able to cope with unknown illumination as in Chapter 4 or with unknown reflectance as in Chapter 5, it always assumes that the input data adheres to a fixed model or behaves in a certain way. If these assumptions are repeatedly violated—beyond statistical measurement errors—in a real scenario, the results become unusable.

Several effects, e.g. shadows, interreflections, image noise, non-linear camera response, non-ideal light sources, etc., are not accounted for in the simple model employed by many photometric stereo methods. They are often treated as outliers, but sometimes they can even be exploited explicitly to obtain additional information. We will discuss some examples of the latter in §6.1 and then look at more general outlier treatment in §6.2.

The ideas from §6.2 can also be exploited in the context of Chapter 5 if non-Lambertian reflectance is considered as an outlier. This is similar in spirit to the concept of Coleman and Jain [1982] introduced in Chapter 3. It requires, however, that sufficiently many images show a surface point without specularity.
6.1 Explicit treatment

6.1.1 Interreflections

Nearly all approaches neglect global illumination effects, i.e. light that reaches a surface patch not directly from the source but after one or more bounces at other surfaces. The actually observed radiance then has a direct and indirect component $L = L_{\text{direct}} + L_{\text{indirect}}$ from which only $L_{\text{direct}}$ follows the model in Equation 2.3. Nayar et al. [1990] discuss the impact of interreflections on photometric reconstruction techniques. They assume the scene to consist of infinitesimal, planar patches of Lambertian reflectance similar to radiosity approaches in computer graphics (Goral et al. [1984]). The geometric relations between patches can be encoded in a kernel matrix $B \in \mathbb{R}^{P \times P}$:

$$
B = \begin{pmatrix}
0 & b_{1,2} & \cdots \\
b_{2,1} & 0 & \cdots \\
& & \ddots & \ddots
\end{pmatrix}
$$

where $b_{i,j}$ encodes the transfer of energy between the interreflecting surface patches $i$ and $j$. So the observed radiance can be written as:

$$
L = (I - RB)^{-1} L_{\text{direct}} = (I - RB)^{-1} RND
$$

(6.1)

where $I$ is the identity matrix, the diagonal matrix $R \in \mathbb{R}^{P \times P}$ contains the albedo and $N \in \mathbb{R}^{P \times 3}$ the normal of each patch, and $D \in \mathbb{R}^{3 \times M}$ the light source directions. As hinted at by Figure 6.1, a standard photometric stereo technique would result in false estimates $\hat{N}, \hat{R}$:

$$
\hat{R} \hat{N} := L \cdot D^{-1} = (I - RB)^{-1} R N \iff RN = (I - RB) \cdot \hat{R} \hat{N}.
$$

(6.2)

Nayar et al. solve this problem by iteratively updating the left hand side with calculations of the right hand side based on estimates from the previous iteration. Chandraker et al. [2005] study a similar problem in the context of uncalibrated photometric stereo. They are able to show that interreflections actually resolve the generalized bas relief ambiguity.
6.1.2 Shadows

Shadows are another source of inconsistencies. One distinguishes cast shadows—generated because light towards the surface point \(X\) is blocked by another point \(\tilde{X}\)—and attached shadows—occurring because the surface faces away from the light source, \(\langle n, \omega_i \rangle \leq 0\). \cite{DaumDudek98} reconstruct surface height based on cast shadows from known point light sources. They derive several constraints based on geometric reasoning such as 'if point \(X\) is shadowed by \(\tilde{X}\), then all surface points on the connecting line must lie below the ray from \(X\) to \(\tilde{X}\)'.

\cite{Chandrakeretal07} use similar constraints for 3D surface reconstruction in their ShadowCuts paper. The approach requires at least four input images of a Lambertian scene illuminated from a combination of four or more light sources, so that each surface point is observed under at least three different, non-shadowed illuminations. Starting from the observation that light source visibility is piecewise constant over a surface, they propose an efficient graph cut-based approach, which determines the per-pixel light source visibility labeling that minimizes the photometric stereo error of the scene. Their variational surface reconstruction approach combines photometric stereo with the constraints given by the shadow labeling.
6.2 General deviations

Kriegman and Belhumeur [2001] instead study attached shadows. They reconstruct the light positions and normals at intersection points of shadow boundaries. Okabe et al. [2009] exploit the binary pattern (shadowed/non-shadowed) at a pixel created by a large sequence of light positions. The similarity between two such patterns is related to the angular difference of the corresponding normals. This insight allows them to recover surface orientation using dimensionality reduction inspired by Sato et al. [2007] (see §7.3 for a detailed description of this paper).

In a real scene, cast and attached shadows will be present at the same time. Sunkavalli et al. [2010] observe that the light source visibility is constant in subregions of the surface. The intensities in each region lie in a three-dimensional subspace as exploited by uncalibrated photometric stereo methods. The overall intensity matrix thus consists of several rank three submatrices which can be obtained using subspace clustering techniques.

6.2 General deviations

6.2.1 Exploiting high inlier to outlier ratios

In contrast to approaches that explicitly treat specific effects, another line of work tries to detect any deviation from the local, Lambertian shading model and then applies photometric stereo only to the remaining intensities. This usually requires many images to ensure that a surface point is observed sufficiently often without shadows or specular behavior.

Such a dense photometric stereo technique is proposed by Wu and Tang [2006]. They formulate a probabilistic image formation model that includes a binary inlier/outlier map as hidden variable. For a single pixel in M images, they compute a set of \( \binom{M}{M-1} \) normals and represent their distribution as a covariance matrix \( B \). The algorithm estimates an optimal \( B \) by weighting down observations that are likely to be outliers according to the map. After updating the weights, the whole process is repeated until convergence. Another example of this category is given by Verbiest and Gool [2008] who also employ expectation maximization
but treat the normal map as a hidden variable. Instead of the per-pixel approach of Wu and Tang, they describe the likelihood of outlier intensities with per-image histograms.

Beljan et al. [2012] also consider sets of normals for a single point computed from \( \binom{M}{3} \) subsets of all available images, but do so in a multi-view setting. The simplified example in Figure 6.2 shows the spread of these normals at different point. For each image triplet, Beljan et al. [2012] compute the set of inlier images whose observations are consistent with the hypothesized normal. Instead of studying the distribution of these normal hypotheses, they apply RANSAC to find the one that has the largest support. The number of inliers then provides a cue to decide whether a voxel belongs to the true surface or not. This leads to good results for the normals but turns out to be not very discriminative in terms of absolute geometry.
6.2. General deviations

6.2.2 Matrix corruptions

A slightly different perspective on photometric stereo with outliers is provided by techniques that analyze the observation matrix for errors and missing entries. Wu et al. [2010] interpret the problem as a low rank matrix recovery task in the presence of sparse corruptions. More formally, they augment the matrix decomposition in Equation 4.2 with an error matrix $E$ that contains all outliers due to shadows or specularities:

$$L = T \cdot S + E =: K + E. \quad (6.3)$$

As we have seen before, the rank of $T \cdot S$ is at most three for a Lambertian surface. Under the assumption that outliers occur infrequently, $E$ is sparse and the task becomes

$$\arg \min_{K,E} [\text{rank}(K) + \alpha \|E\|_0], \quad \text{s.t. } L = K + E \quad (6.4)$$

where $\|\cdot\|_0$ is the number of non-zero entries—and thus not a true norm. Wu et al. [2010] replace this formulation with a convex optimization problem

$$\arg \min_{K,E} [\|K\|_* + \alpha \|E\|_1], \quad \text{s.t. } L = K + E; \quad (6.5)$$

which can be minimized iteratively using Lagrange multipliers with additional penalty terms (Augmented Lagrange Multiplier method, cf. Bertsekas and Rheinboldt [2014]). If the locations of shadows are known beforehand, they enforce the constraints only on the remaining entries.

