Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 216

Perspective Shape from Shading
for Phong-type Non-Lambertian Surfaces

Michael Breuß, Oliver Vogel and Joachim Weickert

Saarbrücken 2008

Perspective Shape from Shading
for Phong-type Non-Lambertian Surfaces

Michael Breuß

Saarland University

Faculty of Mathematics and Computer Science

P.O. Box 15 11 50

66041 Saarbrücken

Germany

breuss@mia.uni-saarland.de

Oliver Vogel

Saarland University

Faculty of Mathematics and Computer Science

P.O. Box 15 11 50

66041 Saarbrücken

Germany

vogel@mia.uni-saarland.de

Joachim Weickert

Saarland University

Faculty of Mathematics and Computer Science

P.O. Box 15 11 50

66041 Saarbrücken

Germany

weickert@mia.uni-saarland.de

Abstract

The shape-from-shading (SfS) problem in computer vision is to compute at hand of the shading variation in a given 2-D image the 3-D structure of depicted objects. We introduce an efficient numerical method for a new perspective SfS model for general non-Lambertian surfaces. First, the modelling process is given in detail. The model is based on the perspective model for Lambertian surfaces recently studied by Prados et al., which we extend by use of the Phong reflection model incorporating ambient, diffuse and specular components. The arising partial differential equation (PDE) is a non-linear time-independent Hamilton-Jacobi equation. In order to compute the sought viscosity supersolution of the PDE, we introduce an artificial time into the equation and solve for the steady state. Based on a multi-scale analysis of the PDE, we construct a fully explicit numerical method and elaborate on its stability. In order to achieve fast convergence of the resulting iterative scheme, a coarse-to-fine strategy combined with a sweeping technique is employed. Numerical experiments show the benefits of our approach: While computational times stay reasonable even for quite large images, a substantial qualitative gain can be achieved by use of the new model. Moreover, the computational technique is relatively easy to implement compared to other approaches in the field.

Introduction

Given exactly one grey value image, the Shape from Shading (SfS) problem consists of computing the 3-D depth of depicted objects. It is a classical problem in computer vision with many potential applications; see e.g. [13, 34, 29] and the references therein for an overview.

The camera model. A basic key ingredient of mathematical models for such computer vision problems is the *camera model*, i.e, the mathematical representation of the projection performed when mapping the 3-D world to 2-D images by the camera; see for instance [7, 11, 31] for detailed discussions of this topic. In 'classical' SfS models, the camera is assumed to perform an *orthographic projection* of the scene of interest. Concerning this type of early models, let us especially mention the pioneering works of Horn [9, 10, 12, 13] who was also the first one modelling the SfS task using a *partial differential equation* (PDE).

Modelling of surface properties. A further important modelling issue is concerned with the *reflectance properties* of depicted objects. This determines how light is reflected by the scene and, consequently, how objects are

perceived when acquired by the camera. The most easily accessible model for this purpose is the model of *Lambertian surfaces*. This model describes surfaces with diffuse light reflection: The perceived light intensity of some point on an object surface depends on the angle between light source direction and surface normal in this point; it is independent from the point of view of an observer. For descriptions of various surface models, see e.g. [8]. The assumption of Lambertian surfaces is the classical choice for modelling SfS, e.g., the mentioned models of Horn [13, 12] also incorporate it, and Lambertian surfaces are also used in many modern works [26, 27, 28, 30, 22, 19, 23, 25, 24, 20, 6]. However, Lambertian models exclude important real-world optical phenomena like e.g. specular light reflections.

For orthographic SfS models, there have been some attempts to extend the range of applicability to non-Lambertian surfaces [2, 14]. A drawback of orthographic SfS models is that they usually suffer from ill-posedness, especially in the form of so-called convex-concave ambiguities [6]. Moreover, the orthographic camera model does not yield reconstruction results of convincing quality in most situations, see e.g. the extensive comparisons with results computed using a perspective model in [19, 25].

Perspective Shape from Shading. The problem of ill-posedness can be dealt with successfully by using a *perspective camera model* instead of an orthographic one, in conjunction with the use of a light source given in one point and a so-called light attenuation term as described below; see e.g. [19, 25]. As we refer in this work to a corresponding model for Lambertian surfaces, let us briefly mention some details on it here.

Assuming a point light source at the optical center, the perspective SFS model described in [19] and investigated in [19, 32] amounts to the *Hamilton-Jacobi equation*

$$\frac{I\mathsf{f}^2}{u}\sqrt{\left[\mathsf{f}^2\left|\nabla u\right|^2 + (\nabla u \cdot x)^2\right]/Q^2 + u^2} = u^{-2}, \tag{1}$$

where $x \in \mathbb{R}^2$ is in the image domain Ω, $|.|$ denotes the Euclidean vector norm, and

- $u := u(x)$ is the sought depth map,

- $I := I(x) = \frac{E(x)}{\sigma}$ is a normalised version of the brightness $E(x)$ of the given grey-value image, whereby σ depends on the albedo of the surface, i.e., the extent to which it diffusely reflects light, as well as on the brightness of the light source,

- f is the focal length relating the optical center of the camera and its retinal plane,

- $Q := Q(x) := \dfrac{\mathsf{f}}{\sqrt{|x|^2 + \mathsf{f}^2}}$.

As indicated, perspective models such as (1) yield superior depth maps compared to orthographic models. However, also in the widely studied model represented by (1) only Lambertian surfaces are taken into account.

On the numerical side, perspective models of the type (1) are dealt with by use of the level set method in [27, 28, 30] and by employing the dynamic programming principle in [19, 25]. As the numerical side is of importance in this paper, let us note that the schemes relying on the dynamic programming principle are generally advocated as they do not need any specific knowledge about the true solution. We will pursue a different approach as in the mentioned works here, discretising directly our new PDE.

Let us also note, that a first yet not very sophisticated attempt to incorporate non-Lambertian surface reflectance into a perspective SfS model is described in [16].

Our contributions. This work represents our results on the modelling as well as on the numerical side. We give in detail the modelling of a new Hamilton-Jacobi PDE for perspective SfS, thereby showing how to incorporate a model for non-Lambertian surfaces. For this, the Phong model well-known in the area of computer graphics [18, 17] is used. We also state clearly which physical assumptions are used that influence the computational setup. The second objective of us is to give a *mathematically justified, efficient* and *easy-to-code* algorithm. We realise this aim by using a finite difference discretisation of the arising PDE together with a sweeping method and a coarse-to-fine strategy. The resulting method is shown to be much faster than the schemes relying on the dynamic programming principle advocated in previous works [19, 25], while it is at the same time much simpler to implement. The finite difference method is analysed rigorously, especially with respect to the choice of the time step size guaranteeing the stability of the method. As shown by numerical experiments, we achieve a considerable gain concerning the quality of computed depth maps compared to non-Lambertian models, and we obtain reasonable results even for simple real-world images.

Relation to previous work. This paper represents a significant extension of our conference contributions [32, 33]. In the mentioned works, we deal with perspective SfS for Lambertian surfaces and a basic discussion of our non-Lambertian model, respectively. In contrast to [33] we elaborate here in full detail on the modelling process, thereby addressing some details important for implementation. On the theoretical side, we extend the previous works by addressing the choice of suitable time step sizes ensuring stability for the general reflection model, incorporating here in this context a multi-scale

3

analysis. Furthermore, we document in this work some further developments on the numerical side. In previous works, we used a semi-implicit scheme in order to handle convective terms and source terms separately, which is a usual proceeding in the field. We improve this numerical scheme using a coarse-to-fine strategy here as well as a sweeping technique combined with a fully explicit ansatz motivated by the multi-scale analysis. This leads to a significant efficiency improvement compared to results in [32, 33] while the implementation effort does not increase dramatically.

Paper organisation. In Section 2, we give the modelling process in detail. Then, in Section 3, we briefly discuss some theoretical properties of the PDE. Especially, we illuminate by use of a multi-scale analysis some fundamental properties of the resulting PDE of practical interest for the numerics. A detailed description of the numerical scheme we use is given in Section 4. This is followed by the stability investigation in Section 5. After an extensive discussion of numerical experiments in Section 6, the paper is finished by concluding remarks.

1   The model

Analogously to the procedure described in [19, 32], let us consider the surface $\mathcal{S}$ representing an object or scene of interest depicted in a given grey value image. Let $\mathcal{S}$ then be parameterised by use of the function $S : \bar{\Omega} \to \mathbb{R}^3, \Omega \subset \mathbb{R}^2$, with

$$S(x) = \frac{\mathsf{f}u(x)}{\sqrt{|x|^2 + \mathsf{f}^2}} \underbrace{(x, -\mathsf{f})^T}_{\in \mathbb{R}^2 \times \mathbb{R}} . \tag{2}$$

As the two columns of the Jacobian $\mathcal{J}[S(x)]$ are tangent vectors to $\mathcal{S}$ at the point $S(x)$, their cross-product is a normal vector to $\mathcal{S}$. One can compute then a normal vector $\vec{\mathsf{n}}(x)$ at the point $S(x)$ as

$$\vec{\mathsf{n}}(x) = \left(\mathsf{f}\nabla u(x) - \frac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2} x \,, \nabla u(x) \cdot x + \frac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2} \mathsf{f} \right)^T . \tag{3}$$

By (2)–(3), we just followed the procedure in [19, 32].

