Introducing More Physics into Variational Depth–from–Defocus

Nico Persch(✉), Christopher Schroers, Simon Setzer, and Joachim Weickert

Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
{persch,schroers,setzer,weickert}@mia.uni-saarland.de

Abstract. Given an image stack that captures a static scene with different focus settings, variational depth–from–defocus methods aim at jointly estimating the underlying depth map and the sharp image. We show how one can improve existing approaches by incorporating important physical properties. Most formulations are based on an image formation model (forward operator) that explains the varying amount of blur depending on the depth. We present a novel forward operator: It approximates the thin–lens camera model from physics better than previous ones used for this task, since it preserves the maximum–minimum principle w.r.t. the unknown image intensities. This operator is embedded in a variational model that is minimised with a *multiplicative* variant of the Euler–Lagrange formalism. This offers two advantages: Firstly, it guarantees that the solution remains in the physically plausible positive range. Secondly, it allows a stable gradient descent evolution without the need to adapt the relaxation parameter. Experiments with synthetic and real–world images demonstrate that our model is highly robust under different initialisations. Last but not least, the experiments show that the physical constraints are essential for obtaining more accurate solutions, especially in the presence of strong depth changes.

1 Introduction

The Depth–from–Defocus Problem. Only points with a certain distance to the lens are imaged completely sharp. This distance depends on the focal settings and is described by the *focal plane*. Points with a larger or smaller distance appear blurred, where the amount of blur increases with the object's offset to the focal plane. The range in which points are imaged acceptably sharp is the *depth–of–field* of the camera. In particular macro photography and microscope imaging suffer from a very limited depth–of–field. In these applications, a common remedy is to capture several images by varying the focal settings. Then each of these images differs in the regions that are projected sharply. Given such an image stack, the *depth–from–defocus* problem consists of inferring the underlying topography (depth map) as well as the sharp image as it would

X. Jiang et al. (Eds.): GCPR 2014, LNCS 8753, pp. 15–27, 2014.
DOI: 10.1007/978-3-319-11752-2_2

have been recorded by a pinhole camera. Essentially this corresponds to inverting the imaging process. This inverse problem is ill–posed and much harder to solve than the forward problem that models the image formation. To deal with this ill–posedness, regularisation is required. Variational formulations offer an elegant approach for this task. In our paper we present a novel variational framework that incorporates important physical properties.

Existing Approaches and Related Work. Instead of inverting the physical imaging process, there are approaches that estimate depth using *in–focus* information. They apply a local sharpness criterion, and the depth is assumed to correspond to the slice of the focal stack where the local sharpness achieves its maximum. The *variance method (VM)* [24] for example uses the local variance as a sharpness criterion.

To our knowledge, the first method that estimates depth using *defocus* information goes back to Pentland [19]. He estimates the amount of blurriness of image features or patches and uses this to infer the local depth. While this seminal work requires a completely sharp image as reference, Subbarao [23] describes a possibility to avoid this restriction. Namboodiri and Chaudhuri [16] assume a constant depth and then use the fact that Gaussian blurring can be expressed by linear diffusion. The extension to the more general case, i.e. allowing variations in the depth profile, is treated in many subsequent work [11,17,18,27]. Here the depth–of–field effect is described by means of an isotropic diffusion process with spatially variant diffusivity as it would occur in an inhomogeneous medium. Extensions to anisotropic diffusion processes also exist [9,13]. The depth–from–defocus problem can alternatively be addressed with Markov random fields [3,7]. In the latter work, Bhasin and Chaudhuri consider a scenario restricted to only two different focal planes. They investigate how the *point spread function (PSF)* has to be iteratively corrected in order to represent the energy distribution at depth discontinuities. Also blind deconvolution approaches such as the one by Chan and Wong [6] can be understood as related work in a broader sense.