Once $K$ is recovered, standard tools such as singular value decomposition can be applied to obtain the factorization $S, T$. Other approaches in this field of research, e.g. by Okatani and Deguchi [2007], Okatani et al. [2011], Eriksson and van den Hengel [2010], directly optimize for the decomposition

$$\arg \min_{S,T} \|H \odot (L - T \cdot S)\| \quad (6.6)$$

with a known matrix $H$ encoding missing entries and $\odot$ being the component-wise multiplication. Such a technique can also be applied to cope with outliers in calibrated photometric stereo as, for example, shown by Ikehata et al. [2012].
Shape reconstruction of Lambertian surfaces under unknown illumination is possible as discussed in Chapter 4. Techniques that handle more complex BRDFs are presented in Chapter 5 but require known light sources or special capture setups. Reconstructing objects with complex BRDFs under unknown illumination is a much harder problem because both constituents interact in a complex way to form the observed appearance.

To a certain extent, ideas from the previous chapters can be applied to this setting with some modifications. For example, §7.1 extends the idea of diffuse and specular separation in §5.1 by applying one of the techniques for general light sources presented in §4.2 to the diffuse component. Similarly, §7.2 is related to the handling of parametrized BRDF models as discussed in §5.2. Incorporating the unknown lighting in such a formulation adds, however, several degrees of freedom that make a robust solution non-trivial. A more general approach is to directly study similarities in the sequence of intensity values at single pixels. Works that exploit such so called appearance profiles are discussed in §7.3.
7.1. Diffuse specular separation

Low-rank factorizations of the intensity matrix as in uncalibrated photometric stereo are not easily possible in the case of non-Lambertian reflectance. A straightforward way to deal with this problem is to separate the diffuse and specular contributions using a technique such as the chromaticity based method by Tan and Ikeuchi [2005] shown in Figure 7.1. Applying one of the methods in Chapter 4 to the diffuse component then yields the desired normal field.

Tan et al. [2007] propose such a system and show that the generalized bas relief ambiguity can be resolved from the additional information present in the specular component. The approach is based on isotropy and reciprocity constraints derived from two or more images, but Tan and Zickler [2009], Tan et al. [2011] show how this can even be reduced to one. Wu and Tan [2013] use a similar strategy of diffuse-specular separation and find the optimal bas relief transform through a constraint on the structure of the BRDF. This requires objects with sufficient distribution of normal directions, e.g. a sphere, and an isotropic BRDF that exhibits a symmetry around the halfway vector.
7.2 Parametric models

A different approach is to define a mathematical model of image formation that encompasses all unknowns and to directly minimize a suitable error, e.g. image intensity differences

$$\arg \min_{N, \rho, L_s} \| I - M(N, \rho, L_s) \|.$$  \hspace{1cm} (7.1)

This is similar in spirit to §5.2 but adds the incoming light distribution $L_s$ to the unknowns. Such an approach comes at the same costs as discussed there: complex optimization problem, susceptibility to local optima, and dependence on a parametric model that might not represent all real-world scenarios.

An example is the technique by Georghiades [2003]. It uses a simplified Torrance-Sparrow BRDF (Torrance and Sparrow [1967]) and a single distant light source. The optimization is, however, restricted to a single BRDF and only allows the diffuse albedo to vary spatially. He is able to show that the generalized bas relief ambiguity is resolved for specular surfaces in almost all cases if at least four images are captured.

7.3 Appearance profiles

We have seen another way to approach photometric reconstruction problems in general in Chapter 5: symmetries and general properties of BRDFs. In the setting discussed here, reasoning about symmetries suffers from unconstrained lighting. A much more common strategy is to exploit general relations between observation vectors or appearance profiles that encode the temporal appearance variation of a single pixel, i.e. columns in Equation 4.1. One example is the work of Shi et al. [2010], that uses appearance profiles to introduce a self-calibrating photometric stereo technique (see the detailed description of the paper in §4.3).

A recurring concept in this context is orientation consistency which states that points with similar normals have similar appearance. Assuming an orthographic camera, i.e. constant $\omega_{\text{out}} = \nu$ for all surface
7.3. Appearance profiles

points, the radiance of a point $X$ with normal $n$ is given by

$$f^{-1}(I) = L = \int \rho(v, \omega_{\text{in}}) L_s(\omega_{\text{in}}) \langle n, \omega_{\text{in}} \rangle \, d\omega_{\text{in}} \quad (7.2)$$

according to Equation 2.2. The right hand side depends only on the normal and BRDF but not on the 3D position. Thus, for a point $\tilde{X}$ on the surface with the same normal and BRDF as $X$ the radiance is the same. Moreover, if the radiance is the same, the pixel intensities are equal, too—even for a non-linear response curve $f$.

[Hertzmann and Seitz 2003] place one or more example objects of known geometry, usually a sphere, in the scene to obtain reference profiles

$$W_i = \begin{pmatrix} I_{1,i,1} & \cdots & I_{1,i,k} \\ \vdots & \ddots & \vdots \\ I_{M,i,1} & \cdots & I_{M,i,k} \end{pmatrix} \quad (7.3)$$
of $k$ basis BRDFs. For a pixel $j$ on the target object, they match its appearance profile

$$w_j := \begin{pmatrix} I_{1,j} \\ \vdots \\ I_{M,j} \end{pmatrix}^T \quad (7.4)$$

against linear combinations $W_i m$ of the reference profiles to find a corresponding point

$$\arg \min_i \| W_i m_j - w_j \|^2, \quad m_j := W_i^T w_j \quad (7.5)$$
on the reference. Based on orientation consistency, both points are then likely to have the same normal. Since normals for the reference object are known, they can simply be transferred to the target, see Figure 7.2.

This approach has the added advantage to require only a minimal amount of calibration. [Ackermann et al. 2010] remove the need for an explicit example object by replacing it with a partial reconstruction of the scene using multi-view stereo techniques. This bootstrapping allows them to apply the example-based photometric stereo idea even in scenarios where physical access to the scene is difficult such as for Internet images.

If the light source positions vary smoothly, derivatives of the appearance profiles can be computed. [Koppal and Narasimhan 2006] obtain
the extrema of appearance profiles from such derivatives under an otherwise unknown light source. They show that most of these extrema are due to constant constellations of viewing direction, light source, and surface normal. This insight makes the set of extrema locations well-suited features for the clustering of surface orientations independent of the BRDF. This process alone, however, cannot recover the relationship between iso-normal clusters, e.g. one facing left and one facing right. Other techniques are needed to assign absolute orientation to the clusters. Those can, however, benefit from the initial clustering as prior knowledge.