New contributions. We assume that the reflectance properties of the considered surface can be described by the Phong reflection model, see e.g. [18], and thus we introduce the *brightness equation*

$$I(x) = k_a I_a + \sum_{\text{lights}} \frac{1}{r^2} \big(k_d I_d \cos \phi + k_s I_s (\cos \theta)^\alpha \big) \tag{4}$$

4

where $I(x)$ is the normalised grey value of the image pixel located at x.

For the convenience of the reader, let us give some comments on the brightness equation (4). I_a, I_d, and I_s are the intensities of the ambient, diffuse, and specular components of light, respectively. Accordingly, the constants k_a, k_d, and k_s with $k_a + k_d + k_s \leq 1$ denote the ratio of ambient, diffuse, and specular reflection. The light contributions of all light sources are added up, i.e., practically one needs to compute the light contribution of each light source separately. Concerning the individual reflection contributions, let us note that the ambient light models light present everywhere in a given scene, i.e., it is a base intensity. The intensity of diffusely reflected light in each direction is proportional to the cosine of the angle ϕ between surface normal and light source direction. The amount of specular light reflected towards the viewer is proportional to $(\cos \theta)^{\alpha}$, where θ is the angle between the ideal (mirror) reflection direction of the incoming light and the viewer direction, α being a constant modelling the roughness of the material. For $\alpha \to \infty$ this describes an ideal mirror reflection.

Note especially, that the so-called light attenuation factor $1/r^2$, where r is the distance between light source and surface, is taken into account. This is noteworthy as including this factor does not only support the well-posedness of the model, it also yields an improved quality of surface reconstructions if comparing with computational results neglecting it; compare [19] for the case of Lambertian surfaces.

In the following, we employ some simplifications. First, we restrict the model to a single light source at the optical center of the camera [19, 32]. As in this case the view direction and light source direction are the same, we obtain $\theta = 2\phi$. Then, while in the original Phong model the light intensities are assumed to be RGB (colour) values, we restrict the model to scalar valued intensities as we only consider grey value images.

With these simplifications, equation (4) becomes

$$I(x) \; = \; k_a I_a + \frac{1}{r^2} \left(k_d (\vec{\mathrm{N}} \cdot \vec{\mathrm{L}}) I_d + k_s (2 (\vec{\mathrm{N}} \cdot \vec{\mathrm{L}})^2 - 1)^{\alpha} I_s \right) , \qquad (5)$$

where $\vec{\mathrm{N}} = \frac{\vec{\mathrm{n}}(x)}{|\vec{\mathrm{n}}(x)|}$ denotes the unit normal vector at the surface at point x, and where $\vec{L}$ is the unit light vector pointing towards the optical center of the camera. In (5), we used the reformulation

$$\cos \theta \; = \; \cos 2\phi \; = \; (\cos \phi)^2 - (\sin \phi)^2 \; = \; 2(\cos \phi)^2 - 1 \; = \; 2(\vec{\mathrm{N}} \cdot \vec{\mathrm{L}})^2 - 1 . \quad (6)$$

Note here, that by the ansatz (5) the Lambertian model (1) is included as a special case, i.e., for $k_a = k_s = 0$ and $k_d = 1$.

As the normalised light source direction $\vec{\mathrm{L}}$ is given by

$$\vec{\mathrm{L}}\left(S(x)\right) \;=\; \frac{1}{\sqrt{|x|^2 + \mathsf{f}^2}}\left(-x, \mathsf{f}\right)^T \,, \tag{7}$$

we can evaluate the inner product $\vec{\mathrm{N}} \cdot \vec{\mathrm{L}}$ as

$$
\begin{aligned}
&\vec{\mathrm{N}} \cdot \vec{\mathrm{L}}\left(S(x)\right) \\
&= \frac{\left(\mathsf{f}\nabla u(x) - \dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}x\,,\; \nabla u(x)\cdot x + \dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}\mathsf{f}\right)^T \cdot (-x, f)^T}{|\vec{\mathrm{n}}(x)|\sqrt{|x|^2 + \mathsf{f}^2}} \tag{8}\\[2mm]
&= \frac{-\mathsf{f}\nabla u(x)\cdot x + \dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}x\cdot x + (\nabla u(x)\cdot x)\mathsf{f} + \dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}\mathsf{f}^2}{|\vec{\mathrm{n}}(x)|\sqrt{|x|^2 + \mathsf{f}^2}} \tag{9}\\[2mm]
&= \frac{\dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}|x|^2 + \dfrac{\mathsf{f}u(x)}{|x|^2 + \mathsf{f}^2}\mathsf{f}^2}{|\vec{\mathrm{n}}(x)|\sqrt{|x|^2 + \mathsf{f}^2}} = \frac{\mathsf{f}u(x)}{|\vec{\mathrm{n}}(x)|\sqrt{|x|^2 + \mathsf{f}^2}}\,. \tag{10}
\end{aligned}
$$

By use of $r = \mathsf{f}u(x)$, we obtain from (5)-(10)

$$I(x) = k_a I_a + \frac{1}{\mathsf{f}^2 u(x)^2}\left(k_d \frac{u(x)Q(x)}{|\vec{\mathrm{n}}(x)|}I_d + k_s\left(\frac{2u(x)^2 Q(x)^2}{|\vec{\mathrm{n}}(x)|^2} - 1\right)^\alpha I_s\right)\,, \tag{11}$$

with

$$|\vec{\mathrm{n}}(x)| \;=\; \sqrt{\mathsf{f}^2|\nabla u(x)|^2 + (\nabla u(x)\cdot x)^2 + u(x)^2 Q(x)^2}\,,\; Q(x) \;=\; \sqrt{\mathsf{f}^2/(|x|^2 + \mathsf{f}^2)}\,. \tag{12}$$

Let us point out here, that (11) is a hyperbolic partial differential equation (PDE) of Hamilton-Jacobi-type. We rewrite (11) yielding the more convenient formulation

$$\left(I(x) - k_a I_a\right)\frac{\mathsf{f}^2 |\vec{\mathrm{n}}(x)|}{Q(x)u(x)} - \frac{k_d I_d}{u(x)^2} - \frac{|\vec{\mathrm{n}}(x)|k_s I_s}{u(x)^3 Q(x)}\left(\frac{2u(x)^2 Q(x)^2}{|\vec{\mathrm{n}}(x)|^2} - 1\right)^\alpha = 0\,. \tag{13}$$

We now employ for our numerical implementation – as it is usual when dealing with this problem – that the surface $\mathcal{S}$ is visible, i.e., it is in the front of the optical center, so that u is strictly positive. Then we use the change

6

of variables $v = \ln(u)$ which especially implies

$$\frac{|\vec{\mathfrak{n}}(x)|}{u(x)} = \frac{\sqrt{\mathsf{f}^2|\nabla u(x)|^2 + (\nabla u(x) \cdot x)^2 + u(x)^2 Q(x)^2}}{u(x)} \tag{14}$$

$$= \sqrt{\frac{\mathsf{f}^2|\nabla u(x)|^2 + (\nabla u(x) \cdot x)^2 + u(x)^2 Q(x)^2}{u(x)^2}} \tag{15}$$

$$= \sqrt{\mathsf{f}^2|\nabla v(x)|^2 + (\nabla v(x) \cdot x)^2 + Q(x)^2}, \tag{16}$$

since $\nabla v(x) = \dfrac{1}{u(x)}\nabla u(x)$. After some further simple computations, we finally obtain the PDE

$$J(x)W(x) - k_d I_d \exp\left(-2v(x)\right) - \frac{W(x)k_s I_s}{Q(x)} \exp\left(-2v(x)\right) \left(\frac{2Q(x)^2}{W(x)^2} - 1\right)^\alpha = 0 \tag{17}$$

where

$$J(x) = (I(x) - k_a I_a)\mathsf{f}^2/Q(x), W(x) = \sqrt{\mathsf{f}^2|\nabla v|^2 + (\nabla v \cdot x)^2 + Q(x)^2}. \tag{18}$$

The PDE (17) is the basis for our numerical implementation. Note that in the Phong model, the cosine in the specular term is usually replaced by zero if $\cos\theta = \frac{2Q(x)^2}{W(x)^2} - 1 < 0$.