Most related to our work are the approaches that jointly estimate the sharp image and the depth by minimising a suitable energy. In [10,14] this problem is stated as the minimisation of Csiszár's information divergence between the recorded focal stack and an appropriate model assumption. While the first approach assumes a locally equifocal surface such that the PSF is shift–invariant, the latter one embeds a shift–variant PSF in the imaging model. When regarding a shift–variant PSF as a 4-D function and a shift–invariant as a 2-D one, Aguet et al. [1] propose a compromise between both: They use a shift–invariant 3-D function defined as a family of 2-D Gaussians with varying standard deviation as a PSF. Compared to a 4-D function, this reduces the complexity by incorporating knowledge about how the PSF adapts depending on depth. However, the proposed formulation does not preserve an important physical property, namely the maximum–minimum principle w.r.t. the image intensities. This causes problems, especially at locations where depth changes occur.

Contributions. To address the aforementioned problems, we propose a novel physically motivated forward operator that preserves the maximum–minimum principle w.r.t. the image intensities. This forward operator is derived as an approximation of the thin lens camera model. Given a sharp image and depth information, the thin lens camera model is the established physically based camera model used in computer graphics for generating photorealistic depth–of–field effects. We show how to invert this forward operator within a variational formulation that allows to jointly obtain the unknown depth and intensity values given a focal stack. As it is our goal to preserve important physical properties, we also have to ensure that our solution contains only positive intensity and depth values. To achieve this we employ the multiplicative Euler–Lagrange formalism. Besides enabling us to restrict our solution to physically plausible values, this formalism offers an additional benefit: It allows us to derive an efficient semi–implicit scheme for finding the sought depth and intensity values. This semi–implicit scheme does not require any adaptation of the relaxation parameter.

Organisation of the Paper. In Sect. 2, we discuss image formation models and derive our novel forward operator. Section 3 then explains the variational formulation that effectively allows to invert our forward operator. Experiments show the benefits of our novel model in Sect. 4. We conclude our paper in Sect. 5.

2 Image Formation Models

Let us first obtain a better understanding of forward operators, i.e. image formation models that allow to generate a focal stack given a sharp image and depth information. To this end, we start by briefly discussing the thin lens camera model.

2.1 Image Formation with a Thin Lens

The thin lens imaging model is illustrated in Fig. 1. It uses a thin circular lens with focal length f. This lens is placed in the optical centre $\boldsymbol{\ell}_0$ at a distance v

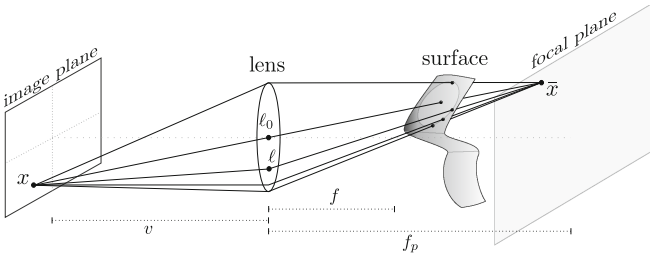

Fig. 1. Thin lens camera model

to the image plane $\Omega_2 \subset \mathbb{R}^2$. Lens and image plane are parallel. The thin lens equation [4]

$$\frac{1}{f_p} = \frac{1}{f} - \frac{1}{v} \tag{1}$$

characterises the imaging process. It involves a virtual *focal plane* that is parallel to the lens at distance f_p. A point $\bar{\boldsymbol{x}}$ within this plane is sharply focused to a single point $\boldsymbol{x}$ on the image plane. For each image point $\boldsymbol{x}$ the corresponding $\bar{\boldsymbol{x}}$ can be obtained by intersecting the ray from $\boldsymbol{x}$ through the optical centre $\boldsymbol{\ell}_0$ with the focal plane.

Generally, the lens focusses a bundle of rays into a single point $\boldsymbol{x}$. This bundle can be described using $\bar{\boldsymbol{x}}$ and all points on the lens. Following [20], we define the *thin lens operator*:

$$\mathcal{F}_{\mathrm{L}}[t, d](\boldsymbol{x}) := \frac{1}{|\mathcal{A}|} \int_{\mathcal{A}} t\Big(T_d\big(\boldsymbol{\ell}, \bar{\boldsymbol{x}}\big)\Big) \, \mathrm{d}\boldsymbol{\ell} \, , \tag{2}$$

where $|\mathcal{A}|$ is the area of the lens and $d : \Omega_2 \to \mathbb{R}^+$ denotes the topography. The function T_d computes the first intersection point of a ray through $\boldsymbol{\ell}$ and $\bar{\boldsymbol{x}}$ with the topography d and t maps these intersection points to intensity values.

