Perspective Shape from Shading with Non-Lambertian Reflectance

Oliver Vogel, Michael Breuß, and Joachim Weickert

Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science, Building E1.1
Saarland University, 66041 Saarbrücken, Germany.
{vogel,breuss,weickert}@mia.uni-saarland.de

Abstract. In this work, we extend the applicability of perspective Shape from Shading to images incorporating non-Lambertian surfaces. To this end, we derive a new model inspired by the perspective model for Lambertian surfaces recently studied by Prados et al. and the Phong reflection model incorporating ambient, diffuse and specular components. Besides the detailed description of the modeling process, we propose an efficient and stable semi-implicit numerical realisation of the resulting Hamilton-Jacobi equation. Numerical experiments on both synthetic and simple real-world images show the benefits of our new model: While computational times stay modest, a large qualitative gain can be achieved.

1 Introduction

Given a single greyscale image, the shape-from-shading (SFS) problem amounts to computing the 3-D depth of depicted objects. It is a classical problem in computer vision with many potential applications, see [1–3] and the references therein for an overview.

In early SFS models, the camera is performing an *orthographic projection* of the scene of interest, which is assumed to be composed of Lambertian surfaces. Let us especially honour the pioneering works of Horn [1, 4] who was also the first to model the problem via the use of a *partial differential equation* (PDE).

For orthographic SFS models, there have been some attempts to extend the range of applicability to non-Lambertian surfaces [5, 6]. However, orthographic models usually suffer from ill-posedness, especially in the form of the so-called convex-concave ambiguities [3]. Moreover, the orthographic camera model yields reconstruction results not of convincing quality in most situations [3].

The problem of ill-posedness can be dealt with successfully by using a *perspective camera model*, see e.g. [3, 7–10]. As our model incorporates for a special choice of parameters the perpective model for Lambertian surfaces widely studied in the literature, let us give some more details on these. Assuming a pinhole camera model and a point light source at the optical center, the perspective SFS model for Lambertian surfaces amounts to the *Hamilton-Jacobi equation*

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

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

- $u := u(x)$ is the sought depth map,
- $I := I(x)$ is the normalised brightness of the given grey-value image
- f is the focal length relating the optical center and its retinal plane,
- $Q := Q(x) := \mathsf{f}/\sqrt{|x|^2 + \mathsf{f}^2}$.

As already indicated, perspective models such as (1) yield superior depth maps compared to orthographic models. However, up to now there does not exist a *reliable* and *easy-to-use* PDE-based model incorporating non-Lambertian surfaces. An interesting, yet not very sophisticated attempt to incorporate other surface models into perspective SFS is given in [11].

Our contribution. In our paper, we introduce a detailed model of a new PDE for perspective SFS incorporating non-Lambertian surfaces. It is clearly stated at which point the Phong reflection model we use for this purpose is taken into account. A second objective is to give an *efficient* and *easy-to-code* algorithm. We realise this aim by using the algorithm of Vogel et al. [12] for perspective SFS and Lambertian surfaces as a basis. As our experiments show, we achieve a considerable gain concerning the *quality* of computed depth maps, and we obtain reasonable results even for simple *real-world images*.

Organisation of the paper. In Section 2, we present the model process in detail. In Section 3, a thorough description of the numerical scheme is given. Following a discussion of experiments in Section 4, the paper is finished by concluding remarks.

2 Description of the Model

Consider the surface $\mathcal{S}$ representing the object or scene of interest depicted in a given image, parameterised by using the function $S : \bar{\Omega} \to \mathbb{R}^3, \Omega \subset \mathbb{R}^2$ [12], 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 gives a normal vector $\mathrm{n}(x)$ at $S(x)$ by

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

Up to this point, the model is largely identical to the one for perspective SFS [3]. However, we assume that the surface properties can be described by a Phong reflection model [13, 14], and thus we introduce the brightness equation

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

Let us comment on equation (4) in some detail: I_a, I_d, and I_s are the intensities of the ambient, diffuse, and specular components of the reflected 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 attenuation factor $1/r^2$, where r is the distance between light source and surface, is taken into account. 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, and α is a constant modeling the roughness of the material. For $\alpha \to \infty$ this describes an ideal mirror reflection.