Some modelling assumptions imposed indirectly by this set-up are pointed out in the section concerned with the derivation of a suitable stability condition.

2 Theoretical discussion

Existence, uniqueness and well-posedness. A point of general interest in the context of this work is the *existence* and *uniqueness* of an analytical solution of the model. For this, we need to consider the PDE together with appropriate boundary conditions φ over a finite domain, i.e., over a given image.

Let us clearly point out, that the appropriate notion of solutions for Hamilton-Jacobi equations of the type discussed here is the notion of *viscosity solutions*, see for instance [5]. Also, it should be understood clearly, that in the case of Hamilton-Jacobi equations there is in general more than one allowed viscosity solution, two of which can be distinguished as the *viscosity sub-* or *supersolution*, respectively. This linguistic usage stems from the fact that the viscosity supersolution is in a pointwise sense larger or equal to all other

Table 1: Units

Term	Unit
$[\mathsf{f}], [x_1], [x_2]$	m
$[v], [k_a], [k_d], [k_s], [I_d], [I_s]$	1
$[v_x], [v_y]$	$1/m$
$[I], [I_a]$	$1/m^2$

possible viscosity solutions. The viscosity subsolution is the analogous lower bound solution in the described sense.

We are especially interested in computing the viscosity supersolution, as by this unique choice we avoid strange artefacts like convex/concave ambiguities. In order to do this, it is recommended for theoretical purposes to set Dirichlet boundary conditions $\varphi \equiv +\infty$ as already elaborated in [19]. Note that this theoretical setting does not imply numerical problems or difficulties; as a practical consequence the boundary condition becomes virtually unimportant, and we can use Neumann boundary conditions in our implementation without any problems. However, let us stress, that this particular consequence arises by an *Eulerian formulation* of the problem. It does not hold, e.g., in the case of the recent semi-Lagrangian method for Lambertian SfS of Cristiani et al. [6].

The Hamiltonian corresponding to the model can be verified to fit into the framework presented in [20, 21], and thus the described boundary value problem is *well-posed*; compare the investigations in [19, 20, 21] to which we have nothing to add here.

Multi-scale analysis. We now give a *multi-scale analysis* of the problem, which illuminates an important construction point of our numerical solver.

In a multi-scale analysis as done, e.g., in the context of computational fluid dynamics [15], the variables of interest are expressed via a reference value times a normalised dimensionless new variable. In many cases, it is then useful to do all computations in the dimensionless variables, multiplying just for the final output the computed values with the corresponding reference values.

As a side effect of such a proceeding, it is possible to check if the dimensions in a new model are put together correctly. Table 2 shows the units of the occuring terms in the PDE.

It follows easily, that $[Q] = 1$ and $[J] = \frac{1}{m^2} m^2 = 1$. For $[W]$, we obtain

$$[W] = \sqrt{\frac{m^2}{m^2} + \left(\frac{m}{m} + \frac{m}{m}\right)^2 + 1^2} = \sqrt{1^2} = 1 \qquad (19)$$

8

which means all occuring terms in the PDE (17) are dimensionless.

Now we focus on the influence of different orders of magnitude of contributions, as this sheds light on the computational set-up. In order to conquer a useful result, it turns out that it makes sense to employ scalings in the point location $x = (x_1, x_2)^T$ as well as in the focal length f. These are parameters by which we may finally describe the ratio of image size and distance of the retinal plane – where x is located – from the optical center. More specifically, we consider

$$x := x_{\text{ref}} \cdot \hat{x} \qquad \text{and} \qquad \mathsf{f} := f_{\text{ref}} \cdot \hat{f}, \qquad (20)$$

with the obvious scalar reference values $x_{\text{ref}}, f_{\text{ref}}$, and where $\hat{x} = (\hat{x}_1, \hat{x}_2)^T$, $\hat{f}$ are the scaled variables.

Note that f is a constant parameter for one single given image, so that it may seem at first glance that it is not a reasonable choice for our analysis. However, introducing a scaling of f as in (20) enables us to compare the difficulty of experiments, see the section on numerical tests.

Pure convection – the term JW. We consider first the terms contributing to the part of (17) corresponding to the *Lambertian model*, i.e., *(i)* $J(x)$, and *(ii)* $W(x)$. For $J(x)$ we obtain

$$\frac{(I(x) - k_a I_a)\,\mathsf{f}^2}{Q(x)}$$

$$= (I(x) - k_a I_a)\left(f_{\text{ref}} \cdot \hat{f}\right)^2 \frac{\sqrt{|x_{\text{ref}} \cdot \hat{x}|^2 + \left(f_{\text{ref}} \cdot \hat{f}\right)^2}}{f_{\text{ref}} \cdot \hat{f}} \qquad (21)$$

$$= f_{\text{ref}}\left(I(x) - k_a I_a\right)\hat{f}^2 \cdot \frac{\sqrt{|\hat{x}|^2 \cdot x_{\text{ref}}^2 + \hat{f}^2 \cdot f_{\text{ref}}^2}}{\hat{f}} \qquad (22)$$

$$= f_{\text{ref}}^2\left(I(x) - k_a I_a\right)\hat{f}^2 \cdot \underbrace{\frac{\sqrt{|\hat{x}|^2 \cdot \gamma + \hat{f}^2}}{\hat{f}}}_{=:\,1/\hat{Q}_\gamma} \quad = \quad f_{\text{ref}}^2 \cdot \frac{(I(x) - k_a I_a)\,\hat{f}^2}{\hat{Q}_\gamma} \qquad (23)$$

with $\gamma := \dfrac{x_{\text{ref}}^2}{f_{\text{ref}}^2}$. Now we turn to $W(x)$. Plugging in the terms from (20), a

simple computation yields

$$
\sqrt{\mathsf{f}^2|\nabla v|^2 + (\nabla v \cdot x)^2 + Q(x)^2}
$$

$$
= \sqrt{\left(f_{\mathrm{ref}} \cdot \hat{f}\right)^2 |\nabla v|^2 + (x_{\mathrm{ref}}\nabla v \cdot \hat{x})^2 + \frac{\left(f_{\mathrm{ref}} \cdot \hat{f}\right)^2}{|x_{\mathrm{ref}} \cdot \hat{x}|^2 + \left(f_{\mathrm{ref}} \cdot \hat{f}\right)^2}} \quad (24)
$$

$$
= f_{\mathrm{ref}} \cdot \sqrt{\hat{f}^2|\nabla v|^2 + \gamma(\nabla v \cdot \hat{x})^2 + \frac{1}{f_{\mathrm{ref}}^2} \cdot \hat{Q}_\gamma^2}. \quad (25)
$$

Putting (23) and (25) together gives

$$
J(x)W(x) = f_{\mathrm{ref}}^3 \cdot \frac{(I(x) - k_a I_a)\,\hat{f}^2}{\hat{Q}_\gamma} \cdot \underbrace{\sqrt{\hat{f}^2|\nabla v|^2 + \gamma(\nabla v \cdot \hat{x})^2 + \frac{1}{f_{\mathrm{ref}}^2} \cdot \hat{Q}_\gamma^2}}_{=:\,W_\gamma}.
$$

$$(26)$$

At this point of the discussion, let us consider the *range of variables* in (26):

- $(I(x) - k_a I_a)$ is in $[0, 1]$.

- $\hat{f}$ is in $(0, 1]$; if not used as a comparison tool, it holds $\hat{f} = 1$.

- Using the point $(0, 0)^T$ in the center of any given image, the components of $\hat{x} = (\hat{x}_1, \hat{x}_2)^T$ are in $[-1, 1]$.

- γ is typically in the range of $\left[\frac{1}{2}, 1\right]$ in our experiments.

- The latter point implies $\hat{Q}_\gamma \in (0, 1]$.

- f_{ref} ist typically of the order 10^2 to 10^3 in our experiments.

These aspects lead to some *conclusions*, which we discuss now in some detail.

(a) By the factor f_{ref}^3 it becomes clear that the PDE is quite stiff. Moreover, this problem becomes even more significant for large focal lengths.

(b) The term $1/f_{\mathrm{ref}}^2 \cdot \hat{Q}_\gamma^2$ is virtually unimportant; its strongest contribution is close to the image center, roughly decreasing quadratically. In our experiments, this means it decreases to approximately $10^{-1} \cdot 1/f_{\mathrm{ref}}^2$ with the distance from the center.

10

(c) The factor $1/\hat{Q}_\gamma$ grows from 1 at the image center to $(2\gamma + \hat{f}^2)/\hat{f}$ at the image boundary. In our experiments, this is roughly a factor of $2 - 3$ for $\hat{x}$ near the image boundary.