Sato et al. [2007] look at appearance profiles as a whole and not just at the locations of extrema. They argue that similar profiles—seen as vectors in $M$-dimensional space—correspond to similar normals, which they justify based on the studies of the Torrance-Sparrow BRDF model (Torrance and Sparrow [1967]). Since normals are defined on the unit sphere $S$ in $\mathbb{R}^3$, this implies that the set of profiles $\mathcal{M}$ is actually a two-dimensional manifold in $\mathbb{R}^M$. Thus, there should exist an embedding $\Phi : \mathbb{R}^M \rightarrow \mathbb{R}^3$ with $\Phi(\mathcal{M}) \subset S$ that preserves the intrinsic structure of $\mathcal{M}$, i.e. points that are close in terms of $\mathcal{M}$ correspond to normals that are close on $S$. The notion of closeness on $\mathcal{M}$ is expressed by the geodesic distance, which can be approximated by Euclidean distances along a sequence of neighboring profiles.

Sato et al. [2007] then employ the Isomap non-linear dimensionality reduction technique by Tenenbaum et al. [2000] to obtain the embed-
7.3. Appearance profiles

ding Φ and afterward force the result vectors to have unit length. The
normal map thus recovered contains only relative orientations, e.g. a
global rotation or reflection are not considered, but can be made unique
if additional constraints, such as contours, are applied. The advantages
of this technique are its generality and a certain robustness against
non-linear camera responses. It makes, however, the assumption of a
spatially constant ratio between diffuse and specular contributions and
requires hundreds of images.

Lu et al. [2013] extend the work by Sato et al. [2007] in several
ways. They show empirically—using the MERL (Matusik et al. [2003])
database of measured BRDFs—that the geodesic distance of the profiles
is proportional to the angular difference of normals

\[ d_{\text{geo}}(w_i, w_j) \propto \cos(\langle n_i, n_j \rangle), \quad \text{for } |\langle n_i, n_j \rangle| > 1 - \delta, \quad (7.6) \]

and that the proportionality factor depends only on the material. They
also propose a way to estimate this factor from intensity samples ob-
tained under uniformly distributed, but unknown, point light sources.
In combination this allows to recover the \( P \times P \) matrix \( N^T N \) contain-
ing the dot products of all pairs of normals. Instead of dimensionality
reduction techniques, their recovery step is based on factorizing this
matrix and removing the arising ambiguity by means of the integrabil-
ity constraint.
Most photometric techniques assume a fixed view point. Information about the orientation and reflectance of a surface is recovered from changes in the illumination. This, however, restricts the result to a single perspective and does not allow a reconstruction of a full, e.g. 360°, object model. Furthermore, varying view points provide additional information. For example, observing an object from the front and side lets us estimate its absolute position in the scene and not just the surface orientation. Multiple views also introduce redundancy in surface coverage, which can be exploited to increase robustness.

If the illumination is fixed while the camera moves, we leave the scope defined in §1.1. Nevertheless, §8.1 discusses some stereo algorithms because they can seen as complementary tools to photometric reconstructions as noted by [Horn et al. 1978]. Stereo often performs poorly on surfaces that exhibit only little texture variation. On the other hand, it provides a good estimate of the absolute geometry and not only recovers surface orientation that has to be integrated as in §2.5. Photometric techniques work well even in uniform regions but can suffer from low-frequency distortions. Some of the approaches in later sections will therefore combine both techniques to counter their respective disadvantages.
8.1 Fixed illumination

We will categorize the techniques relying mainly on shading variations according to the magnitude of the assumed motion of the scene components. For small, differential changes in camera and lighting discussed in §8.2, the situation is not too different from a fixed camera and can be addressed by per-pixel perturbations as in optical flow methods. A second strategy is to recover an initial proxy and adjust its shape based on photometric reasoning. This idea also gets applied in §8.3 where camera baselines are much wider and the changes in lighting are not restricted to smooth variations.

8.1 Fixed illumination

Reconstructions from two or more views are commonly addressed by stereo or multi-view stereo methods. The idea is to find properties that are invariant—or almost invariant—to view-point changes such as the color of a Lambertian surface point $X$ under constant illumination. If $X$ projects to pixel $\tau_1(X)$ in the first image and $\tau_2(X)$ in the second, their color will be the same. For a hypothesized point $\tilde{X}$ that does not lie on the surface, the color of its projections $\tau_1(\tilde{X})$ and $\tau_2(\tilde{X})$ will differ in general. Thus, it is possible to detect the correct position by checking the consistency of projections with respect to the invariant. Figure 8.1 illustrates this concept. For increased robustness, small surfaces patches are often used in favor of single points. The image intensities for a projected patch $\mathcal{P}$ in two cameras $i, j$ are then compared using metrics such as the sum of squared differences

$$d_{i,j}(\mathcal{P}) = \sum_{X \in \mathcal{P}} [I_i(\tau_i(X)) - I_j(\tau_j(X))]^2$$

(8.1)

or the normalized cross correlation. For multiple image pairs, a hypothesis is either retained or discarded based on an aggregation of the scores $d_{i,j}$.

Seitz et al. [2006] give a good introduction to the aspects of different multi-view stereo approaches and provide an overview over the literature in this field. We note that almost all techniques assume a Lambertian surface. Exceptions are the works by Jin et al. [2003], Soatto et al. [2003], Jin et al. [2005], which use a rank constraint on the radiance
8.2 Differential motion

Some of the first steps towards multi-view reconstructions that consider shading are connected to the analysis of densely sampled video sequences. Carceroni and Kutulakos [2001] use multiple video streams at calibrated positions to recover the motion, shape, and Phong reflectance parameters (Phong [1975]) of deformable surfaces under known point light sources.

Zhang et al. [2003] introduce a framework that encompasses the similarity of patches seen from multiple cameras and their appearance change caused by varying illumination. The authors recover the direction of a distant light source, the surface normals, and absolute depth—i.e. the distance from a camera center—in an alternating optimization.
A general theory for shape reconstruction from differential object motion is derived by Chandraker et al. [2013]. It extends optical flow to objects of unknown BRDF under arbitrary illumination.

These works rely on relatively dense temporal sampling and smooth object motion in order to linearize the relationship of image intensities $\Delta I = I_i - I_{i+1}$. Thus, it is questionable whether they are also suited for wide-baseline, multi-view setups.

Lim et al. [2005], while also considering video frames as input, do not explicitly exploit this linearization. They initialize a piece-wise planar surface by triangulating tracked scene points. Projecting it into all views yields an intensity matrix, which can be factorized to obtain normals from uncalibrated photometric stereo. The positions of the tracked points allow a disambiguation of the general bas relief transform. Then, the surface is updated by integrating the normals into a depth map. This surface is again projected into all views to create a new intensity matrix. The authors iterate these steps until convergence.

Joshi and Kriegman [2007] follow the same line of thought. They derive a consistency measure for Lambertian surfaces which allows them to estimate a coarse depth map. Projecting this depth map into all images then yields an intensity matrix, which can be factorized. They resolve the ambiguities by comparing the pseudo normals with the normals obtained by differentiating the depth map. Lastly, the normals and depth map are both integrated into a final surface.

Their consistency measure is based on hypothesizing planar patches, which they project into all images and form an intensity matrix for each patch. As we have seen in §2.2, this matrix has rank three for a Lambertian surface. If the patch was at an incorrect depth, the error of a rank three approximation to this matrix should be high. Minimizing this error combined with a straightforward smoothing does not yield high quality depth maps, but is sufficient as initialization.

We observe that these techniques alternate between a depth estimation and a normal estimation step. This is because photometric stereo requires known pixel correspondences to associate the intensities to a surface point. In traditional photometric techniques, these are trivially defined by the fixed view point assumption.
8.3 Wide baseline

The technique by Joshi and Kriegman [2007] was used in a video-based reconstruction setup but does not explicitly exploit differential motion. This example leads us to another group of algorithms that do not make such strong assumptions about the changes in object, camera, or light positions. We summarize the setups of some these in Table 8.1. They allow for wider baselines between views and increased freedom in light source movement.