We restrict the model to a single light source at the optical center of the camera [15]. As in this case the view direction and light source direction are the same, we obtain $\theta = 2\phi$. Moreover we restrict the model to scalar valued intensities as we only consider greyscale images. Then equation (4) becomes

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

with $\mathrm{N} = \frac{\mathrm{n}(x)}{|\mathrm{n}(x)|}$ being the unit normal vector at the surface at point x, and L is the unit light vector which points towards the optical center of the camera. The scalar products $N\cdot L$ arise since $\cos\phi = \mathrm{N}\cdot\mathrm{L}$ and $\cos\theta = \cos 2\phi = 2(\cos\phi)^2 - 1 = 2(\mathrm{N}\cdot\mathrm{L})^2 - 1$. As the normalised light source direction is given by

$$\mathrm{L}\left(S(x)\right) = \left(|x|^2 + \mathsf{f}^2\right)^{-1/2}(-x, \mathsf{f})^T\,, \tag{6}$$

we can evaluate the scalar products yielding

$$\mathrm{N}\cdot\mathrm{L}\left(S(x)\right) = \mathsf{f}u(x)\left(|\mathrm{n}(x)|\sqrt{|x|^2+\mathsf{f}^2}\right)^{-1}\,. \tag{7}$$

By use of $r = \mathsf{f}u(x)$, we obtain from (5-7)

$$I(x) = k_a I_a + \frac{1}{\mathsf{f}^2 u(x)^2}\left(k_d\frac{u(x)Q(x)}{|\mathrm{n}(x)|}I_d + k_s\left(\frac{2u(x)^2Q(x)^2}{|\mathrm{n}(x)|^2}-1\right)^\alpha I_s\right)\,, \tag{8}$$

where $|\mathrm{n}(x)| = \sqrt{\mathsf{f}^2|\nabla u(x)|^2 + (\nabla u(x)\cdot x)^2 + u(x)^2Q(x)^2}$ and with

$$Q(x) = \sqrt{\mathsf{f}^2/(|x|^2+\mathsf{f}^2)}\,. \tag{9}$$

The PDE (8) is of Hamilton-Jacobi-type. We now rewrite (8), yielding

$$(I(x) - k_aI_a)\frac{\mathsf{f}^2|\mathrm{n}(x)|}{Q(x)u(x)} - \frac{k_dI_d}{u(x)^2} - \frac{|\mathrm{n}(x)|k_sI_s}{u(x)^3Q(x)}\left(\frac{2u(x)^2Q(x)^2}{|\mathrm{n}(x)|^2}-1\right)^\alpha = 0\,. \tag{10}$$

We assume – as usual when dealing with this problem – that the surface $\mathcal{S}$ is visible, so that u is strictly positive. Then we use the change of variables $v = \ln(u)$ which especially implies

$$\frac{|\mathrm{n}(x)|}{u(x)} = \sqrt{\mathsf{f}^2|\nabla v(x)|^2 + (\nabla v(x)\cdot x)^2 + Q(x)^2}\,, \tag{11}$$

since $\nabla v(x) = \frac{1}{u(x)} \nabla u(x)$. Neglecting the notational dependence on $x \in \mathbb{R}^2$, we finally obtain the PDE

$$JW - k_d I_d e^{-2v} - \frac{W k_s I_s}{Q} e^{-2v} \left(\frac{2Q^2}{W^2} - 1 \right)^{\alpha} = 0 \tag{12}$$

where

$$:= J(x) = (I(x) - k_a I_a) \mathsf{f}^2 / Q(x) \,, \tag{13}$$

$$W := W(x) = \sqrt{\mathsf{f}^2 |\nabla v|^2 + (\nabla v \cdot x)^2 + Q(x)^2} \,. \tag{14}$$

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

3 Numerical Method

In order to solve the arising boundary value problem (12), we employ the *method of artificial time*. As it is well-known, this is one of the most successful strategies to deal with static Hamilton-Jacobi equations, see e.g. [16] and the references therein. This means, we introduce a pseudo-time variable t riting $v := v\,(x,t)$, and we iterate in this pseudo-time until a steady state defined by $v_t = 0$ is attained.