(d) If $\hat{x}$ is close to the image corners, i.e., for $\hat{x} \approx (\pm 1, \pm 1)^T$, we obtain a notable contribution from the term $\gamma(\nabla v \cdot \hat{x})^2$. Simplifying terms below the root, we obtain at the image corners approximately $\sqrt{\hat{f}^2 + \gamma} \cdot |\nabla v|$. In our experiments, $\sqrt{\hat{f}^2 + \gamma}$ amounts to a factor around $5/4$. Hence, approaching the boundary the term $\gamma(\nabla v \cdot \hat{x})^2$ may become slightly more important than the term $\hat{f} |\nabla v|^2$.

These conclusions show that there are spatially dependent factors involved within the computation which are especially important close to the boundary, see (c)-(d). Especially, a multiplier of the considerable number f_{ref}^3 arises, increasing its importance. This fact will influence the construction of our scheme, see the next paragraph.

Let us also note that the term $1/f_{\mathrm{ref}}^2 \cdot \hat{Q}_\gamma^2$ is negligible compared to all other contributions. Thus, we may safely neglect it during the computation of a suitable time step size.

The specular term. We now consider the contribution due to the specular term. For this, let us note that it can be rearranged as

$$
\frac{W(x)k_s I_s}{Q(x)} \exp\left(-2v(x)\right) \left(\frac{2Q(x)^2}{W(x)^2} - 1\right)^\alpha
$$
$$
= k_s I_s \exp\left(-2v(x)\right) \underbrace{\left(\frac{2Q(x)^2}{W(x)^2} - 1\right)^\alpha}_{=:\, A} \cdot \frac{W(x)}{Q(x)} . \qquad (27)
$$

Thus, it can be written in the same fashion as $J(x)W(x)$ which can be arranged as

$$
J(x)W(x) = (I(x) - k_a I_a)\mathsf{f}^2 \cdot \frac{W(x)}{Q(x)} . \qquad (28)
$$

A quick glance at the content of the bracket in A in (27) reveals

$$
\frac{2Q^2}{W^2} - 1 = \frac{2Q^2 - \mathsf{f}^2|\nabla v|^2 + (\nabla v \cdot x)^2 + Q^2}{W^2} = \frac{Q^2 - \mathsf{f}^2|\nabla v|^2 + (\nabla v \cdot x)^2}{W^2} . \qquad (29)
$$

Analogously to the procedure leading to (25), we may draw a factor f_{ref}^2 out of the nominator and the denominator in (29). Thus, in total no contribution in terms of a factor f_{ref} arises by A. Therefore, we will not consider A further

11

and leave it unchanged, just looking after the other factors in the specular term.

Note that considering the terms in A it is clear that A will finally be in the range of $[-1, 1]$, whereby a negative number will be mapped to zero. Also, as α is usually quite large, a strong damping except for points with highlights is usual with respect to the specular contribution.

By simply identifying terms in (26) and (28), we obtain from (27)-(29)

$$
\begin{aligned}
& k_s I_s \exp\left(-2v(x)\right) \left(\frac{2Q(x)^2}{W(x)^2} - 1\right)^\alpha \cdot \frac{W(x)}{Q(x)} \\
= \; & f_{\mathrm{ref}} \cdot \frac{k_s I_s}{\hat{Q}_\gamma} \exp\left(-2v(x)\right) \underbrace{\left(\frac{2Q(x)^2}{W(x)^2} - 1\right)^\alpha}_{=: A} \cdot \hat{W}_\alpha \, .
\end{aligned}
\tag{30}
$$

Let us stress, that (29)-(30) reveals an important working mechanism of our model. As the factor f_{ref}^3 in (26) is a considerable order of magnitude larger than the factor f_{ref} arising in (29), in most parts of a given grey value image the Lambertian SfS model dominates the reconstruction of the 3-D shape. Only in parts where the factor $\exp\left(-2v\right)$ in (29) is of the same order of magnitude as f_{ref}^2 we obtain an important contribution of the specular term. Let us briefly comment on the logic behind the latter argumentation, sharpening by this an intuitive understanding of our new model. At a specular reflection, i.e., at a highlight, the grey value can be expected to be near 255 (white). The SfS assumption, especially the light attenuation term, implies that a bright image location is closer to the camera than a dark image region. Thus, the value u describing the shape at a highlight will be close to zero, as u is given in terms of the distance to the retinal plane. By the change of variables $v = \ln u$ used before (14), this means that at a bright image point with $0 < u \ll 1$ we have $v \ll 0$, and $\exp\left(-2v\right)$ will be very large. The minus sign before the specular term then implies, that a surface point with a highlight is pushed a bit into the direction of the darker background.

The diffuse part. Dividing equation (17) by f_{ref}^3, we observe that also a contribution by the *diffuse term* mainly arises at highlights. The smallness of this term renders it quite unimportant compared with other contributions, so that we deal with it explicitly, which is in contrast to previous works [19, 32] where the intuition to discretise source terms implicitly was employed.

3 Description of the numerical scheme

Artificial time. In order to solve the arising boundary value problem, we employ the *method of artificial time*. This means, we introduce a pseudo-

time variable t writing $v := v(x, t)$, $x := (x_1, x_2)^T \in \mathbb{R}^2$, and we iterate in this pseudo-time until a steady state defined by $v_t = 0$ is attained. This proceeding has the following properties desired by us:

- It fits our philosophy as it allows to define an easy-to-code numerical solver.

- Beginning to evolve in time with an initial state close to the viscosity supersolution of the PDE, see [1, 5, 19], the iterates reliably converge to an approximation of the sought viscosity supersolution.

Convection dominant form. We incorporate the information on the relative size of terms we obtained in the last paragraph by dividing the whole equation by f_{ref}^3. This emphasizes the dominance of the transport term JW over the sources present in all points but highlights.

Setting up the discretisation. Dropping for abbreviation the dependence on spatial and temporal variables, we thus opt to solve the time-dependent PDE reading as

$$
v_t = \underbrace{f_{\mathrm{ref}}^{-3} JW - f_{\mathrm{ref}}^{-3} \frac{W k_s I_s}{Q} \exp\left(-2v\right) \left(\frac{2Q^2}{W^2} - 1\right)^{\alpha} - f_{\mathrm{ref}}^{-3} k_d I_d \exp\left(-2v\right)}_{=:\, B}.
$$

$$(31)$$

Notation. In order to present the basic components of our numerical method, we employ the following notation:

- $v_{i,j}^n$ denotes the approximation of $v(ih_1, jh_2, n\tau)$,

- i and j are the coordinates of the pixel (i, j) in x_1- and x_2-direction, respectively,

- h_1 and h_2 are the corresponding mesh widths,

- τ is a time step size which needs to be chosen automatically or by the user.

Temporal discretisation. For the discretisation of the time derivative $v_t(x, t)$ we use the well-known Euler forward method, i.e.,

$$
\frac{v_{i,j}^{n+1} - v_{i,j}^n}{\tau}
$$

$$(32)$$

approximates $v_t(x, t)$ at $(x, t) = (ih_1, jh_2, n\tau)$.

As indicated, the value of τ can be defined by the user, however, the restriction applies that the scheme shall be stable. In Section 4, we will derive a stability condition on the time step size.

In defining τ we apply our knowledge from the multi-scale analysis and compute it in a *pointwise* fashion, i.e., $\tau = \tau_{i,j}$. The reason for this is the spatially dependent factor analysed in the previous section, see the conclusions (c) and (d) there. If we would define τ as a constant number globally for the whole image, our iterates would satisfy the stability condition near the image boundary which is the critical region in this context, but at the price of converging very slowly in the center region of the image. Note also, that our aim is to compute the steady state defined by $v_t = 0$, so that a pointwise choice of τ is feasible without introducing a consistency problem. This also holds with respect to the factor f_{ref}^{-3}, as in the steady state one may safely multiply the PDE with f_{ref}^3.

Spatial discretisation. Let us now consider the spatial terms. Here, the discretisation will be more intricate. The discretisation of $I(x)$ and $Q(x)$ is simple as these terms can be evaluated pointwise at all pixels (i, j). As a building block for the discretisation of *spatial derivatives* incorporated in W, we use the stable upwind-type discretisation of Rouy and Tourin [26], reading

$$v_{x_1}(ih_1, jh_2, \cdot) \;\; \approx \;\; \min\left(0, \frac{v_{i+1,j} - v_{i,j}}{h_1}, \frac{v_{i-1,j} - v_{i,j}}{h_1}\right), \qquad (33)$$

$$v_{x_2}(ih_1, jh_2, \cdot) \;\; \approx \;\; \min\left(0, \frac{v_{i,j+1} - v_{i,j}}{h_2}, \frac{v_{i,j-1} - v_{i,j}}{h_2}\right). \qquad (34)$$

Note, that in (33)-(34) the time level is not yet specified. The reason for this is that we use a combination of *(i)* a sweeping technique and *(ii)* a Gauß-Seidel-type iteration in order to accelerate convergence, leading to a different choice of values for each sweeping direction.