Another difference is that the approaches discussed before employ a local surface representation, i.e. defined per view. We will now look at methods that rely on a global geometric model of the surface. This has the advantage of providing readily available occlusion information while the surface is deformed by the optimization.

Voxel carving

Weber et al. [2002] use a voxel representation and consider objects on a turntable. They carve away voxels for which the predictions of the Lambertian model—given light source positions—disagree with the recorded intensities even for the best fitting normal. The estimated albedo and normals can then be used to render the scene with novel lighting.

Yang et al. [2003] develop a consistency measure based on the behavior in color space of a surface point observed from multiple views: it varies linearly from the diffuse color to the color of the light source. In contrast to other voxel-based formulations, the approach additionally employs a smoothness constraint. This is not straightforward to define for a voxel representation, and the authors resort to extending the disparity gradient introduce by Burt and Julesz [1980], which is originally defined between two images only. Their method does not recover normals and actually considers mostly a setting with fixed illumination.

Treuille et al. [2004] also reconstruct a voxel representation. They consider multiple light directions and require one or more example objects in the scene similar to Hertzmann and Seitz [2005]. Their consistency measure compares appearance profiles for a candidate voxel
<table>
<thead>
<tr>
<th>Reference</th>
<th>Camera</th>
<th>Turntable</th>
<th>Illumination</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weber et al. [2002]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>P Y (3)</td>
</tr>
<tr>
<td>Yang et al. [2003]</td>
<td>Y (8)</td>
<td>Y/N</td>
<td>N</td>
<td>A Y (4) equal light color</td>
</tr>
<tr>
<td>Treuille et al. [2004]</td>
<td>Y</td>
<td>Y</td>
<td>N (Y\textsuperscript{a})</td>
<td>P Y (3) example object</td>
</tr>
<tr>
<td>Ackermann et al. [2014]</td>
<td>N</td>
<td>N</td>
<td>N P</td>
<td>N example object</td>
</tr>
<tr>
<td>Vogiatzis et al. [2006]</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>P Y (3)</td>
</tr>
<tr>
<td>Hernandez et al. [2008]</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>P Y (3)</td>
</tr>
<tr>
<td>Birkbeck et al. [2006]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>P Y (2)</td>
</tr>
<tr>
<td>Yoshiyasu and Yamazaki [2011]</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>P N attached flash</td>
</tr>
<tr>
<td>Wu et al. [2011]</td>
<td>Y (8-20)</td>
<td>N</td>
<td>N</td>
<td>A/P Y dome</td>
</tr>
<tr>
<td>Park et al. [2013]</td>
<td>Y (2)</td>
<td>Y</td>
<td>N</td>
<td>P Y (13)</td>
</tr>
<tr>
<td>Yoon et al. [2010]</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y (?)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Lighting is known for experiments that have shadow handling enabled

Table 8.1: A summary of works on wide baseline multi-view methods and the setups used in their practical implementation. The assumptions and theoretical models of the actual algorithms are often less restrictive and could probably be applied in more general settings, too. We list the number of cameras and whether they are fixed or move freely. In the case of a single, fixed camera multiple views are usually generated by rotating the object on a turntable (Y=yes, N=no). For the illumination, we list whether it is known, a point (P) or area (A) light, and if it physically moved between shots. We also mark the light source as unknown if it is estimated by the technique along with the object shape, i.e. without separate calibration. The number of lights in a turntable setting can also mean that a single light source was moved to a different location after a full or a third of a turn.
Figure 8.2: Illustrating the voxel-based reconstruction by Treuille et al. [2004].

**Left:** One of the input images taken from multiple viewpoints and with varying illumination. The inset shows the reference sphere that is photographed along with the target object under the same lighting conditions. **Middle:** Due to the discrete voxel grid, the actual surface geometry cannot represent fine details. **Right:** The detail information encoded by the normals can be exploited for rendering.

against those on the reference objects. Figure 8.2 illustrates that the reconstructed normals contain much more details than the coarse voxel approximation. This information can, however, only be exploited during rendering because normals and voxels are not fused into a single surface representation. A further restriction concerns the camera positions, which need to be separated from the scene by a hyperplane to ensure the correct order during voxel processing. These restrictions are addressed by Ackermann et al. [2014] who use a continuous, local representation instead of a global voxel model. Their approach might, however, lead to slight inconsistencies between views.

**Mesh deformation**

Vogiatzis et al. [2006] and Hernandez et al. [2008] represent the global model as a triangle mesh with attached normals, which circumvents the drawback of the method by Treuille et al. [2004]. The approach in both works assumes Lambertian surfaces and treats specular reflections as outliers. It relies heavily on object silhouettes to recover the camera parameters, light directions $D_k$, and an initialization of the mesh $\mathcal{P}$. The mesh is then projected into the images to obtain the intensities $I_{i,k}$ per face and light direction. From these, normals $N$ are computed.
with a standard photometric stereo approach
\[
\arg \min_N E_{\text{photo}}(N; \mathcal{P}) = \arg \min_N \sum_i \sum_k \left( \langle n_i, D_k \rangle - I_{i,k} \right)^2.
\] (8.2)

This step is alternated with a vertex displacement that forces the face normals \( \tilde{\mathbf{N}} \) defined by the mesh vertices \( \mathcal{P} \) to conform with the reconstructed normals \( \mathbf{N} \)
\[
\arg \min_P E_{\text{disp}}(\mathcal{P}; N) = \arg \min_P \sum_i \| \tilde{n}_i(P) - n_i \|^2.
\] (8.3)

[Birkbeck et al. 2006] also filter out specular reflections during shape recovery but fit a specular Blinn-Phong model (Blinn 1977) once the final geometry is available. Thus, the final output of this algorithm is not just the object shape but also reflectance parameters. To achieve this, they obtain the input images and the required light calibration from a very controlled capture system. All three techniques can draw on initializations that are already quite close to the true surface.

[Yoshiyasu and Yamazaki 2011] show that a similar approach can succeed with a much simpler initialization, e.g. a set of spheres approximating the object as in Figure 8.3. It also operates on mesh vertices, but converts them into an implicit surface and back to a mesh during the optimization. This allows to handle a greater variability in topology—implicit surfaces can, for example, self-intersect. Similar to [Hernandez et al. 2008], their technique requires silhouettes to be extracted. In fact, if the mesh deformation is constrained solely by silhouettes—ignoring the photometric term completely—it already provides very good results. The photometric cue is much weaker and only serves to add some details. Such combinations are also used in dynamic performance capture systems as proposed by [Ahmed et al. 2008, Vlasic et al. 2009].

While these approaches rely on silhouettes, the general idea of deforming an initial reconstruction based on photometric information can also be applied in other settings. [Wu et al. 2011] reconstruct an initial mesh using multi-view stereo and refine it with normals from uncalibrated photometric stereo. Another way of initialization is to assume a mostly planar target as suggested by [Ruijters and Klein 2009]. They represent geometry variations as a height field and recover spatially
Multi-view settings

Figure 8.3: Overview of the multi-view reconstruction approach by Yoshiyasu and Yamazaki [2011]. A basic mesh gets deformed iteratively to fulfill silhouette and photometric constraints (©2011 IEEE. Reprinted, with permission, from Yusuke Yoshiyasu and Nobutoshi Yamazaki. Topology-adaptive multi-view photometric stereo. In IEEE Conference on Computer Vision and Pattern Recognition [2011]).

