Making Shape from Shading Work for Real-World Images

Oliver Vogel, Levi Valgaerts, Michael Breuß, and Joachim Weickert

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

Abstract. Although shape from shading (SfS) has been studied for almost four decades, the performance of most methods applied to real-world images is still unsatisfactory: This is often caused by oversimplified reflectance and projection models as well as by ignoring light attenuation and nonconstant albedo behavior. We address this problem by proposing a novel approach that combines three powerful concepts: *(i)* By means of a Chan-Vese segmentation step, we partition the image into regions with homogeneous reflectance properties. *(ii)* This homogeneity is further improved by an adaptive thresholding that singles out unreliable details which cause fluctuating albedos. Using an inpainting method based on edge-enhancing anisotropic diffusion, structures are filled in such that the albedo does no longer suffer from fluctuations. *(iii)* Finally a sophisticated SfS method is used that features a perspective projection model, considers physical light attenuation and models specular highlights. In our experiments we demonstrate that each of these ingredients improves the reconstruction quality significantly. Their combination within a single method gives favorable perfomance also for images that are taken under real-world conditions where simpler approaches fail.

1 Introduction

An ultimate goal in computer vision is the 3-D reconstruction of our real world based on 2-D imagery. Although tremendous progress has been achieved when reconstructing a 3-D surface from *multiple* images [1], problems are much more severe when only a *single* image is available and the illumination is known. In our paper we address this so-called *shape-from-shading (SfS)* problem by introducing a novel framework that is particularly tailored to the difficulties one has to face in real-world scenarios.

In the SfS problem, one usually assumes that a three-dimensional surface is illuminated by a single light source whose direction is known. The goal is to reconstruct this 3-D surface from the brightness variations within a single 2-D image. It is evident that this is a very difficult task that requires a number of additional, simplifying model assumptions in order to become tractable.

The investigation of SfS models was pioneered by Horn [2]. His orthographic camera model and his Lambertian surface assumption became characteristic for numerous early SfS algorithms; see e.g. [3] for a survey. Another milestone in the development of SfS models are the approaches of Prados et al. [4], Tankus et al. [5], and Cristiani et al. [6]. They replaced the orthographic camera model by a pinhole camera model performing a perspective projection, and they assumed that the light source is located at the optical centre. Moreover, a light attenuation term is considered in [4]. These ideas have been further extended by Ahmed et al. [7] and by Vogel et al. [8]. In these works, the Lambertian reflectance model is replaced by the more realistic model of Oren and Nayar [9], which is particularly useful for skin surfaces, or by the Phong model from computer graphics [10], which models specular highlights. Many experts agree that Lambertian assumptions do not model realistic surfaces in an appropriate way [7, 11, 9].

Although this development shows a clear evolution of SfS models towards more realistic assumptions, most of these papers work on synthetic data. The few ones that use real-world data sets usually do not consider more realistic effects such as highlights or inhomogeneous reflectance properties as part of their models. In view of these difficulties, it is not surprising that in order to make SfS methods work in real-world applications, they had to be combined with external expertise provided e.g. by face databases and machine learning techniques [12] or by user-specified constraints [13].

Our Contribution. The goal of our paper is to show that by a more sophisticated approach, SfS works for a larger class of real-world images, even when no substantial a priori knowledge is available. To this end we combine three successful concepts:

- In order to extract the object of interest for the SfS process we segment the image with two level set approaches: the region-based Chan-Vese segmentation model [14] and the edge-based geodesic active contour model [15, 16].
- We detect fluctuations in the albedo by a local adaptive thresholding [17] and eliminate them by inpainting with edge-enhancing anisotropic diffusion [18].
- We use the non-Lambertian, perspective SfS model of Vogel et al. [8] that belongs to the most realistic SfS techniques and takes into account highlights.

Related Work. In our experiments we demonstrate that it is exactly the combination of the *three* successful concepts segmentation, albedo handling and non-Lambertian SfS that is crucial for the performance of our method. However, some related ideas with *two* combined concepts have been proposed in the literature. Concerning the combination of inpainting and SfS, Prados et al. [19] applied an algorithm of Tschumperlé and Deriche [20] for inpainting the eyes and the eyebrows for facial Lambertian SfS. Jin et al. [21] have combined a segmentation step with 3-D reconstruction of Lambertian surfaces. Their method also exploits multiple views.