Let us elaborate on this as follows, beginning with the description how to incorporate the Gauß-Seidel-idea. Notice that at pixel (i, j) the data

$$
\begin{array}{ccc}
 & v_{i,j+1} & \\
v_{i-1,j} & v_{i,j} & v_{i+1,j} \\
 & v_{i,j-1} &
\end{array}
\qquad (35)
$$

are used in (32)-(33) because of the upwind differences. Let us assume for the moment that we iterate pixel-wise *from left to right* and, beginning with the top line, *from top to bottom* over the computational grid. Thus, ascending in i and descending in j, we incorporate the already updated values into the

scheme. This procedure yields the formulae

$$
v_{x_1}(x,t)\big|_{(x,t)=(ih_1,jh_2,t)} \approx \min\left(0, \frac{v_{i+1,j}^n - v_{i,j}^n}{h_1}, \frac{v_{i-1,j}^{n+1} - v_{i,j}^n}{h_1}\right), \quad (36)
$$

$$
v_{x_2}(x,t)\big|_{(x,t)=(ih_1,jh_2,t)} \approx \min\left(0, \frac{v_{i,j+1}^{n+1} - v_{i,j}^n}{h_2}, \frac{v_{i,j-1}^n - v_{i,j}^n}{h_2}\right). \quad (37)
$$

Let us emphasize, that the data $v_{i,j+1}^{n+1}$ and $v_{i-1,j}^{n+1}$ in (36)-(37) are already computed via the described procedure, so that they are fixed and one can safely use them for the computation of $v_{i,j}^{n+1}$.

Sweeping. We now turn to the sweeping technique which we adopt from the works [35, 36]. The idea behind this technique is as follows. For hyperbolic equations like the Hamilton-Jacobi PDEs we are dealing with, information is transported along characteristics. Thus, iterating in only one manner – e.g., always ascending in i and descending in j as described above – information flow may not be realised numerically in an optimal way.

Staying within the Gauß-Seidel-type framework of our example, in a pixel (i,j) updated values are always taken into account just from the left, see (36), and from above, see (37). Therefore, information is numerically propagated faster in the directions "from left to right" and "from top to bottom" as in all other possible directions. As a remedy, it is obvious to switch the iteration directions in the following cyclic fashion:

1. Left $\rightarrow$ Right, and Top $\rightarrow$ Bottom

2. Top $\rightarrow$ Bottom, and Right $\rightarrow$ Left

3. Right $\rightarrow$ Left, and Bottom $\rightarrow$ Top

4. Bottom $\rightarrow$ Top, and Left $\rightarrow$ Right

Exactly this procedure is called *sweeping*. As is easily seen, defining the sweepings as above different values $v_{i\pm1,j\pm1}^{n+1}$ are needed to be taken into account in (33)-(34) depending on the sweeping direction.

Source terms. Being a factor within the part of the scheme where upwinding is employed, we discretise $k_s I_s e^{-2v}$ at pixel (i,j) using the known data at point (i,j), i.e., setting $k_s I_s e^{-2v_{i,j}^n}$. Note, that this is also adequate as the corresponding specular term in practice only yields notable contributions at highlights, compare our discussion in Section 3.

Finally, let us consider the source term $f_{\text{ref}}^{-3} k_d I_d e^{-2v}$. Source terms like this typically enforce the use of very small time step sizes when evaluated *explicitly*, leading in this case to very long computational times. However, we have

seen in our multi-scale analysis, that this source term generally yields a very small contribution, and even at highlights its contribution is of one order in f_{ref} lower than the specular contribution. Thus, we discretise it explicitly, i.e.,

$$f_{\text{ref}}^{-3} k_d I_d e^{-2v(x,t)}\big|_{(x,t)=(ih_1,jh_2,n\tau)} \approx f_{\text{ref}}^{-3} k_d I_d e^{-2v_{i,j}^n} . \tag{38}$$

Scheme summary. Letting $\hat{B}$ denote the discretised version of term B from (31), we obtain pixelwise the update formula

$$v_{i,j}^{n+1} = v_{i,j}^n - \tau \hat{B} - \tau f_{\text{ref}}^{-3} k_d I_d e^{-2v_{i,j}^n} . \tag{39}$$

Thus, by (39) we obtain a completely explicit scheme which needs to evaluated pixelwise, iterating over the grid.

Coarse to fine. In addition to the sweeping technique, we employ a cascading multigrid method, compare [4]. This actually amounts to a coarse-to-fine algorithm, where we interpolate the input image linearly on coarser grids from the original input image. In the restriction step, we take the maximum over all neighbouring depths as initialization for new grid points. This ensures consistency of our scheme, since we rely on being above the viscosity solution. This accelerates the convergence of our iterative solver.

4   The choice of the time step size

Conditional stability. The method of artificial time we use is relatively easy to implement, as it does not rely on solving systems of linear or nonlinear equations. This is very important, as it enables a large spectrum of image processing researchers to access and use the model. However, the price one has to pay for this desirable property is that a consistent explicit method like (39) is typically *conditionally stable*, i.e., the use of a too large time step size leads to a blow-up of numerical solutions.

In practice, estimates for an upper bound on the time step size are often too pessimistic for direct use. In our case we will employ – motivated by the discussion in Section 3 – a local estimate leading to a convenient pixelwise computation of a suitable time step size. Moreover, as we will also see, a stability investigation as performed here encourages to write down clearly the underlying modelling assumptions in the discrete setting.

Discrete minimum-maximum principle. A meaningful stability notion in the context of the discussed PDE is a *discrete minimum-maximum-principle*, i.e., the solution obeys an upper and lower bound determined by the data. To ensure the validity of this criterion, we impose a local version of it, namely that the update is not allowed to become smaller or larger

than all neighbouring data taken into account within the stencil. If this corresponding condition is met pixelwise, the validity of the global discrete minimum-maximum-principle is implied as can easily be seen. Let us remark, that this stability notion relates to the corresponding property of viscosity solutions of the discussed PDE [5].

Let us stress here, that the model (31) is too nonlinear to obtain an elegant and nice formula for a stability condition on τ. A mathematically correct, but impractical means of computing a suitable τ would be to compute pointwise all occuring terms in their discrete representation, and compute τ so that the discrete minimum-maximum-principle holds. Besides the arising numerical effort to check several inequalities, this brute force method does not emphasise an understanding of the underlying process. In particular, it turns out that one can gain valuable insights into our model by use of some slight simplifications.

In order to realise the described concept, we first consider the update formula (39) without the diffuse source term:

$$v_{i,j}^{n+1} = v_{i,j}^n - \tau \hat{B} \,. \tag{40}$$

As in

$$\hat{B} \;=\; f_{\mathrm{ref}}^{-3} JW - f_{\mathrm{ref}}^{-3} \frac{W k_s I_s}{Q} \exp\left(-2v\right) \left(\frac{2Q^2}{W^2} - 1\right)^\alpha \,, \tag{41}$$

the term $f_{\mathrm{ref}}^{-3} JW$ and the specular term arise additively, and we can proceed by considering the contributions separately adding up the resulting constraints in the end. The diffuse part is discussed separately at the end of this paragraph.

Convection. Taking into account only the term $f_{\mathrm{ref}}^{-3} JW$, a local discrete minimum-maximum principle is satisfied provided τ satisfies the condition $\left|\tau f_{\mathrm{ref}}^{-3} JW\right| \leq \delta v$ with

$$\delta v := \max\left(\left|\frac{v_{i+1,j} - v_{i,j}}{h_1}\right|, \left|\frac{v_{i-1,j} - v_{i,j}}{h_1}\right|, \left|\frac{v_{i,j+1} - v_{i,j}}{h_2}\right|, \left|\frac{v_{i,j-1} - v_{i,j}}{h_2}\right|\right) \,. \tag{42}$$

As indicated before, we do not specify the time level in (42) because of the employed sweeping algorithm. The data in δv are taken from time level n or $n + 1$, depending on the sweeping direction, respectively.