A direct simulation of geometric optics is possible with raytracing methods [8]. However, this is computationally very expensive because a large amount of blur requires processing a huge number of rays per pixel. Therefore, instead of directly considering the thin lens camera model, researchers are interested in finding approximations as alternatives for the simulation of photorealistic depth–of–field effects [2,5,21]. Similarly, we are also interested in finding a good approximation of the thin lens operator, however, with the additional requirement that it well fits into a variational framework. To this end, let us first rewrite the thin lens camera model with a spatially variant convolution.

2.2 Spatially Variant Convolution

Given a topography d, the thin lens camera model can be expressed with a spatially variant point spread function (PSF) $H_d : \Omega_2 \times \Omega_2 \to \mathbb{R}_0^+$:

$$\mathcal{F}_{\mathrm{H}}[u, d](\boldsymbol{x}) := \int_{\Omega_2} H_d(\boldsymbol{x}, \boldsymbol{y}) \; u(\boldsymbol{y}) \; \mathrm{d}\boldsymbol{y} \, , \tag{3}$$

where u corresponds to the image as it would be captured with a pinhole camera, and $\boldsymbol{x}$ describes the location within the 2-D image plane. From (2) it is straightforward to see that the thin lens camera model fulfils a maximum–minimum principle w.r.t. t. Accordingly, H_d has to preserve this w.r.t. the intensity values of the sharp image, i.e.

$$\int_{\Omega_2} H_d(\boldsymbol{x}, \boldsymbol{y}) \; \mathrm{d}\boldsymbol{y} = 1 \quad \forall \boldsymbol{x} \in \Omega_2 \, . \tag{4}$$

This guarantees that each intensity value of the resulting image lies between the minimum and maximum intensity value of the sharp image. Equation (3) can be

understood as a weighted average of the sharp image intensities. To obtain the weights of the PSF, raytracing techniques may be used. However, this is similar to computing the thin lens camera model directly. Thus, let us investigate a different more efficient way to approximate the weights of the PSF.

2.3 Approximation of the PSF

In the thin lens model, a point on the surface spreads its intensity to a *circle of confusion* on the image plane [23]. For the moment, let us assume that the surface is equifocal ($d = \text{constant}$), i.e. aligned parallel to the lens. Then Eq. (3) can be expressed in terms of a convolution with a spatially invariant kernel $h_d : \Omega_2 \subset \mathbb{R}^2 \to \mathbb{R}_0^+$ instead of a spatially variant one H_d. This comes down to reducing the PSF from a 4-D to a 2-D function. More precisely, in case of a circular lens, the kernel corresponds to a pillbox function and its radius is related to the constant topography. However, in practise it can be a better choice to use a Gaussian PSF instead of a pillbox when taking into account the wave character of light [19]. The standard deviation of the Gaussian replaces the radius of the pillbox. In the general case of non–constant topographies, the standard deviation changes with the depth of each surface point. Following this idea, Aguet et al. [1] express the imaging process as

$$\mathcal{F}_{\mathrm{U}}[u, d](\boldsymbol{x}, z) := \int_{\Omega_2} h(\boldsymbol{x} - \boldsymbol{y}, z - d(\boldsymbol{y}))\ u(\boldsymbol{y})\ \mathrm{d}\boldsymbol{y}\ , \tag{5}$$

where z represents a given focal plane and $h : \Omega_3 \subset \mathbb{R}^3 \to \mathbb{R}_0^+$. Therewith, they lift h_d from a 2-D function to a 3-D one, composed of 2-D PSFs varying in their standard deviation, along the third dimension.