Paper Organization. In Section 2, we present more details on the key concepts of our combined method. An evaluation of their individual usefulness is given in Section 3. The paper is concluded by a summary with outlook in Section 4.

2 Our Three-Stage Approach

Let us now have a more detailed look at the three key concepts that are combined within our SfS framework in order to exclude the background, to handle albedo variations and to deal with non-Lambertian surfaces.

2.1 Finding the Region of Interest – Segmentation

In a first step we separate the object of interest from the background. This is necessary since both have incompatible reflectance properties. For this task we use the active contour model of Chan and Vese [14]. This is a classic level-set-based method that exploits the grey-value difference between object and background.

The Chan-Vese model segments the image domain $\Omega \subset \mathbb{R}^2$ into two regions by minimising the difference between the image intensity $f(\mathbf{x}) : \Omega \to \mathbb{R}$ and its average value in each region. Additional constraints are imposed on the length of the region boundary C and on the area inside C. This comes down to minimising the energy

$$\begin{aligned}\mathcal{E}\,(C, c_1, c_2) &= \mu \text{ length}(C) + \nu \text{ area}(insideC) \\ &\quad + \int_{inside(C)} (f - c_1)^2 \mathrm{d}\mathbf{x} + \int_{outside(C)} (f - c_2)^2 \mathrm{d}\mathbf{x},\end{aligned} \tag{1}$$

where c_1 and c_2 are the average values of f inside and outside C, and $\mu \geq 0$ and $\nu \geq 0$ are weighting parameters. These weights are important to tune the object detection: A large μ will give a coarse segmentation, while a small μ will detect fine details. As a region-based segmentation model, the Chan-Vese method is fast and robust with respect to initialisation and noise.

In order to further improve the localization of the object contour, we use the Chan-Vese result as initialisation for the edge-based geodesic active contour model [15, 16]. The governing evolution equation is given by

$$\begin{aligned}&\partial_t \phi = |\nabla \phi| \text{ div}\left(g\left(|\nabla f_\sigma|\right) \frac{\nabla \phi}{|\nabla \phi|}\right) && \text{on } \Omega \times [0, \infty), \\ &\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}) && \text{on } \Omega,\end{aligned} \tag{2}$$

where $\phi(\mathbf{x}, t)$ is a level-set function, ϕ_0 a suitable initialisation and $\nabla = (\partial_x, \partial_y)^\top$ is the gradient operator. The edge stopping function g draws the contour towards nearby edges in the presmoothed image f_σ, which is obtained by convolving f with a Gaussian with standard deviation σ. The function $g(s^2)$ is decreasing in s. In our application we choose the Perona-Malik diffusivity $g_{PM}(s^2) = (1 + s^2/\lambda^2)^{-1}$, where $\lambda > 0$ is some contrast parameter [22]. If the object is bounded by a pronouced edge, the edge-based active contours will generally result in a sharper segmentation.

2.2 Ensuring a Homogeneous Albedo – Inpainting by Edge Enhancing Diffusion

Generally, real-world objects do not have a constant albedo. To apply SfS we need to ensure that the albedo does not vary within the segmented contour.

In our approach we detect regions of differing albedo and fill in neighborhood information to obtain homogeneous reflectance properties.

In order to identify regions with fluctuating albedo we use an adaptive thresholding algorithm that works on local windows [17]. Adaptive thresholding is robust with respect to varying illumination conditions within the scene and is widely used in document analysis. Note that by slightly enlarging the identified regions by morphological erosion we can improve the subsequent interpolation result, preventing artifacts at the boundaries.