Varying BRDFs similar to the approach by Goldman et al. [2005]—but in a multi-view setting. This approach even incorporates interreflection effects in the micro-structure and is related to BTF (Bidirectional Texture Function) acquisition techniques.

A disadvantage of approaches that project vertices into images is that the mesh resolution is not necessarily related to image resolution. Either a huge number of vertices has to be used or details in the images might be lost. Park et al. [2013] therefore propose a pipeline that uses mesh parametrization to define warpings from images onto a planar mesh representation. Thus, fine detail—in the form of a displacement map—can be recovered even with few mesh faces. The base mesh is obtained by merging multi-view stereo reconstructions from all views. Again, specular reflectance is treated as an outlier and the underlying photometric stereo technique is the one developed by Hayakawa [1994].

Parametrized models

Similar to § 5.2 and § 7.2, heavily model-based approaches can also be extended to multi-view settings. A very good example is provided by Yoon et al. [2010]. Their generative model consists of a finite number of distant light sources and shadow maps, the surface represented as a level set, reflectance parameters for the Blinn-Phong model (Blinn [1977]), and a set of pinhole cameras. This allows to render synthetic images of the scene given the current parameters.
Visibility is handled by projecting the global surface into each image. Furthermore, their technique requires a model of the background scene to prevent the surface from shrinking to an empty set during the optimization. They also introduce an auxiliary normal field to decouple the appearance computation from the surface gradient for increased stability. The final optimization minimizes the error of all images compared to the renderings from the current parameters. It has to be performed in an alternating fashion—like in most other works—because of its complexity and the coupling of shape and reflectance in the image formation.

Special setups

Like single view photometric stereo the multi-view case often requires quite elaborate setups.

Lu and Little [1995] present a solution in a rather special setting: the light source and camera are co-located and the object spins around a known axis. They derive surface orientations pointing towards the light from intensity maxima and then exploit the known rotation angle of the object to track them. The tracked points allow the reconstruction of a one-dimensional slice of the BRDF. Once this is known, the surface orientation of all other surface points can be computed from its inverse.

Zhou et al. [2013] use a ring light system with 72 LEDs to capture 30 images under varying illumination for each camera position. This allows them to cope with arbitrary isotropic BRDFs by applying the technique of Alldrin and Kriegman [2007] in each view. The resulting iso-depth contours are then associated with sparse structure from motion points to obtain absolute depth values. The authors propagate depth along contours from each view to merge them into a globally consistent model. Once the geometry is acquired, a set of basis BRDFs and their mixing weights can be estimated because illumination is known from a calibration step.

Tunwattanapong et al. [2013] also require multiple light settings for each view. The incoming radiance distributions in such a sequence are shaped as spherical harmonics and created by an rotating arc of LEDs. This setup allows them to recover reflectance parameters in
addition to surface orientation. The results from each camera position are then fused with a multi-view stereo reconstruction in a final surface optimization, which again alternates between several stages.

Less restrictive than these systems is the setup proposed by Pater-son et al. [2005]. It consists of a calibration target and a flash light that is rigidly attached to the camera. They assume an almost planar target which allows them to establish pixel correspondences by warping the images. The recovery of normals and parameters of a modified Torrance-Sparrow BRDF model is alternated with a re-estimation of this warp.

Another technique that relies on a special capture setup to acquire series of images in each of multiple cameras is presented by Schuster [2010]. He proposes a photo-consistency measure based on the fitting error of a Cook Torrance BRDF at each hypothetical voxel and extracts a surface using graph cuts (Boykov and Kolmogorov 2003, Lempitsky and Boykov 2007). As initialization, he constructs the visual hull to obtain visibility information and computes normals based on Helmholtz stereopsis. This requires the light and camera to swap positions, which is approximated in this case by a hemispherical dome of 151 cameras equipped with flash lights.

A similar setup, but augmented with several projectors for structured light acquisition, is used by Weinmann et al. [2012]. They acquire multi-camera, multi-light image sequences of an object rotating on a turntable. The surface is then extracted from an error function that combines structured light consistency and Helmholtz stereopsis into a single variational formulation.

Finally, the technique by Aliaga and Xu [2010] covers a whole range of single or multiple camera and projector setups. They combine Lambertian photometric stereo and structured light to recover oriented 3D points of the object and the geometric calibration of the system. Even for a single camera, they are able to formulate multi-view constraints on the depth of the points. That is possible by exploiting the fact that a projector can act as a virtual camera if pixel correspondences are established.
8.3. Wide baseline

Over-complete representations

We observe that several of the discussed multi-view photometric stereo methods (e.g. Joshi and Kriegman [2007], Hernandez et al. [2008], Yoshiyasu and Yamazaki [2011]) can be modeled by up to three general error terms

$$E_1(X), \ E_2(N), \ E_3(X, N).$$

(8.4)

The first one is tied to multi-view constraints such as silhouettes. The second one finds optimal normals based on the current depth estimate. It is associated to photometric constraints, i.e. surface appearance. The third one updates the geometry based on these normal estimates. These terms are either combined into a sum or optimized one after the other.

Such a scheme uses a two-fold representation of the shape that includes a 3D proxy and an auxiliary normal field. It is an overcomplete representation since in the ideal case the normal field should be defined by the derivative of the geometry. This is sometimes described as a convenient formulation for optimization or as a necessity for stability, cf. Jin et al. [2004], Yoon et al. [2010]. It has the additional advantage to allow different resolutions in the two representations as exploited by Park et al. [2013]. Often, the representations are coupled by a soft constraint in the spirit of $$E_3 = \|\partial X - N\|$$. A popular method to perform that merging step is presented in the work of Nehab et al. [2005]. In the future, a better understanding of the benefits and drawbacks of such an overcomplete formulation would be an interesting research topic.
The previous chapters show that many of the state of the art approaches draw on ideas and methods that have been developed ten or twenty years ago, e.g. matrix factorization or reflectance symmetries. We observe that today, more focus is put on integrating these ideas into ready to use systems and increasing their applicability through combination with other techniques. Two important aspects are the robustness with regard to non-ideal input images and the removal of requirements imposed on the capture setup. Thus, images downloaded from the Internet provide an interesting testbed for new developments. We emphasize this relevance by discussing some of the advances in computer vision related to Internet data §9.1 and reconstructions under natural illumination §9.2.

The scope of this chapter is slightly broadened and incorporates exemplary techniques from areas such as intrinsic image factorization or inverse rendering. This is justified because ideas from other fields on how to address some of the general challenges in uncontrolled settings, e.g. radiometric camera calibration, hopefully provide insights for photometric reconstructions as well.
9.1 Internet data

9.1.1 Image collections

Snavely et al. [2006] show that it is possible to apply robust structure from motion to images from online photo sharing sites, e.g. Flickr. They recover the geometric camera parameters and a sparse set of 3D feature points. This allows a user to explore the scene by 3D browsing of the images and provides a novel way to communicate the impressions of one or several observers. Tompkin et al. [2012] extend 3D browsing to video sequences.