2.4 Our Modification

The formulation above is problematic if partial occlusions occur, which is expected to happen due to depth changes. The forward operator then effectively performs spatially variant 2-D convolutions with unnormalised kernels. This results in a violation of the maximum–minimum principle w.r.t to the images intensities. To avoid this, we propose to replace (5) by the novel forward operator

$$\mathcal{F}_{\mathrm{N}}[u, d](\boldsymbol{x}, z) := \frac{\mathcal{F}_{\mathrm{U}}[u, d](\boldsymbol{x}, z)}{\int_{\Omega_2} h(\boldsymbol{x} - \boldsymbol{x'}, z - d(\boldsymbol{x'}))\ \mathrm{d}\boldsymbol{x'}}\ . \tag{6}$$

The normalisation function guarantees the maximum–minimum principle, and thus handles regions where partial occlusions appear in a more appropriate way. While this normalisation may look like a small modification at first glance, it can have a large impact on the quality of the result: Fig. 2 depicts the behaviour of the different forward operators. We see that in regions where depth changes are present, partial occlusions appear. Applying an unnormalised forward operator

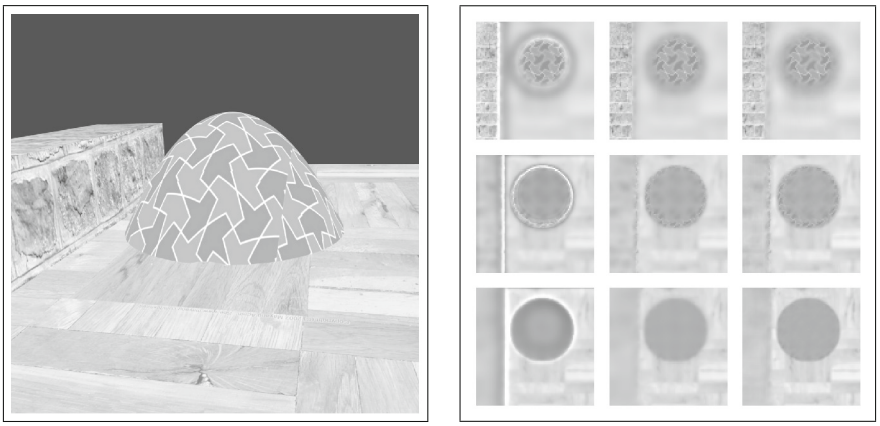

Fig. 2. (a) Left box: Our 3-D synthetic test model. **Right box:** Results of different forward operators (from left to right) with changing focal planes (from top to bottom). **(b) Left column:** Unnormalised forward operator of [1]. **(c) Centre column:** Our normalised forward operator (6). **(d) Right column:** Thin lens camera model (2).

results in bright overshoots followed by dark shadows (Fig. 2(b)). These local violations of the maximum–minimum principle lead to wrong model assumptions, which produce results that are not photorealistic. On the other hand, comparing Fig. 2(c, d) shows that our normalised approach comes very close to the physically well–founded thin lens camera model which allows to create realistic depth–of–field effects.

3 Variational Formulation

So far we have discussed image formation models, i.e. operators that can create stacks of blurred images if we know the sharp image and the depth. In this section we are interested in inverting this process, i.e. given an image stack, we wish to jointly estimate the depth map and the sharp image.

3.1 Variational Model

Let $u_{\mathrm{R}} : \Omega_3 \to \mathbb{R}^+$ be the stack of recorded 2-D images that vary in their focal plane. The sought, sharp image $u : \Omega_2 \to \mathbb{R}^+$ in combination with the depth map $d : \Omega_2 \to \mathbb{R}^+$ can then be estimated as a minimiser of the energy

$$E(u, d) = M(u, d) + \alpha\, S(|\nabla d|). \tag{7}$$

The data term M enforces the similarity between the recorded stack and the forward operator applied to the unknown sharp image u and depth d. To penalise

deviations from the model assumptions we choose a quadratic cost function which is optimal for Gaussian distributed noise:

$$M(u,d) = \int_{\Omega_3} \Big(\underbrace{u_{\mathrm{R}} - \mathcal{F}_{\mathrm{N}}[u,d]}_{=:e[u,d]} \Big)^2 \mathrm{d}\boldsymbol{x}\ \mathrm{d}z\ . \tag{8}$$