Thus, for $v = v(x,t)$, we opt to solve the time-dependent PDE

$$v_t \; = \underbrace{JW - k_s I_s e^{-2v} \frac{W}{Q} \left(\frac{2Q^2}{W^2} - 1 \right)^{\alpha}}_{=:A} - k_d I_d e^{-2v} \,. \tag{15}$$

Discretisation. We employ the following standard 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, h_1 and h_2 are the corresponding mesh widths, and τ is a time step size which needs to be chosen automatically or by the user.

We approximate the *time derivative* v_t by the Euler forward method, i.e.,

$$v_t(x,t)|_{(x,t)=(ih_1, jh_2, n\tau)} \approx \frac{v_{i,j}^{n+1} - v_{i,j}^n}{\tau} \,. \tag{16}$$

Let us now consider the *spatial terms*. 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 [16]:

$$v_{x_1}(ih_1, jh_2, \cdot) \approx h_1^{-1} \max\left(0,\, v_{i+1,j} - v_{i,j},\, v_{i-1,j} - v_{i,j}\right) \,, \tag{17}$$

$$v_{x_2}(ih_1, jh_2, \cdot) \approx h_2^{-1} \max\left(0,\, v_{i,j+1} - v_{i,j},\, v_{i,j-1} - v_{i,j}\right) \,. \tag{18}$$

Note that in (17), (18) the time level is not yet specified, as we wish to employ a Gauß-Seidel-type idea, compare e.g. [12]. To this end, notice that at pixel (i,j)

the data $v_{i,j+1}, v_{i-1,j}, v_{i,j}, v_{i+1,j}, v_{i,j-1}$ are used. Iterating pixel-wise over the computational grid, ascending in i and descending in j, we incorporate already updated values into the scheme. This yields the formulae

$$v_{x_1}(x,t)|_{(x,t)=(ih_1,jh_2,n\tau)} \approx h_1^{-1} \max\left(0,\, v_{i+1,j}^n - v_{i,j}^n,\, v_{i-1,j}^{n+1} - v_{i,j}^n\right) , \quad (19)$$

$$v_{x_2}(x,t)|_{(x,t)=(ih_1,jh_2,n\tau)} \approx h_2^{-1} \max\left(0,\, v_{i,j+1}^{n+1} - v_{i,j}^n,\, v_{i,j-1}^n - v_{i,j}^n\right) . \quad (20)$$

Let us emphasize that the data $v_{i,j+1}^{n+1}$ and $v_{i-1,j}^{n+1}$ in (19),(20) are already computed via the described procedure, so that they are fixed and one can safely use them for the computation of $v_{i,j}^{n+1}$.

Being a factor before W, we discretise $k_s I_s e^{-2v}$ at pixel (i,j) using known data as it is the case in the remainder of this term, i.e., setting $k_s I_s e^{-2v_{i,j}^n}$.

Finally, let us consider the source term $k_d I_d e^{-2v}$. Source terms like this enforce the use of very small time step sizes when evaluated *explicitly*, leading to very long computational times. Thus, we discretise it *implicitly* by

$$k_d I_d e^{-2v(x,t)}|_{(x,t)=(i,j,n\tau)} \approx k_d I_d e^{-2v_{i,j}^{n+1}} . \quad (21)$$

Letting $\hat{A}$ denote the discretised version of term A from (15), we obtain by (16) the update formula

$$v_{i,j}^{n+1} = v_{i,j}^n - \tau \hat{A} - \tau k_d I_d e^{-2v_{i,j}^{n+1}} \quad (22)$$

which has to be solved for $v_{i,j}^{n+1}$. We treat the nonlinear equation (22) by use of the classical one-dimensional Newton method, which convergences in practice in 3-4 iterations.

To summarize our algorithm, we (i) initialize $v := -0.5 \log I\mathsf{f}^2$, (ii) iterate using equation (22) until the pixel-wise error in v is smaller than some predefined small constant ϵ in every pixel.

Note that it is possible to do a rigorous stability analysis, yielding a reliable estimate for τ useful for computations, compare [12] for an example of such a procedure.

4 Experiments

Let us now present experiments on both synthetic and real images. In all these experiments we use the method developed above. Let us note, that the method was compared to other recent methods in the field in the context of perspective SFS with Lambertian reflectance [12]. In the latter context, the proposed algorithm has turned out to be by far the most efficient numerical scheme.