Since known camera parameters are a prerequisite for many reconstruction approaches, Snavely et al. also paved the way for works in that area. For example, Goesele et al. [2007] present a multi-view stereo approach for community photo collections. It relies heavily on the selection of suitable images to achieve robustness against scale differences or occluding clutter. Furukawa and Ponce [2010] achieve robustness through several filter steps that reason about visibility and consistency of 3D patches. These approaches recover the full scene geometry instead of a sparse feature set and can convey the 3D impression of the scene even better. Goesele et al. [2010] exploit this as proxy geometry in an image-based rendering system to achieve a more convincing browsing experience in the presence of unreliable data. Agarwal et al. [2009] and Frahm et al. [2010] extend the reconstruction approaches to whole city areas using clustering techniques.

These techniques focus on the reconstruction of the 3D geometry and try to be robust against appearance changes. Such changes, e.g. due to varying illumination or viewpoint, are at the heart of photometric techniques and convey additional information. Garg et al. [2009] extend results from Belhumeur and Kriegman [1998] about the space of possible images for a given scene to photo collections. They project all images onto a 3D model and factorize the matrix that arises from stacking intensities at each vertex. The factorization can be interpreted as basis images for the scene that capture variation along different axes. While these axes often correspond to a certain meaning, e.g. mean image and shading variations in the Lambertian case, there are no clear
semantics for decompositions in more complex scenes.

A more usable separation of appearance for uncontrolled scenes is achieved by [Haber et al. 2009]. They build on a 3D model as produced by [Goesele et al. 2007] and recover the reflectance—encoded as mixing weights per surface point with respect to a selection of Cook-Torrance BRDFs—and illumination $L_{s,i}$ in each view of an image collection. A linearization

$$\rho = \sum_k \rho_k \Psi_k, \quad (9.1)$$

$$L_{s,i} = \sum_l L_{s,i,l} \Psi_l \quad (9.2)$$

based on Haar-Wavelets lets them formulate the outgoing radiance as

$$L_i = \sum_{k,l} \rho_k L_{s,i,l} \int \Psi_k(\omega) \Psi_l(\omega) \, d\omega = (\rho_1, \ldots) \cdot B \cdot \left( L_{s,i,1} \right) \quad (9.3)$$

The final optimization over all surface points $X_j$ and images

$$\arg \min_{j,i} \sum (I_i(X_j) - L_i(X_j))^2 \quad (9.4)$$

alternates between both unknowns $\rho, L_s$ and yields a decomposition that gives plausible relighting results. Figure 9.1 compares the rendering results to one of the input images which was not used during reconstruction.

[Diaz and Sturm 2011] also use a 3D model to estimate illumination, but additionally consider the camera response function. They represent this function as a linear combination of the EMOR basis (Empirical Model of Response: [Grossberg and Nayar 2003, Grossberg and Nayar 2004])

$$f = h_0 + \sum_{i=1} \alpha_i h_i \quad (9.5)$$

and include the coefficients $\alpha_i$ in the lighting optimization. While this work recovers only the diffuse albedo, it is one of the few that actually take non-linear response curves beyond a straightforward gamma correction into account.
9.1. Internet data

Figure 9.1: Haber et al. [2009] first recover the geometry, e.g. using multi-view stereo, and then estimate an environment map (left) and the reflectance properties. Combining these constituents, they achieve synthetic renderings (right) that are quite similar to the actually observed image intensities (middle). This approach works even for uncontrolled image collections as found on the Internet and can be used for relighting or appearance editing (©2009 IEEE. Reprinted, with permission, from Tom Haber, Christian Fuchs, Philippe Bekaert, Hans-Peter Seidel, Michael Goesele, and Hendrik P.A. Lensch. Relighting objects from image collections. In IEEE Conference on Computer Vision and Pattern Recognition [2009], pages 627–634).

9.1.2 Webcam data

Another growing source of Internet data are video sequences, e.g. from webcams. Jacobs et al. [2007a] collect images from hundreds of webcams and discover that the coordinates of a principal component analysis (PCA) for different cameras behave similarly over time. This allows them to define a canonical basis and assign pixel labels with a common meaning in all cameras. In subsequent works, Jacobs et al. [2007b, 2010, 2013a,b] are able to estimate the location of an unknown camera by comparing PCA coordinates with those of known webcams. The depth of individual scene points can be recovered from correlations induced by cloud shadows. Additionally, the movement of the clouds themselves yields cues to predict vanishing points and thus calibrate the camera geometrically. Lalonde et al. [2008, 2010] demonstrate how the sun and sky illumination can be exploited for such a calibration as well. They relate azimuth and zenith angles in a geo-reference coordinate system to pixel coordinates in the camera. Inserting this transformation into the analytical sky model presented by Perez et al. [1993] yields a function that predicts radiance for sky pixels. The camera calibration is
then found by an optimization that compares the predicted radiance values to observed pixel intensities.

Apart from calibration and geo-location, webcams are also used to decompose scene appearance and enable editing of its components. Sunkavalli et al. [2007] analyze the intensity profiles of pixels in dense video to identify transitions from shadow to sunlight. They find that profiles within a scene are similar up to scale and a shift in time. This allows them to define time-varying basis curves for the sun and sky contribution. The corresponding coefficients with respect to this basis are recovered similar to Lawrence et al. [2006] and constitute a decomposition into ambient light, albedo, and sun reflectance. Sunkavalli et al. [2008] additionally recover partial surface normals—projected on the plane of solar movement—based on color changes. Lalonde et al. [2009] estimate the sky appearance and sun visibility in each frame for a collection of webcams. These cues are used to match the illumination conditions between different cameras and thus enable the transfer of correctly lit objects. The authors also insert synthetic objects by illuminating them according to the parameters of the underlying sky model in each image.

9.1.3 Recovering surface orientation

Only few approaches focus on surface orientation as an important part of scene appearance. Shen and Tan [2009] use a decomposition of illumination as proposed by Basri and Jacobs [2001b] to estimate weather conditions in Internet photo collections. This also includes computation of surface normals, but only at sparse feature points that have been matched between images. Their approach relies on a selection of suitable features with sufficiently Lambertian behavior and support in close-by views.

Ackermann et al. [2012] aim for a more complete decomposition of scene appearance. They rely on a fixed camera position as provided by Internet webcams. At each pixel $j$, the direct sun light interacts with a mixture of basis materials $\rho_k$ modulated by a binary shadow function $\xi$. The sky is modeled by a per-pixel additive contribution $\beta$ which gives
9.1. Internet data

Figure 9.2: Exemplary results for the webcam-based reconstruction by Ackermann et al. [2012]. Left to right: One of the input images, the recovered normal map, three basis BRDFs below their respective material maps, and a synthetic rendering for a novel sun position (©2012 IEEE. Reprinted, with permission, from Jens Ackermann, Fabian Langguth, Simon Fuhrmann, and Michael Goesele. Photometric stereo for outdoor webcams. In IEEE Conference on Computer Vision and Pattern Recognition [2012], pages 262–269).

an overall radiance of

\[ L_{i,j} = l_i \left( \sum_k \gamma_{j,k} \rho_k(n_j, \omega_i) \xi_{i,j} + \beta_j \right) \]  

(9.6)

for relative sun intensity \( l_i \) and sun direction \( \omega_i \). The non-linear dependence within this large set of parameters, including normals \( n_j \), yields a complex optimization problem. The authors decompose it into three steps that each keep some of the variables fixed. The final result is shown in Figure 9.2 and yields a decomposition that can be used for relighting. This work relies heavily on a preprocessing pipeline that selects suitable images from a webcam stream and calibrates the camera.