The next step is to interpolate the image in these regions. For this task we choose edge-enhancing anisotropic diffusion (EED) [23]. It was shown to perform better for image inpainting and scattered data interpolation than other PDE-based methods [18]. The main idea behind EED is to allow smoothing within homogeneous regions and along image edges, but to reduce smoothing across them. To this end it makes use of a diffusion tensor. In the region that we want to inpaint we solve the steady-state diffusion equation

$$0 = \operatorname{div}\left(g(\nabla u_\sigma \nabla u_\sigma^\top)\nabla u\right), \tag{3}$$

with the boundary conditions specified by the surrounding data. Here u_σ is a smoothed version of the evolving image u, obtained by convolving it with a Gaussian of standard deviation σ. The scalar-valued diffusivity g is applied to the eigenvalues of the structure tensor $\nabla u_\sigma \nabla u_\sigma^\top$, while leaving its eigenvectors unchanged. This way, the first eigenvector of the diffusion tensor is parallel to the edge detector ∇u_σ. The desired filter effect comes from the fact that the corresponding eigenvalue is given by $g(|\nabla u_\sigma|^2)$, such that smoothing is reduced at edges, where $|\nabla u_\sigma|$ is large. The second eigenvector is orthogonal to ∇u_σ with corresponding eigenvalue 1. For the diffusivity g one typically chooses the Charbonnier diffusivity $g_C(s^2) = (1+s^2/\lambda^2)^{-1/2}$, with contrast parameter $\lambda > 0$.

The interpolated image can be seen as an albedo-corrected version of the original image, which now satisfies the assumption of a surface with homogeneous reflectance properties.

2.3 3-D Reconstruction – Shape from Shading

Finally, we need to reconstruct the modified image from Section 2.2 within the segmentation region obtained in Section 2.1. For this, we use the method of Vogel et al. [8] incorporating the Phong reflectance model since real-world objects feature non-Lambertian surfaces [11]. The model is formulated in terms of the Hamilton-Jacobi equation

$$\frac{I - k_a I_a}{Q}\mathsf{f}^2 W - k_d I_d e^{-2v} - \frac{W k_s I_s}{Q} e^{-2v}\left(\frac{2Q^2}{W^2} - 1\right)^\alpha = 0\,, \tag{4}$$

where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ is in the image domain, and $u > 0$ with $v := \ln(u)$ is the sought depth map. The other terms in (4) are given as follows. $I := I(\mathbf{x})$ is the brightness normalised to the interval $[0, 1]$, and f is the focal length denoting the distance between the optical centre of the camera and the 2-D retinal plane. The terms Q and W are given as

$$Q := \frac{\mathsf{f}}{\sqrt{x^2 + y^2 + \mathsf{f}^2}}\,, \tag{5}$$

$$W := \sqrt{\mathsf{f}^2 |\nabla v|^2 + (\nabla v \cdot \mathbf{x})^2 + Q^2}\,. \tag{6}$$

Note that in (4), the underlying brightness equation reads as

$$I \;=\; k_a I_a + \sum_{\text{light sources}} \frac{1}{r^2}\big(k_d I_d \cos\phi + k_s I_s (\cos\theta)^\alpha\big)\,. \tag{7}$$

Here, ϕ is the angle between the surface normal at the point $\tilde{u} := (\mathbf{x}, u(\mathbf{x})) \in \mathbb{R}^3$ and the light source direction as seen from $\tilde{u}$. The amount of specular light reflected towards the camera 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 at $\tilde{u}$. The parameter α models the roughness of the material: For $\alpha \to \infty$ one would obtain a model for a perfect mirror. I_a, I_d, and I_s are the intensities of the ambient, diffuse, and specular components of light, respectively. 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 [10].

For solving the PDE (4), we use the algorithm proposed by Breuß et al., for details see [24].

3 Real-World Experiments

In this section, we evaluate our proposed framework on real-world images. Figure 1 (a) shows a picture of a cup taken with a digital camera in our office environment. The image has size 408×306 with quadratic pixels of 1.61 μm side length. The focal length of the camera is 70.2 mm.

Figure 1 (b) shows a reconstruction of this surface using the SfS method of Prados and Faugeras [4]. Note that this is already an advanced SfS method, which uses a perspective projection model on Lambertian surfaces and considers the physical light attenuation term. The parameters used for the reconstruction were $\mathsf{f} = 5435 = 70.2\ mm/1.61\ \mu m$ and $\gamma = 100000$, where γ is the calibration parameter used in the model of Prados et al. [4, 19]. Note that this parameter can be chosen arbitrarily, since it will only scale the reconstruction uniformly in all dimensions.

We can clearly see that the reconstruction fails completely in the background, at the transition from foreground to background, and at textures and highlights on the cup. Now, we demonstrate step by step how our proposed framework helps to improve this reconstruction.