Using synthetic test scenes, the ground truth is known, so that we can compare the reconstructions with it and get a quantitative measure of the error.

Reality, however, is different: Recent SFS methods consider only Lambertian surfaces, while in reality such surfaces do not exist. Although the Phong model is only an approximation to the physical reality, surfaces rendered with it appear much more realistic. In order to evaluate the benefit of the Phong-based approach for real-world images, we only consider synthetic surfaces rendered with the Phong model and $k_s > 0$. We use Neumann boundary conditions in all experiments.

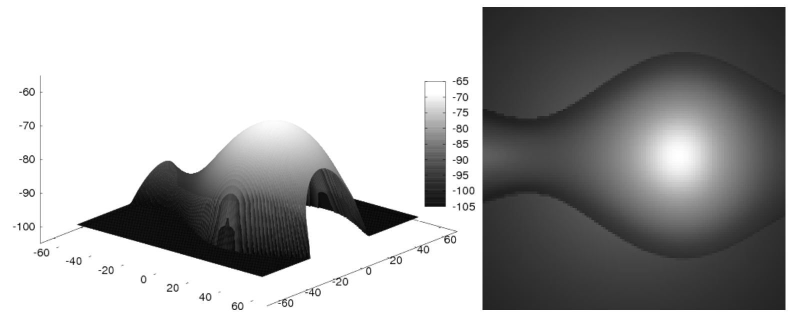

Fig. 1. The vase: Ground truth surface and rendered image.

The Vase Experiment. In our first experiment, we use a classic SFS test scene: The vase [2]. Figure 1 shows the ground truth together with a rendered version. It has been rendered by ray-tracing the surface which complies to our model with $\mathsf{f} = 110$, $h_1 = h_2 = 1$, $I_d = I_s = 3000$, $k_d = 0.6$, $k_s = 0.4$, $k_a = 0$, 128×128 pixels, and $\alpha = 4$.

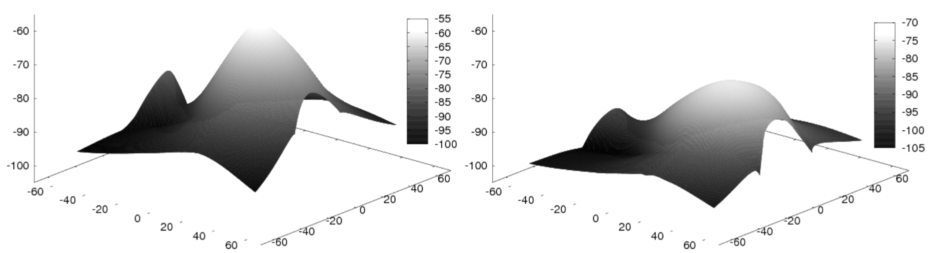

Fig. 2. Reconstructions of the vase. Left: Lambertian model, Right: Phong model.

In Fig. 2 we find reconstructions using our new model and the Lambertian model from [3, 12] employing the known parameters. The reconstruction with the Lambertian model looks distorted. The whole surface is pulled towards the light source, located at $(0, 0, 0)$. The effect is most prominent at specular highlights. At the boundary of vase and background, we observe a strong smoothing, which is normal when using a Lambertian model. The shape of the reconstruction using our new model is quite close to the original shape. It has the right distance from the camera and the correct size. The boundary of the vase is smooth, too, but the transition is clearly sharper than the one obtained using the Lambertian model. In Tab. 1, we compare the reconstruction with the ground truth. We observe a

Table 1. Average L_1 errors for the vase experiment.

	Phong model	Lambertian model
Error in u	0.042	0.088
Error in v	0.049	0.101
Relative error in u	4.62%	9.66%

considerable improvement using our model. The computational times are about one minute, which is just a bit slower than the Lambertian algorithm. Note that we omitted the relative error in v since $v = \log u$.

Let us now evaluate the performance of our method on noisy input images. Figure 3 shows the vase input image distorted by Gaussian noise with standard deviations $\sigma = 5$ and $\sigma = 10$. This is quite a lot of noise for shape from shading applications.