Employing the notation $\nabla \hat{v}$ for the discretisation of ∇v within W, we compute the following estimates:

$$|\nabla \hat{v}|^2 \leq \left(\sqrt{2\delta v^2}\right)^2 = 2\delta v^2 \quad \text{and} \quad (\nabla \hat{v} \cdot x)^2 \leq ((x_1 + x_2)\,\delta v)^2 = (x_1 + x_2)^2 \,\delta v^2 \,, \tag{43}$$

where we make use of the pixel coordinates $x = (x_1, x_2)$. At the pixel with index (i, j) this is equal to (ih_1, jh_2), i.e., a scaling by use of x_{ref} is not taken into account in the above formulation. Note, that we could estimate globally $(x_1 + x_2)^2 = x_{\text{ref}}^2 (\hat{x}_1 + \hat{x}_2)^2 \leq x_{\text{ref}}^2 (1 + 1)^2 = 4x_{\text{ref}}^2$, but instead we follow a local formulation taking into account $\hat{x} = (\hat{x}_1, \hat{x}_2)$. Denoting then by $\hat{W}$ the discretised version of W and using $\hat{Q}_{i,j}^2 \leq 1$, we arrive at

$$\hat{W} \leq f_{\text{ref}} \cdot \sqrt{2\hat{f}^2 \delta v^2 + \gamma (\hat{x}_1 + \hat{x}_2)^2 \delta v^2 + \frac{1}{f_{\text{ref}}^2}} \,. \tag{44}$$

Employing then the complete expression $f_{\text{ref}}^{-3} JW$ at pixel (i, j) as well as the discussion below (26), we obtain

$$
\begin{aligned}
f_{\text{ref}}^{-3} JW \big|_{(i,j)} &\leq \frac{(I_{i,j} - k_a I_a) \hat{f}^2}{\hat{Q}_\gamma} \cdot \sqrt{2\hat{f}^2 \delta v^2 + \alpha (\hat{x}_1 + \hat{x}_2)^2 \delta v^2 + \frac{1}{f_{\text{ref}}^2}} \\
&\approx \frac{(I_{i,j} - k_a I_a) \hat{f}^2}{\hat{Q}_\gamma} \cdot \sqrt{2\hat{f}^2 \delta v^2 + \alpha (\hat{x}_1 + \hat{x}_2)^2 \delta v^2} \\
&= \frac{(I_{i,j} - k_a I_a) \hat{f}^2}{\hat{Q}_\gamma} \cdot \sqrt{2\hat{f}^2 + \alpha (\hat{x}_1 + \hat{x}_2)^2} \cdot \delta v \,. \tag{45}
\end{aligned}
$$

See also the remark at the end of this paragraph with respect to the slight simplification used above. Note that $\hat{Q}_\gamma$ depends on (i, j). The resulting expression (45) implies for all reasonable $\delta v \neq 0$ the pointwise stability condition

$$\tau \leq \frac{\hat{Q}_\gamma}{(I_{i,j} - k_a I_a) \hat{f}^2 \sqrt{2\hat{f}^2 + \gamma (\hat{x}_1 + \hat{x}_2)^2}} \,. \tag{46}$$

Specular source term. We now turn to the discretised specular source term

$$f_{\text{ref}}^{-2} \frac{\hat{W}_\gamma k_s I_s}{\hat{Q}_\gamma} \exp\left(-2v_{i,j}^n\right) \left(\frac{2Q_{i,j}^2}{\hat{W}^2} - 1\right)^\alpha , \tag{47}$$

see in particular (30) and (41).

The following time step size estimate is interesting, as it nicely illustrates the influence of some underlying modeling assumptions. As noted below (18), the term $(\cdot)^\alpha$ is nonnegative. Thus, one could estimate the power of $2Q_{i,j}^2/\hat{W}^2 - 1$ via

$$\frac{2Q_{i,j}^2}{\underbrace{f^2 |\nabla \hat{v}|^2 + (\nabla \hat{v} \cdot x)^2}_{\geq 0} + Q_{i,j}^2} - 1 \leq \frac{2Q_{i,j}^2}{Q_{i,j}^2} - 1 = 1 \tag{48}$$

18

by $1^\alpha = 1$. However, let us point out clearly, that this estimation is at first glance very pessimistic in a local setting as exactly the strong damping via $(\cdot)^\alpha$ is responsible for the negligible size of contributions of the specular term at non-highlights within computations. Thus, the above simplification may be understood as a reasonable estimate at highlights. This fact needs to be kept in mind in interpreting other occuring contributions.

Diffuse source term. It remains to discuss the factor $\exp\left(-2v_{i,j}^n\right)$ which proves to be a bit more intricate. As $v = \ln u$, we take into account the

$$\text{Modelling assumption No. 1} \qquad u \geq u_{\min} > 0 \,, \tag{49}$$

where $u_{\min}$ denotes the minimal distance an object is away from the camera. Note, that this assumption is not really a restriction, as *(i)* by it, we avoid division by zero, and *(ii)* the situation that an object is directly before the camera is not of practical interest. We would also like to impose

$$\text{Modelling assumption No. 2} \qquad \forall\, i,j \,:\, \mathsf{f}\sqrt{I_{i,j} - k_a I_a} \geq 1 \,. \tag{50}$$

As I is normalised, we have substituting $\mathsf{f}^2 u^2$ for r^2

$$I_{i,j} - k_a I_a \leq \frac{1}{\mathsf{f}^2 u^2} \,, \tag{51}$$

see (4). After some trivial manipulations, the latter inequality leads to

$$u \leq \frac{1}{\mathsf{f}\sqrt{\beta}} \,, \quad \text{where} \quad \beta := \min_{i,j}\left(I_{i,j} - k_a I_a\right) \,, \tag{52}$$

and where we set $\beta > 0$, again in accordance with (4). Thus, by (52) we impose an upper bound $u_{\max}$ with $u_{\max} \leq 1$, i.e, we assume that there are no black pixels in the input image for the SfS process. Note, that this assumption is usually made for SfS algorithms without a detailed explanation, see e.g. [19, 25]. As a second consequence of the basic relation (51), we can fix $u_{\min}$ via

$$\mathsf{f}^2 \max_{i,j}\left(I_{i,j} - k_a I_a\right) = \frac{1}{u_{\min}^2} \,. \tag{53}$$

In summary, it follows

$$u \in [u_{\min}, u_{\max}] \;\Rightarrow\; \ln(u) \in [\ln\left(u_{\min}\right), \ln\left(u_{\max}\right)] \subset (-\infty, 0] \,. \tag{54}$$

For a global estimate of the exponential term in (47), we thus could employ

$$\exp\left(-2v_{i,j}^n\right) \leq \exp\left(-2\ln\left(u_{\min}\right)\right) = \frac{1}{u_{\min}^2} \,. \tag{55}$$

In a similar fashion as in (45), we can then find the estimate

$$\tau \leq \frac{f_{\text{ref}}^2 u_{\min}^2 \hat{Q}_\gamma}{k_s I_s \sqrt{2\hat{f}^2 + \gamma \left(\hat{x}_1 + \hat{x}_2\right)^2}} \, . \tag{56}$$

Note that, by (53), the number $f_{\text{ref}}^2 u_{\min}^2$ is in general larger than 1, so that the specular part only implies a restriction on τ for reasonably large numbers I_s, i.e., at highlights. Also, a highlight implies that u is identical or close to $u_{\min}$, so that also the numerator is concerned with this situation. Thus, a necessary condition for (56) to hold is given by

$$\tau \leq \frac{\hat{Q}_\gamma}{k_s I_s \sqrt{2\hat{f}^2 + \gamma \left(\hat{x}_1 + \hat{x}_2\right)^2}} \, . \tag{57}$$

At this point of the discussion, we may summarise the conditions (46) and (57) as

$$\tau \leq \frac{\hat{Q}_\gamma}{\sqrt{2\hat{f}^2 + \gamma \left(\hat{x}_1 + \hat{x}_2\right)^2}} \cdot \left(\frac{1}{k_s I_s} + \frac{1}{\left(I_{i,j} - k_a I_a\right) \hat{f}^2}\right) \, . \tag{58}$$

It remains to discuss the influence of the diffuse term $f_{\text{ref}}^{-3} k_d I_d e^{-2v_{i,j}^n}$. As seen in (55), one could estimate $e^{-2v_{i,j}^n}$ by $u_{\min}^{-2}$; then (53) shows that $f_{\text{ref}}^{-2} u_{\min}^{-2}$ is identical to $\max_{i,j}\left(I_{i,j} - k_a I_a\right)$, which should be close to 1 in many situations of interest. A factor f_{ref}^{-1} remains, implying that the diffuse term yields a minor contribution at specular highlights, while at all other points the contribution is damped exponentially.

Let us remark here, that instead of making use of the scaling argument to neglect $1/f_{\text{ref}}^2$ in (45), one could estimate the root from above by a Taylor linearisation at $1/f_{\text{ref}}^2$. Then this term would be counted to the sources, having the same order of magnitude as the diffuse part. This proceeding would appeal to be a bit more rigorous, but it is also longer and bears no additional content.

In a similar fashion as the slight simplification leading to (45) and (56), we neglect the minor diffuse term, defining for practical purposes τ as

$$CFL \cdot \frac{\hat{Q}_\gamma}{\sqrt{2\hat{f}^2 + \gamma \left(\hat{x}_1 + \hat{x}_2\right)^2}} \cdot \left(\frac{1}{k_s I_s} + \frac{1}{\left(I_{i,j} - k_a I_a\right) \hat{f}^2}\right) \tag{59}$$

where $CFL := 0.7 < 1$ to ensure stability.