In a similar setting, Abrams et al. [2012] recover dense surface normals, albedo, and the radiometric camera calibration from outdoor webcams. They exploit the known sun position with respect to a georeferenced coordinate system in order to arrive at a traditional photometric stereo problem for Lambertian surfaces. This task becomes more complex through an ambient term and the non-linear response, which is represented as a combination of EMOR (Grossberg and Nayar 2003, Grossberg and Nayar 2004) basis curves. Abrams et al. [2012] split the corresponding optimization into two alternating steps and approximate each of them with a linear subproblem which can be solved efficiently. This allows them to use hundreds of images compared to...
the careful selection of about 50 images by Ackermann et al. [2012].
The increase in data gives better robustness to outliers that are not captured by their image formation model.

More recently, Shan et al. [2013] have applied a related outdoor lighting model for photometric reconstruction on community photo collections. They first recover a mesh using a pipeline of structure from motion and multi-view stereo as described in §9.1.1. Then, they estimate the sun direction and sky contributions in each image and finally compute the diffuse albedo and normal at each vertex in a non-linear optimization. The evaluation is performed in the form of a so called Visual Turing Test, i.e. renderings and original images are presented to humans who have to specify which one is more realistic. Several low resolution images actually pass that test. Visual artifacts and not reconstructed content such as people, however, reveal the limitations of current approaches towards relightable appearance reconstruction on Internet data.

Finally, Jung et al. [2015] perform photometric stereo using a handful of outdoor images of a scene captured on a single sunny day by a static camera. Their key observation is that modelling the complete sky as light source (using a variant of an analytic sky model by Preetham et al. 1999) allows to overcome the restriction that the movement of the sun on a single day contains not enough variance to recover full normals. Using their analytic sky model and a set of 1000 normal hypotheses, they can compute 1000 corresponding radiance profile candidates per pixel (up to a scaling factor), and compute their correlation with the observed values in the images captured at different times of day. Per-pixel surface normals are then assigned using an MRF-based optimization that provides regularization and allows to fill holes at pixels that received too few valid observations.

9.2 Natural illumination

Compared to traditional approaches that require laboratory conditions, Internet images are at least two levels more challenging. It makes sense to also think about in-between settings, e.g. a controlled outdoor
dataset, which are not yet solved entirely.

Sato and Ikeuchi [1994a, 1995] apply their separation of diffuse and specular components based on illuminant color (Sato and Ikeuchi 1994b) to outdoor images under clear skies. They normalize image regions to simulate a uniform albedo and obtain its value from the brightest pixels in the image sequence. Then, the angle between the normal and the solar plane is obtained at each pixel simply as its maximal intensity divided by the albedo. Narasimhan et al. [2002] study appearance variation due to weather changes in long sequences of controlled outdoor data. Atmospheric scattering which depends on the distance of the scene point from the camera allows them to compute approximate depth for each pixel.

Hung et al. [2015] capture the environmental illumination from a mirror sphere. Assuming a calibrated camera, they discretize the rendering equation as

\[ I = \sum_j \rho c_j \langle n, \omega_j \rangle, \]

where \( c_j \) is the light intensity from direction \( \omega_j \). For each surface point, they compute the pair \((n, \rho)\) that best explains the observations \( I_i \) in multiple images according to Equation 9.7. These ideas are closely related to the work by Yu et al. [2013], who also use a mirror sphere and discretize the illumination into a set of point light sources. They assume a Lambertian reflectance, treat specularities as outliers, and employ a simple heuristic to estimate self occlusion. Their results on outdoor images expose problems due to insufficient variation of light directions. This might be one of the reasons why most current work on outdoor reconstructions, e.g. Oxholm and Nishino [2014], is focused on fixed illumination settings. Especially techniques that decompose the illumination and reflectance in images of a known shape, such as presented by Dong et al. [2014b], show some promising advances.

If the whole light field of the scene is known, it can be decomposed into a basis of spherical harmonics. Ramamoorthi and Hanrahan [2001] show that those frequencies of illumination and reflectance that are also present in the outgoing light field—which they interpret as a spherical convolution—can be recovered from its coefficients. This task is much harder if only parts of the light field, e.g. a single image, is available.
and requires additional constraints or regularizing assumptions. For example, any image of a sphere can be explained not only by the true illumination and BRDF, but also by a perfect mirror illuminated with that image.

Experiments conducted by [Fleming et al. 2003] indicate that humans are able to match reflectance properties independent of illumination and that they do so based on prior knowledge about the behavior of natural illumination. Therefore, they exploit the statistics of natural illumination to define a prior on the possible lighting $L_s$. Their idea is to marginalize the posterior $p(L_s, R|I) \propto p(I|L_s, R)p(R)p(L_s)$ over the lighting to obtain the probability $p(R|I)$ of the reflectance $R$ given an image $I$. Computing the mean of this distribution amounts to selecting a reflectance that not only explains the image for a single illumination, but for all illuminations—according to their probability. [Lombardi and Nishino 2012] also consider objects of known shape but additionally recover illumination explicitly.

[Laffont et al. 2012] reconstruct a point cloud of a scene using multi-view stereo. This serves as proxy to compute the contributions of the
sun, sky, and indirect lighting based on a measured environment map. Once sun visibility and reflectance are computed for each 3D point, all this information is projected into the image and propagated over all pixels. Laffont et al. [2013] compute an intrinsic image decomposition but allow for varying illumination between images, which makes the technique applicable to Internet photo collections. They observe that one of the main problems with that kind of data is the unknown camera response. Figure 9.3 lists the components of their decomposition. Lee et al. [2012] also decompose multiple images but do so for a video stream with fixed illumination. Again, their algorithm is supported by a priori known shape information—acquired from a depth camera.
10

Conclusion

In this chapter, we first summarize in §10.1 the core ideas of modern photometric stereo techniques following the main structure of the review. We show how approaches can be judged by their generality with respect to lighting, reflectance and viewpoint distribution as well as by their robustness with respect to different real-world effects. We then discuss future research directions for photometric stereo and related approaches in §10.2 and conclude in §10.3 with thoughts on the future use of photometric stereo.

10.1 Summary

(Lambertian) photometric stereo is a simple yet very powerful technique to reconstruct a surface from a couple of images taking under varying illumination. Its essence can be captured in a handful of formulae as described at the beginning of Chapter 2. Despite its simplicity, photometric stereo yields high quality reconstructions if the scene and the capture setup conform to the basic assumptions underlying the technique. As described in Chapter 3, most of the fundamental concepts of photometric stereo were developed in seminal works during
10.1. Summary

The 1980s. Most of its limitations and shortcomings were also already known at that time. These include, e.g. the need for specific capture settings and calibrated capture equipment as well as the fact that outliers such as shadows or surface interreflections need to be treated very carefully. Most of the research in the last decades focussed on addressing these restrictions in order to generalize photometric stereo to less constrained and less controlled settings. One could therefore ask the question: “When will photometric stereo be solved?”

One answer to this question is that photometric stereo is solved once we can take any arbitrary set of images of an object or scene with large enough input variation and reconstruct a perfect model of its surface orientation from these images. This problem can be decomposed into two almost independent issues—on one hand handling general lighting, reflectance and viewpoints and on the other hand being robust to various effects. We will in the following categorize and discuss these using Figure 10.1 as visual aid.