Exactly like $\mathcal{F}_{\mathrm{U}}$, our forward operator $\mathcal{F}_{\mathrm{N}}$ from (6) is linear in u but nonlinear in d, and the data term is convex in u but nonconvex in d. Especially in homogeneous regions, a minimiser of the data term alone is non–unique. To avoid such ambiguities, we add a regularisation term S that penalises large gradient magnitudes in the depth field:

$$S(|\boldsymbol{\nabla} d|) = \int_{\Omega_2} \Psi(|\boldsymbol{\nabla} d|^2)\ \mathrm{d}\boldsymbol{x}\ , \tag{9}$$

where $\Psi : \mathbb{R} \to \mathbb{R}^+$ is a positive increasing function imposing (piecewise) smoothness. For the results in Sect. 4, we employ the Whittaker–Tikhonov penaliser $\Psi(s^2) = s^2$ which corresponds to homogeneous diffusion [25,29]. Other regularisers such as total variation (TV) [22] are also appropriate. The amount of smoothness can be steered by the regularisation parameter $\alpha > 0$.

3.2 Minimisation

Euler-Lagrange Equations. A minimiser (u,d) of the energy (7) must necessarily fulfill the *Euler–Lagrange* equations

$$\frac{\delta E}{\delta u} = 0 \quad \text{and} \quad \frac{\delta E}{\delta d} = 0 \tag{10}$$

and its corresponding natural boundary conditions. The established approach to find the functional derivatives $\frac{\delta E}{\delta u}$ and $\frac{\delta E}{\delta d}$ is given by applying the classical (additive) Euler–Lagrange formalism [12]. To obtain the functional derivative $\frac{\delta E}{\delta u}$, one uses the definition

$\langle \frac{\delta E}{\delta u}, v \rangle = \delta_v E$, where $\langle \cdot, \cdot \rangle$ denotes the standard inner product, and

$$\delta_v E := \frac{\partial}{\partial \epsilon} E(u + \epsilon v,\ d)|_{\epsilon=0} \tag{11}$$

acts like a "directional derivative" in the direction of a function v. Then we obtain

$$\frac{\delta E}{\delta u}(\boldsymbol{x}) = -2\ \Big(\overline{e} * h^* \Big)(\boldsymbol{x}, d(\boldsymbol{x})) \tag{12}$$

with E from Eq. (7). Here we have introduced the abbreviation $\overline{e}\ := N^{-1} \cdot e$, where e is related to the data term (8) and N corresponds to the normalisation function, i.e. the denominator in (6). The operator $*$ expresses a 3-D convolution, and $h^*(x) := h(-x)$. Analogously applying the same formalism w.r.t. the depth d allows to compute the functional derivative

$$\frac{\delta E}{\delta d}(\boldsymbol{x}) = 2\,\Big(\Big(\overline{e} * h_z^*\Big)(\boldsymbol{x}, d(\boldsymbol{x}))\cdot u - \Big(\big(\overline{e}\cdot\mathcal{F}_{\mathrm{N}}[u,d]\big) * h_z^*\Big)(\boldsymbol{x}, d(\boldsymbol{x})) - \alpha\cdot\operatorname{div}\Big(\Psi'(|\boldsymbol{\nabla} d|^2)\;\boldsymbol{\nabla} d\Big)\Big). \tag{13}$$

Enforcing Positivity. Since negative intensities as well as negative depth values are not physically plausible, we would like to modify our approach in such a way that both quantities are strictly constrained to be positive. The *multiplicative* Euler–Lagrange formalism offers an interesting and efficient way to achieve this [28]. Here, a multiplicative perturbation is used instead of an additive one. Thus, we consider

$$\delta_v^* E := \frac{\partial}{\partial\epsilon} E(u + \epsilon u\cdot v,\, d)|_{\epsilon=0} \tag{14}$$

for minimisation w.r.t. u, and an analog expression for d. This gives the following functional derivatives:

$$\frac{\delta^* E}{\delta u} = u\cdot\frac{\delta E}{\delta u} \quad\text{and}\quad \frac{\delta^* E}{\delta d} = d\cdot\frac{\delta E}{\delta d}. \tag{15}$$

There are two different ways to understand why the multiplicative Euler–Lagrange formalism restricts the solution to positive values [28]: The first explanation interprets the multiplicative Euler–Lagrange formalism via the reparametrisations $u = \exp(w)$ and $d = \exp(z)$. Moving unwanted values to infinite distance is the second explanation. To this end, one can show that the multiplicative functional gradients $\frac{\delta^* E}{\delta u}$ and $\frac{\delta^* E}{\delta d}$ occur within the additive formalism when one replaces the Euclidean metric $\mathrm{d}u$ by a hyperbolic one, i.e. $\mathrm{d}u/u$.