In the next experiment, we perform a segmentation as proposed in Section 2.1. Using the parameters $\nu = 0, \mu = 10$ for Chan-Vese postprocessed by geodesic active contours with $\lambda = 3.6$, we obtain the segmented cup shown in Figure 2

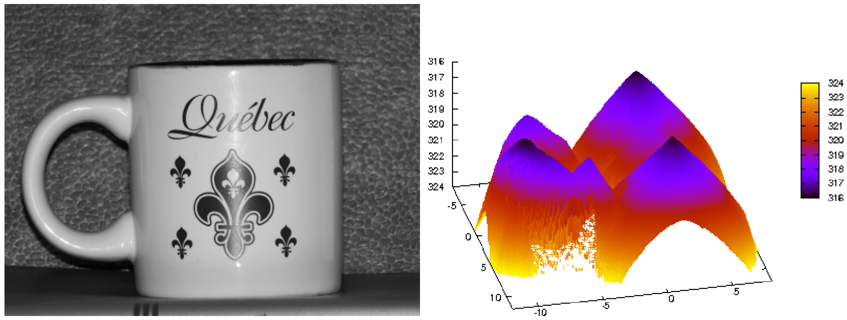

Fig. 1. (a) Photograph of a cup. (b) Lambertian reconstruction.

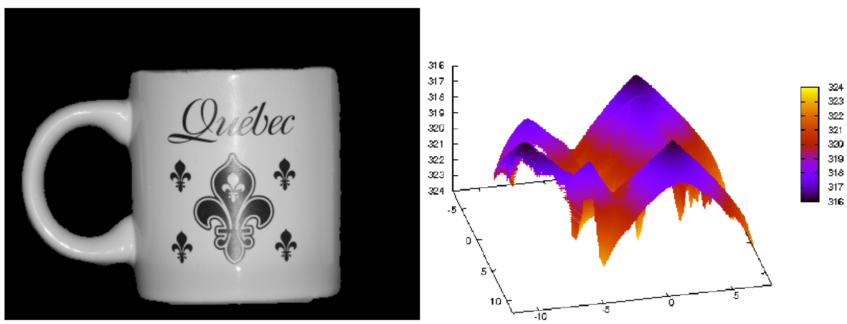

Fig. 2. (a) Segmented version of the cup image. (b) Lambertian reconstruction of the segmented image.

(a). Now we reduce the reconstruction to only this area. The resulting surface using a Lambertian model for reconstruction is shown in Figure 2 (b). Clearly, this improves the reconstruction of the cup. It is still oddly shaped, but on its boundaries, the reconstruction is substantially better.

In the next experiment, we adapt the albedo in the textured regions using the procedure described in Section 2.2. We perform an adaptive thresholding on the image within the cup area, taking a 100×100 window. This gives the inpainting region, which is the black template in Figure 3 (a).

After a morphological erosion of this inpainting region in order to enlarge its size, we apply EED with the parameters $\lambda = 2$ and $\sigma = 0.3$ to inpaint the image there. The inpainted image is shown in Figure 3 (b). This image can be regarded as a constant-albedo version of the original image, within the segmented area. Note that this image still contains specular highlights.

Now, we reconstruct the surface from the segmented and inpainted data. Figure 3 (c) shows the corresponding reconstruction. We still use the Lambertian model by Prados et al here. The shape of the cup obtained by this Lambertian model looks quite reasonable. However, the cup is estimated much too close to