5 Experiments

In this section, we give both quantitative and qualitative experimental evaluations of our method on synthetical and real images.
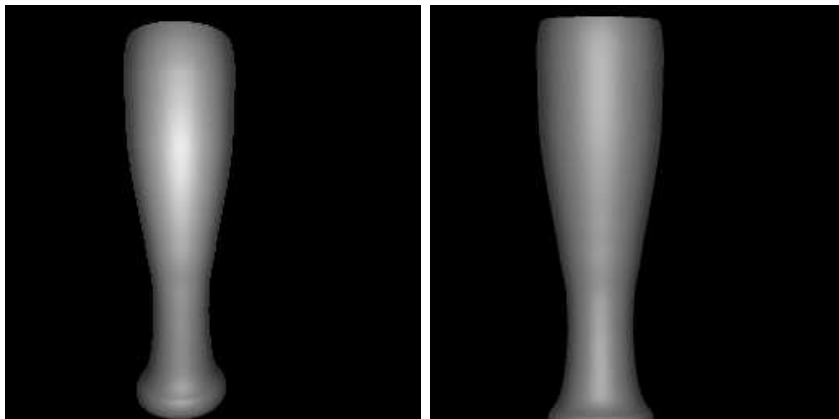

Figure 1: Synthetic wheat beer glass input images. Left: $f = 251.7$, Right: $f = 1000$.

Synthetic experiment. This test case addresses two qualitatively different situations, corresponding to different ratios $x_{\text{ref}}/f_{\text{ref}}$; see the discussion in Section 2.

Figure 1 shows images of a synthetic wheat beer glass, rendered with two different focal lengths. The rendering parameters were $f = 1000$, $h_1 = h_2 = 1$, $k_d = 0.7$, $k_s = 0.3$, $I_d = I_s = 80$ and $f = 221.7$, $h_1 = h_2 = 1$, $k_d = 0.7$, $k_s = 0.3$, $I_d = I_s = 4$, respectively. The smaller focal length leads to extreme perspective distortion. While perspective helps to make the problem well-posed, too much perspective distortion can make the reconstruction more difficult, e.g. because of occlusions at object boundaries.

Figure 2 shows the ground truth surfaces for both input images, i.e., the surfaces that would be obtained when using the known correct depths. In the following evaluation, we compare the reconstructed depth with the correct depth from this known surface.

In Figure 3 we find reconstructions of the synthetic beer glass images using both the Lambertian surface model [19, 32] as well as the reconstruction using the method presented in this paper. Parameters of the reconstruction are the correct ones we know from the rendering process. As we can see, the reconstruction for both input images looks much better using the non-Lambertian method. The Lambertian reconstruction looks distorted at the specular highlights. In addition to that, it is estimated much too close to
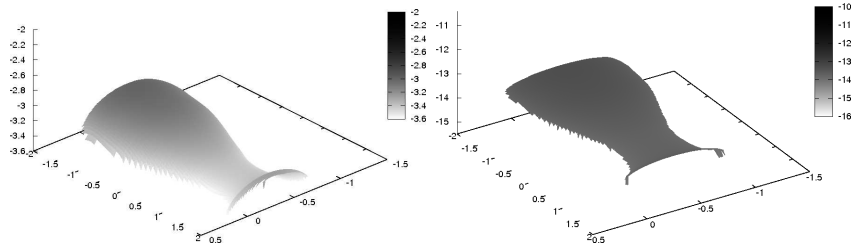
Figure 2: Ground truth surfaces for the beer glass experiment. Left: $f = 251.7$, Right: $f = 1000$.
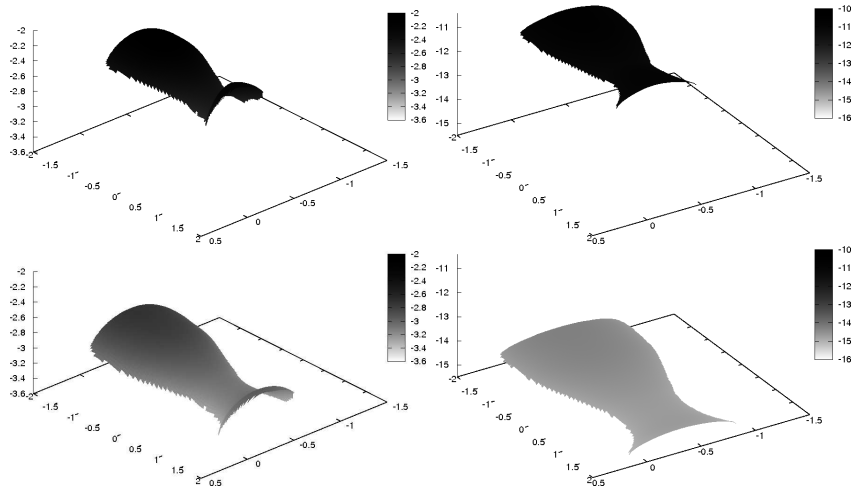

Figure 3: Reconstructed surfaces of the beer glass experiment. Left: $f = 251.7$, Right: $f = 1000$. Top row: Lambertian surface model, Bottom row: Phong surface model.

the camera. In contrast, the reconstruction using the Phong surface model is much more accurate for both input images.

Comparing the average L_1 error relative to the actual depth, we get the values shown in Table 2. Using the Phong model for the reconstruction clearly gives better results than using a Lambertian surface model, for both input images. Note, however, that despite the fact that perspective is helpful for the reconstruction, the reconstruction with the larger focal length has a smaller error. Some parts of the surface might be occluded in the image with large perspective distortion, which leads to missing information about the surface. Since the background is black, the reconstruction will be as steep as possible at the boundary of the glass, since the attenuation term implies

an infinite depth there. For small focal lengths, the attenuation factor plays a larger role in the reconstruction, which is mainly useful for transitions between different objects. It is a tradeoff between a larger impact of the attenuation factor and potential occlusions, but in general a very small focal length makes the reconstruction difficult.

Table 2: Relative L_1 depth errors for the beer glass experiment.

	f = 251.7	f = 1000
Lambertian	30.79%	20.67%
Phong	9.59%	7.00%

Finally, Table 3 shows the computational times of the reconstruction of our proposed method compared to the recent state-of-the-art (Lambertian) method by Prados et al. [19, 25]. We can see that our method is considerably faster than this Lambertian reference method, although more terms are considered due to the more complicated surface model. The reconstruction times were obtained using a straightforward implementation in C of both methods on a standard PC (Linux, Pentium IV, 3.2 GHz, 2 GB RAM). Employing a coarse-to-fine strategy improves the times even further. In each experiment, we stopped the iteration as soon as the maximum change in the depth u was smaller than 10^{-7}.

Table 3: Computation times for the beer glass experiment.

	f = 251.7	f = 1000
Prados et al. (Lambertian)	263s	273s
Our method (Phong)	172s	164s
Our method (Phong) + coarse-to-fine	111s	104s

In summary, this experiment shows that using the Phong surface model for shape reconstruction from shading on images including specular highlights drastically improves the reconstruction accuracy. Also, we see that our method is significantly faster than the state-of-the-art Lambertian method in the field.

The real beer glass experiment. In the second experiment, we test the algorithm on a *real-world image*. Figure 4 shows an image of a beer mug. The image was taken with a normal digital camera with built-in camera flash. The surface was approximately 50 centimeters away from the camera.

Since the writings on the surface of the mug would violate the assumption that objects have uniform reflectance properties, we interpolate the image

Figure 4: Real-world experiment: Original image, interpolation mask and input image.

using image inpainting with homogeneous diffusion [3] at these locations. Figure 4 shows the interpolation mask and the resulting input image.

Figure 5 shows the reconstruction using both a Lambertian surface model and the Phong model with the estimated parameters $f = 3000$, $h_1 = h_2 = 1$, $I_s = I_d = I_a = 2000$, $k_a = 0.1$, $k_d = 0.6$, $k_s = 0.3$. Again, we observe a distortion of the surface at specular highlights. This distortion is much stronger for the Lambertian reconstruction than for the reconstruction using the Phong model. Still, the reconstruction is not perfect even with the Phong model, but much closer. Also, the step at the lower half of the mug is far too steep in the Lambertian reconstruction, while it is recovered pretty well using the Phong model.