Research towards general lighting lead to uncalibrated photometric stereo algorithms that are oblivious to the lighting setup. A recurring method in this context is a low-rank matrix factorization as introduced...
in Chapter 4. Simultaneously, research has advanced towards removing the need to know the reflectance properties of the target object. General invariants of the BRDF or reflection geometry define constraints that are exploited by the techniques in Chapter 5. Any algorithm in these two fields is based on assumptions and these might not be fulfilled in practice. Whether this presents a serious problem depends on the specific application. One way to handle them is to use robust methods. Chapter 6 addresses this using techniques for individual error sources and methods that are robust to outliers in general.

Combining the concepts introduced in Chapters 4-6 allows to tackle reconstruction tasks where both, lighting and reflectance, are unknown. One solution to this problem is to model every aspect of light transport explicitly. In addition, Chapter 7 discusses works that exploit orientation consistency and appearance profiles. These techniques are independent of a parametric BRDF or lighting model but require typically a large amount of input data.

Chapter 8 adds a new perspective by considering the camera position as an additional variable. In this chapter, we collect multi-view techniques regardless of their assumptions about illumination and BRDF which leads to a broad variety of approaches. Again, some ideas from the single-view setting can be transferred, e.g. heavily parametrized models, while others are newly introduced such as reconstruction by deformation of an initial proxy mesh.

All these works demonstrate how far the field of photometric reconstructions has evolved since its beginnings. Even very general algorithms in terms of input data are, however, still restricted to capture under controlled settings such as a laboratory. Missing robustness, requirements on the capture setup, or an implicitly assumed camera calibration reduce their applicability. Related fields in computer vision have seen a definite shift towards uncontrolled scenarios and applications in consumer products. Chapter 9 discusses these advances and some of the first approaches to exploit photometric cues in natural environments and even on Internet data.
10.2 Discussion and outlook

We now discuss various aspects that are (in our subjective opinion) important for the future development of photometric stereo.

10.2.1 Broadening the applicability

The advances in generalizing photometric stereo with respect to unknown light or reflectance have revealed valuable insights during the last decades. We believe, however, that significant advancements will in the future most likely not be made along a single of these axes. Instead, it seems more promising to make photometric techniques ready for applications outside the lab. The approaches for general lighting and reflectance in Chapter 7 provide a starting point towards that end. But to achieve a noticeable increase in applicability a greater focus on uncontrolled settings and real-world capture conditions, e.g. a consumer camera with an attached flash unit, is required.

Another trend that can be deduced from Chapter 8 is the development of multi-view techniques. Varying camera positions are also an important step towards less controlled scenarios. Currently, most of these techniques are, however, not as general in the assumed lighting model as those in Chapter 7. Furthermore, the applicability of photometric reconstruction would benefit from research towards better radiometric auto-calibration methods. Images taken with a consumer camera by a casual user are meant (and made by the camera) to look good not to be well suited for reconstruction.

Finally, we see the transition from reconstructing single target objects to complete scenes, e.g. a room, as a worth-while advancement. While some methods, e.g. voxel-based multi-view approaches, should be able to handle that case in theory, they are traditionally evaluated on single objects only. Multiple objects introduce many discontinuities in geometry and appearance which pose novel and interesting challenges.

10.2.2 How general do we have to be?

Many researchers would agree that bringing the image creation models used for photometric stereo closer to reality, e.g. by incorporating
interreflections, while still being able to invert them or to find suitable invariants is a worthwhile goal. It remains, however, unclear how well the model’s assumptions must match reality in order to achieve good results over a wide range of objects. For example, some techniques relying on a Lambertian reflectance model have shown success even when applied to specular materials by treating specular reflections as outliers. Still, a drawback of very general methods might be that they are outperformed by more specialized techniques, which can exploit prior knowledge in their respective target areas.

Another question is how to find the best balance between building explicit models and using more data-driven approaches. For example, appearance profiles do not depend on a strict lighting model but might carry not enough information if only a few input images are available. Parametric models can be a way to work around a lack of input data by enforcing stricter constraints at the cost of reduced flexibility.

Finally, no matter how sophisticated the theory underlying an approach is it will never completely match reality. There will always be a case where observations do not follow all assumptions made by the algorithm. These outliers can misguide the technique to a wrong solution if they are not properly handled, e.g. using ideas from §6.2. The challenge is to find the balance between sufficient robustness and accidentally throwing away too much information.

10.2.3 Combining different techniques

It is highly unlikely that a single algorithm or key idea will be widely or even universally applicable. Even for different techniques from the same general class, e.g. uncalibrated photometric stereo, it often depends on the specific scene which one technique performs best. Thus, combinations of several approaches with complementary advantages and disadvantages seem like a promising direction. For example, the boundaries between purely texture-based stereo approaches and photometric techniques already vanish as discussed in Chapter 8. Another strategy could be to use a different algorithm for different parts of a scene or depending on the object type. Then, the interesting challenge would be to choose the best suited technique automatically.
10.2.4 Data issues

In general, the usefulness of input images is strongly related to the observable shading variations. If the shading does not change, e.g. because the illumination remains fixed with respect to the surface, there is no chance of success for any photometric approach (unless shape from shading techniques with strong assumptions on the surface properties can be employed). This might be one of the reasons why everyday scenarios are rarely considered by photometric stereo methods. Images captured in an office or outdoors will not display sufficient lighting variation unless taken at different times of day. Such issues also arise in a lab setting, but achieving consistent results over a large range of capture conditions and target objects becomes much harder for uncontrolled scenarios.

The opposite of missing input data occurs for some of the approaches in Chapter 9. On Internet images, e.g. months or even years of webcam videos, the vast amount of data would cause many traditional techniques to reach the limits of computational resources such as memory or processing time. It is therefore crucial to reduce the amount of input data by selecting suitable images for reconstruction. Selection strategies not only ensure the computational feasibility but can also increase robustness by removing non-ideal, e.g. under-exposed, images. Estimating the value of an image for the overall reconstruction is not yet considered by most approaches. Instead, they typically assume that all images are equally important and beneficial.

10.3 Looking ahead

Photometric stereo, which can be simply summarized as a way to reconstruct surface orientation from a set of images with varying illumination, has historically been a very important technique. Its core ideas can be applied to a wide variety of domains including surface reconstruction, rendering, relighting as well as industrial inspection. The transition towards more general algorithms and less controlled capture conditions will make it ready for the demands of future application scenarios. We are thus very confident that photometric stereo will also
Conclusion

play a crucial role in the upcoming years and decades, most likely as an integral part of comprehensive and versatile reconstruction systems.
Acknowledgements

We would like to thank Reinhard Klein for his valuable advice and suggestions, including the suggestion to publish this review. We also thank Mate Beljan, Simon Fuhrmann, Stepan Konrad, Arjan Kuijper, Fabian Langguth, Nils Moehrle, Martin Ritz and Andre Stork for their contributions to individual parts of this review. Special thanks go to the various authors of related works that gave us kind permission to use some of their figures and images. Furthermore, we would like to acknowledge the valuable input and constructive criticism provided by the reviewers. We also thank the editor Brian Curless and the publisher James Finlay for their support during the production of this review. Last but not least, we thank our families and friends who strongly supported us on our sometimes difficult path towards writing this review.

This work was supported in part by the European Commission’s Seventh Framework Programme under grant agreements no. ICT-611089 (CR-PLAY), as well as the DFG Emmy Noether fellowship GO 1752/3-1.


References


Yudeog Han, Joon-Young Lee, and In So Kweon. High quality shape from a single RGB-D image under uncalibrated natural illumination. In IEEE International Conference on Computer Vision, 2013.


References


References


References


References