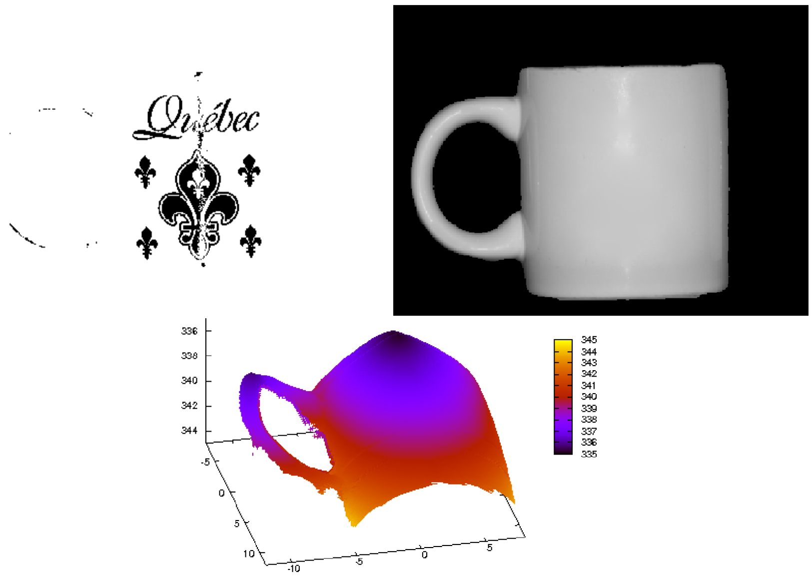

Fig. 3. (a) Inpainting region obtained by adaptive thresholding. (b) Inpainted image. (c) Reconstruction of the inpainted image using a Lambertian model

the camera, in particular at specular highlights. Note that the handle, which is pointing slightly towards the background in the original image, is pulled to the front.

As a final step, we switch to the more advanced SfS model of Vogel et al. [8], which assumes Phong reflectance properties. With the parameters $I_s = I_d = 100000$, $k_d = 0.6$, $k_s = 0.4$, $\alpha = 6$, we obtain the reconstruction shown in Figure 4 (a). The parameters have been estimated manually, where only α and the ratio between k_d and k_s is really relevant. The magnitude of I_s and I_d will only scale the reconstruction. This yields a fairly realistic reconstruction of the cup. Its shape is recovered well, as is its size and the distance to the camera. The handle is now approximately at the correct position, and even at specular highlights the reconstruction is satisfactory. Compared to the results without any preprocessing in Figure 1 (b), the reconstruction quality is improved dramatically. Figure 4 (b) shows the recovered shape rendered with the texture from the input image.

To show the applicability of our framework to other images, we applied it to two other real-world images shown in Figure 5. The impact of the different steps of our framework for these experiments is similar to those of the first experiment. This will be investigated in more detail in future work. The first image shows a computer mouse on a table. Mouse and table obviously have different materials, and the logo of the manufacturer on the mouse has a different colour than the rest of the mouse. The gap between the buttons makes the reconstruction additionally difficult, since we have shadows there, which contradict the model

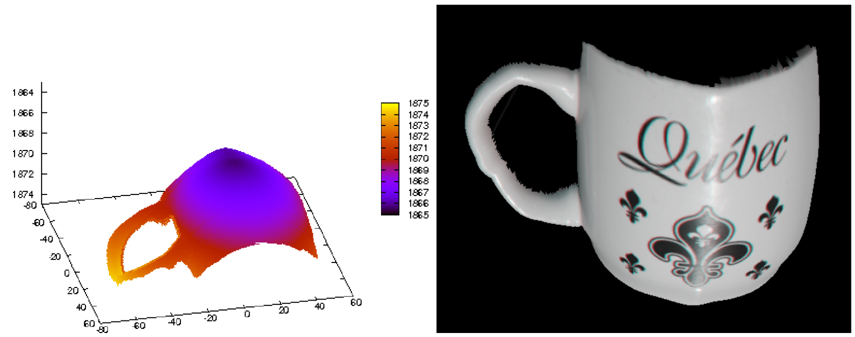

Fig. 4. (a) Reconstruction of the cup using the Phong model. (b) Rendered version of the final reconstruction.

assumptions. Since for this example, foreground and background have similar brightness, we made use of the hue channel in the segmentation step. Figure 6 (a) shows the reconstruction of the mouse. The mouse is recovered very well, including the slots on the buttons, and nearly perfect even at the gaps between the buttons.

Figure 5 (b) is a photograph of a book. The background is quite inhomogeneous and would lead to distortions of the shape if reconstructed unsegmented. The book has some texture on it in different colours and brightnesses. The reconstruction in Figure 6 (b), however, is quite convincing.