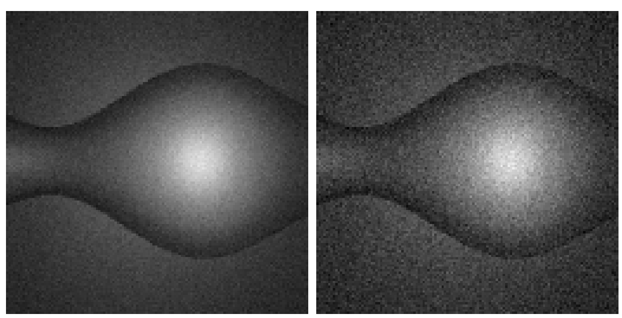

Fig. 3. The vase: Noisy input images. Left: $\sigma = 5$, Right: $\sigma = 10$.

Figure 4 shows the reconstructions of both noisy images using our new method. In both cases, the general shapes are preserved quite well, only the fine structure of the surface looks a bit noisy. Despite the strong noise in the input data, we get good reconstructions. The error levels in Tab. 2 support this impression. For the image distorted with only a little noise, we get almost the same reconstruction error as in the noiseless experiment. With the second image, the error is a bit higher, but almost the same. The method performs very stable under noise. For experiments with noise on Lambertian surfaces compare [3, 12].

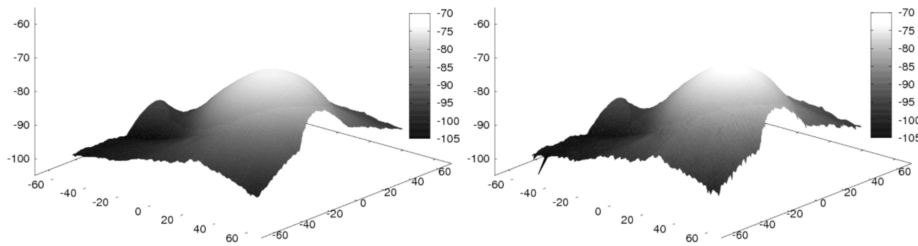

Fig. 4. The vase: Reconstructions of noisy data. Left: $\sigma = 5$, Right: $\sigma = 10$.

Table 2. Average L_1 errors for the vase experiment on noisy images.

	Noise, $\sigma = 5$	Noise, $\sigma = 10$
Error in u	0.047	0.050
Error in v	0.051	0.054
Relative error in u	5.33%	5.49%

The Cup Experiment. In Fig. 5, we see a photograph of a plastic cup, taken with a normal digital camera with flashlight. In this image, several model assumptions are violated.

- We do not have the same surface throughout the image.
- The flashlight is not located at the optical center, creating shadows in the image.
- Some parts of the scene reflect on the surface of the cup, especially on the left.
- The image was taken in a room that was darkened, but certainly was not pitch black, and there was some ambient light reflected from the walls of the room.

Fig. 5. Photograph of a plastic cup.

Figure 6 shows reconstructions using Phong and Lambertian models, respectively. Parameters for the Phong reconstruction were $\mathsf{f} = 1500$, $h_1 = h_2 = 1$, $I_s = I_d = I_a = 2000$, $k_a = 0.1$, $k_d = 0.5$, $k_s = 0.4$, and $\alpha = 4$. We used some ambient lighting to compensate for ambient light present in the room. For the Lambertian reconstruction we used the same parameters, but with $k_s = 0$. In the Phong reconstruction, the plate in the background is flat like it should be. In the Lambertian reconstruction, we see artifacts at specular highlights on the cup where the surface is pulled towards the light source: The cup is estimated much too close to the camera (not observable in Fig. 6). In the Phong reconstruction, we hardly see artifacts. At the specular highlights, we have an almost normal shape. Even the handle of the cup and the plate are recovered very well.