Additionally, the Lambertian method estimates the cup to be much closer to the camera than it actually is, which is not observable in the shown images. This results in a larger perspective distortion of the background, which is reconstructed much better if using the Phong method. Note that in this image, several model assumptions are violated, there is more than one surface in the image, and the camera flash is not located at the optical center, causing shadows that could not occur if this model assumption was fulfilled. This shadow results in different reconstructions of both sides of the mug, on the
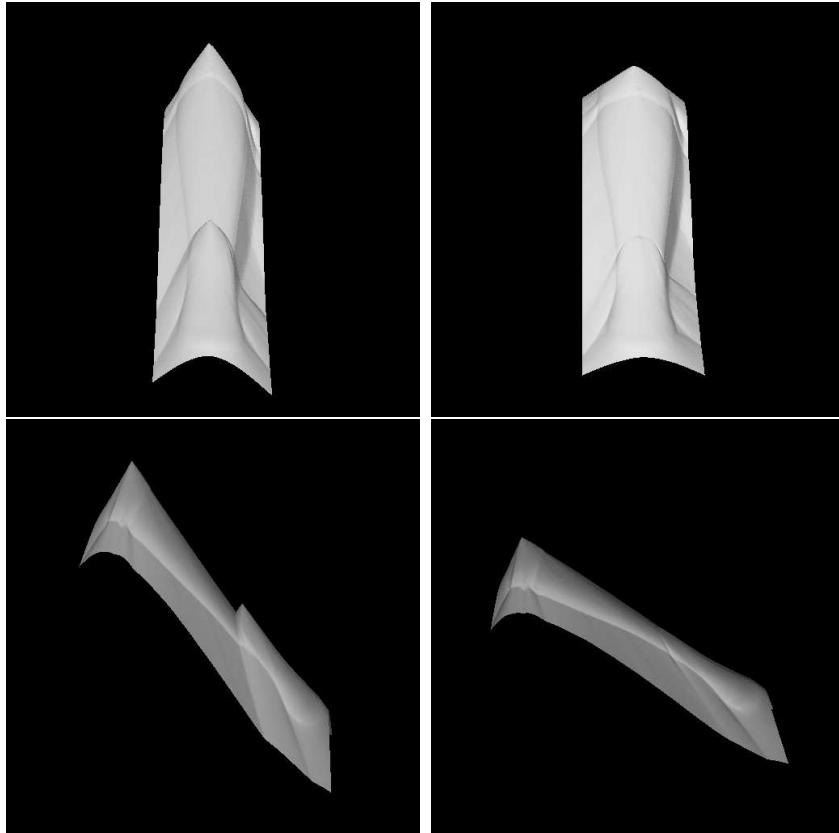
Figure 5: Reconstructions of the beer mug. Left: Lambertian surface model. Right: Phong surface model.

shadow side, the reconstruction is much steeper. This is natural, since the darker pixel implies the depth to be larger because of the light attenuation term.

In conclusion, the reconstruction using the Phong model is much better than the one using the Lambertian model, in particular at specular highlights. This illustrates that on real-world scenes, where a lot of highlights occur, our new model can perform significantly better than other approaches in the field.

6   Summary

We have introduced an important improvement in the promising class of perspective shape from shading models, incorporating the possibility to use

more realistic surface reflectance properties than before. Also, we have discussed in detail an efficient yet relatively simple and easy-to-code algorithm, which is very important for a practical use of the model.

Our developments in the near future extending the discussed model will center around a reasonable parameter estimation for the purposes of SfS, especially concerning the albedo of objects. Furthermore, we seek further improvements on the numerical side, although the proposed scheme seems to be a good choice with respect to the balance of efficiency and coding simplicity.

7 Acknowledgements

References

[1] R. Abgrall and V. Perrier. On the numerical approximation of first order hamilton jacobi equationson the numerical approximation of first order hamilton jacobi equations. *International Journal of Applied Mathematics and Computer Sciences*, 17(3):403–412, 2007.

[2] S. Bakshi and Y.-H. Yang. Shape from shading for non-Lambertian surfaces. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 130–134, Austin, TX, November 1994. IEEE Computer Society Press.

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. SIGGRAPH 2000*, pages 417–424, New Orleans, July 2000.

[4] F. Bornemann and P. Deuflhard. Cascadic multigrid methods. In R. Glowinski, J. Periaux, Z. Shi, and O. Widlund, editors, *Domain Decomposition Methods in Sciences and Engineering*, pages 205–212, Chichester, New York, 1997. John Wiley & Sons.

[5] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.

[6] E. Cristiani, M. Falcone, and A. Seghini. Some remarks on perspective shape-from-shading models. In *Proc Scale Space and Variational Meth-*

ods in Computer Vision, volume 4485 of *Lecture Notes in Computer Science*, pages 276–287, Berlin, May-June 2007. Springer.

[7] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993.

[8] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips. *Grundlagen der Computergraphik*. Addison–Wesley, Bonn, 1994.

[9] B. K. P. Horn. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Department of Electrical Engineering, MIT, Cambridge, MA, 1970.

[10] B. K. P. Horn. Obtaining shape from shading information. In P. H. Winston, editor, *The Psychology of Computer Vision*, chapter 4, pages 115–155. McGraw-Hill, New York, NY, 1975.

[11] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[12] B. K. P. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5:37–75, 1990.

[13] B. K. P. Horn and M. J. Brooks. *Shape from Shading*. Artificial Intelligence Series. MIT Press, 1989.

[14] K. Lee and C.-C. J. Kuo. Shape from shading with a generalized reflectance map model. *Computer Vision and Image Understanding*, 67(2):143–160, August 1997.

[15] A. Meister. *Asymptotic Expansions and Numerical Methods in Computational Fluid Dynamics*, pages 63–137. Narosa Publishing House, New Delhi, 2003.

[16] T. Okatani and K. Deguchi. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer Vision and Image Understanding*, 66(2):119–131, May 1997.

[17] B. T. Phong. Illumination for computer-generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[18] B. T. Phong. *Illumination of Computer-Generated Images*. PhD thesis, Department of Computer Science, University of Utah, July 1975.

[19] E. Prados. *Application of the theory of the viscosity solutions to the Shape from Shading problem.* PhD thesis, University of Nice Sophia-Antipolis, Oct. 2004.

[20] E. Prados, F. Camilli, and O. Faugeras. A unifying and rigorous shape from shading method adapted to realistic data and applications. *Journal of Mathematical Imaging and Vision*, 25(3):307–328, 2006.

[21] E. Prados, F. Camilli, and O. Faugeras. A viscosity solution method for shape-from-shading without image boundary data. *Mathematical Modelling and Numerical Analysis (M2AN)*, 40(2):393–412, 2006.

[22] E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. In *Proc Ninth International Conference on Computer Vision*, volume 2, pages 826–831, Nice, France, October 2003. IEEE Computer Society Press.

[23] E. Prados and O. Faugeras. Unifying approaches and removing unrealistic assumptions in shape from shading: Mathematics can help. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, pages 141–154, Berlin, 2004. Springer.

[24] E. Prados and O. Faugeras. A generic and provably convergent shape-from-shading method for orthographic and pinhole cameras. *International Journal of Computer Vision*, 65(1-2):97–125, 2005.

[25] E. Prados and O. Faugeras. Shape from shading: A well-posed problem? In *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 870–877, San Diego, CA, June 2005. IEEE Computer Society Press.

[26] E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM Journal of Numerical Analysis*, 29(3):867–884, 1992.

[27] A. Tankus, N. Sochen, and Y. Yeshurun. A new perspective [on] shape-from-shading. In *Proc. Ninth International Conference on Computer Vision*, volume 2, pages 862–869, Nice, France, Oct. 2003. IEEE Computer Society Press.

[28] A. Tankus, N. Sochen, and Y. Yeshurun. Perspective shape-from-shading by fast marching. In *Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 43–49, Washington, DC, June-July 2004. IEEE Computer Society Press.

[29] A. Tankus, N. Sochen, and Y. Yeshurun. Reconstruction of medical images by perspective shape-from-shading. In *Proc. 2004 International Conference on Pattern Recognition*, volume 3, pages 778–781, Cambridge, UK, Aug. 2004.

[30] A. Tankus, N. Sochen, and Y. Yeshurun. Shape-from-shading under perspective projection. *International Journal of Computer Vision*, 63(1):21–43, June 2005.

[31] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, Englewood Cliffs, 1998.

[32] O. Vogel, M. Breuß, and J. Weickert. A direct numerical approach to perspective shape-from-shading. In H. Lensch, B. Rosenhahn, H.-P. Seidel, P. Slusallek, and J. Weickert, editors, *Vision, Modeling, and Visualization*, pages 91–100, Saarbrücken, Germany, November 2007.

[33] O. Vogel, M. Breuß, and J. Weickert. Perspective shape from shading with non-Lambertian reflectance. In G. Rigoll, editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 517–526, Berlin, June 2008. Springer.

[34] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

[35] Y.-T. Zhang, H. Zhao, and J. Qian. High order fast sweeping methods for static hamilton-jacobi equations. *Journal of Scientific Computing*, 29(1):25–56, 2006.

[36] H. Zhao. A fast sweeping method for eikonal equations. *Mathematics of Computation*, 74(250):603–627, 2004.