4 Conclusions and Outlook

The key message of our paper is the proof that shape from shading is possible under the difficult conditions of real-world images, even without the need to include knowledge-based techniques. This has been achieved by a sophisticated three-stage model that incorporates object segmentation, albedo inpainting and non-Lambertian shape from shading. Our experiments demonstrate that shape from shading has the potential of becoming a serious alternative in computer vision systems when other techniques are difficult to apply. In our future work we will focus on exploring this potential further.

References

1. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1., New York, NY, IEEE Computer Society Press (June 2006) 519–528
2. Horn, B.K.P.: Obtaining shape from shading information. In Winston, P.H., ed.: The Psychology of Computer Vision. McGraw-Hill, New York, NY (1975) 115–155

Fig. 5. (a) Photograph of a computer mouse on a table. (b) Photograph of a book.

3. 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
4. Prados, E., Faugeras, O.: Shape from shading: A well-posed problem? In: Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2., San Diego, CA, IEEE Computer Society Press (June 2005) 870–877
5. Tankus, A., Sochen, N., Yeshurun, Y.: Shape-from-shading under perspective projection. International Journal of Computer Vision **63**(1) (June 2005) 21–43
6. Cristiani, E., Falcone, M., Seghini, A.: Some remarks on perspective shape-from-shading models. In Sgallari, F., Murli, F., Paragios, N., eds.: Scale Space and Variational Methods in Computer Vision. Volume 4485 of Lecture Notes in Computer Science. Springer, Berlin (May-June 2007) 276–287
7. Ahmed, A., Farag, A.: A new formulation for shape from shading for non-Lambertian surfaces. In: Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. IEEE Computer Society Press, New York, NY (June 2006) 17–22
8. Vogel, O., Breuß, M., Weickert, J.: Perspective shape from shading with non-Lambertian reflectance. In Rigoll, G., ed.: Pattern Recognition. Volume 5096 of Lecture Notes in Computer Science., Berlin, Springer (June 2008) 517–526
9. Oren, M., Nayar, S.: Generalization of the Lambertian model and implications for machine vision. International Journal of Computer Vision **14**(3) (1995) 227–251
10. Foley, J., van Dam, A., Feiner, S., Hughes, J.: Computer Graphics: Principles and Practice. Addison-Wesley (1996)
11. Harrison, V.G.W.: Definition and Measurement of Gloss. Printing & Allied Trades Research Association (PATRA) (1945)
12. Smith, W.A.P., Hancock, E.R.: Facial shape-from-shading and recognition using principal geodesic analysis and robust statistics. International Journal of Computer Vision **76**(1) (2008) 71–93

Fig. 6. (a) Reconstruction of the computer mouse. (b) Reconstruction of the book.

13. Zhang, L., Dugas-Phocion, G., Samson, J.S., Seitz, S.M.: Single view modeling of free-form scenes. Journal of Visualization and Computer Animation **13**(4) (2002) 225–235
14. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (February 2001) 266–277
15. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22** (1997) 61–79
16. Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., Yezzi, A.: Conformal curvature flows: from phase transitions to active vision. Archive for Rational Mechanics and Analysis **134** (1996) 275–301
17. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognition **33**(2) (2000) 225–236
18. Weickert, J., Welk, M.: Tensor field interpolation with PDEs. In Weickert, J., Hagen, H., eds.: Visualization and Processing of Tensor Fields. Springer, Berlin (2006) 315–325
19. Prados, E., Camilli, F., Faugeras, O.: A unifying and rigorous shape from shading method adapted to realistic data and applications. Journal of Mathematical Imaging and Vision **25**(3) (2006) 307–328
20. Tschumperlé, D., Deriche, R.: Vector-valued image regularization with PDEs: A common framework for different applications. In: Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1. IEEE Computer Society Press, Madison, WI (June 2003) 651–656
21. Jin, H., Cremers, D., Wang, D., Yezzi, A., Prados, E., Soatto, S.: 3-d reconstruction of shaded objects from multiple images under unknown illumination. International Journal of Computer Vision **76**(3) (2008) 245–256
22. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990) 629–639
23. Weickert, J.: Theoretical foundations of anisotropic diffusion in image processing. Computing Supplement **11** (1996) 221–236
24. Breuß, M., Vogel, O., Weickert, J.: Efficient numerical techniques for perspective shape from shading. In: Algoritmy, Podbanske, Slovakia (March 2009) 11–20