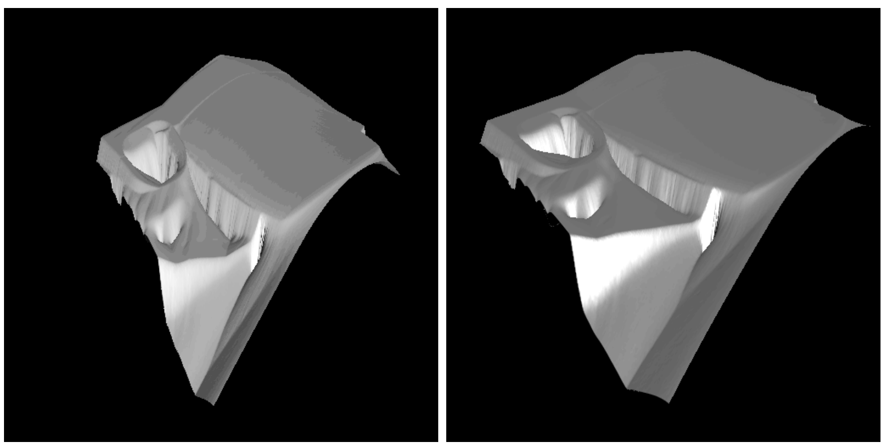

Fig. 6. Reconstructions of the cup. Left: Lambertian model. Right: Phong model.

5 Summary

We have introduced a Phong-type perspective SFS model that yields much more accurate results than its Lambertian counterpart. Moreover, we have seen that also real-world images and noisy images can be tackled successfully by our new model. This is an important step towards making SFS ideas applicable to a larger range of real-world problems.

Our ongoing work is concerned with more complicated illumination models as this step will improve the applicability of SFS to real-world problems, as well as algorithmic speedups. We also propose to use the current model with the presented technique for automated document analysis, which will be the subject of future work.

References

1. Horn, B.K.P., Brooks, M.J.: Shape from Shading. Artificial Intelligence Series. MIT Press (1989)
2. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from shading: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(8) (1999) 690–706
3. Prados, E.: Application of the theory of the viscosity solutions to the Shape from Shading problem. PhD thesis, University of Nice Sophia-Antipolis (2004)
4. Horn, B.K.P.: 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)
5. Bakshi, S., Yang, Y.H.: Shape from shading for non-Lambertian surfaces. In: Proc. IEEE International Conference on Image Processing. Volume 2., Austin, TX, IEEE Computer Society Press (1994) 130–134
6. Lee, K., Kuo, C.C.J.: Shape from shading with a generalized reflectance map model. Computer Vision and Image Understanding **67**(2) (1997) 143–160
7. Okatani, T., Deguchi, K.: Reconstructing shape from shading with a point light source at the projection center: Shape reconstruction from an endoscope image. In: Proc. 1996 International Conference on Pattern Recognition, Vienna, Austria (1996) 830–834
8. Tankus, A., Sochen, N., Yeshurun, Y.: Perspective shape-from-shading by fast marching. In: Proc. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1., Washington, DC, IEEE Computer Society Press (2004) 43–49
9. Tankus, A., Sochen, N., Yeshurun, Y.: Shape-from-shading under perspective projection. International Journal of Computer Vision **63**(1) (2005) 21–43
10. Cristiani, E., Falcone, M., Seghini, A.: Some remarks on perspective shape-from-shading models. In: Proc Scale Space and Variational Methods in Computer Vision. Volume 4485 of Lecture Notes in Computer Science., Berlin, Springer (2007) 276–287
11. Okatani, T., Deguchi, K.: 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) (1997) 119–131
12. Vogel, O., Breuß, M., Weickert, J.: A direct numerical approach to perspective shape-from-shading. In Lensch, H., Rosenhahn, B., Seidel, H.P., Slusallek, P., Weickert, J., eds.: Vision, Modeling, and Visualization, Saarbrücken, Germany (2007) 91–100
13. Phong, B.T.: Illumination of Computer-Generated Images. PhD thesis, Department of Computer Science, University of Utah (1975)
14. Phong, B.T.: Illumination for computer-generated pictures. Communications of the ACM **18**(6) (1975) 311–317
15. Prados, E., Faugeras, O.: Unifying approaches and removing unrealistic assumptions in shape from shading: Mathematics can help. In Pajdla, T., Matas, J., eds.: Computer Vision – ECCV 2004, Part IV. Volume 3024 of Lecture Notes in Computer Science., Berlin, Springer (2004) 141–154
16. Rouy, E., Tourin, A.: A viscosity solutions approach to shape-from-shading. SIAM Journal of Numerical Analysis **29**(3) (1992) 867–884