3.3 Discretisation and Implementation

The multiplicative approach presented above does not only guarantee the positivity of our solution, it also enables us to introduce an efficient semi–implicit iteration scheme. To this end, we consider a gradient descent scheme with the multiplicative gradient $\frac{\delta^* E}{\delta u}$ from Eq. (15):

$$\frac{u^{k+1} - u^k}{\tau} = 2\; u^{k+1}\Big(\overline{e}^k * h^*\Big)(\boldsymbol{x}, d)\;, \tag{16}$$

where τ is the relaxation parameter, and the upper index denotes the iteration level. For the multiplicative gradient descent w.r.t. the depth map d, we use the semi-implicit approach

$$\frac{d^{k+1} - d^k}{\tau} = -2\,\Big(\Big(\overline{e}^k * h_z^*\Big)(\boldsymbol{x}, d^k)\; u - \Big(\Big(\overline{e}^k\cdot\mathcal{F}_{\mathrm{N}}[u, d^k]\Big) * h_z^*\Big)(\boldsymbol{x}, d^k)\Big)\cdot d^{k+1} + 2\alpha\cdot\operatorname{div}\Big(\Psi'(|\boldsymbol{\nabla} d^k|^2)\;\boldsymbol{\nabla} d^{k+1}\Big)\cdot d^k\;. \tag{17}$$

This semi–implicit scheme is less sensitive sensitive w.r.t. the relaxation parameter such that it can remain fixed during the iterations. Therefore, we can

refrain from backtracking line–search or complicated methods such as Brent's algorithm that combines the bisection method, the secant method and inverse quadratic interpolation. Such methods are required when using the standard additive Euler–Lagrange formalism for this problem and are avoided by our approach.

Since we deal with digital images, we replace continuous functions by their discrete counterparts and derivatives by finite differences. The 3-D convolution is implemented in the Fourier domain, using the Fast Fourier Transform and the convolution theorem. While Eq. (16) can be solved directly, Eq. (17) requires to solve a nonsymmetric linear system of equations. We solve the latter one iteratively with a Jacobi algorithm. We use an alternating minimisation scheme [15] as it is commonly used e.g. in blind deconvolution problems [6]: Keeping the solution for one sub–problem fixed (e.g. recovering the sharp image u), the other problem (e.g. the estimation of the depth) is solved with a fixed number of gradient descent steps. After that, roles are exchanged. To account for the nonconvexity, we apply a coarse–to–fine strategy where the solution of the downsampled problem serves as initialisation on the next finer one.

4 Experiments

Synthetic Data. In our first experiment, we generated a stack of images with varying focal planes. This is achieved by rendering the 3-D model from Fig. 2(a) with the thin lens camera model (lens diameter $D = 2.69\,\text{cm}$, distance to image plane $v = 35\,\text{mm}$). In total we have rendered 20 images where the distance of the focal plane to the lens changed in equidistant steps from $f_p = 3\,\text{cm}$ to $f_p = 7\,\text{cm}$. Figure 2(d) shows 3 different slices of this rendered focal stack.

In Fig. 3, we compare the results of different approaches to estimate the topography. For the variance method two undesired hills in front of and behind the hemisphere arise (Fig. 3(a)). Using the forward operator $\mathcal{F}_U$ leads to a violation

Table 1. Quantitative comparison. Error measurement of the estimated topography and the sharp image to its ground–truth. We consider the mean squared error (MSE) as well as the structural similarity (SSIM) [26]. We compare the variance method (VM) with and without Gaussian post–smoothing with variance σ, the operator $\mathcal{F}_\text{U}$, and our max–min–preserving imaging model $\mathcal{F}_\text{N}$. The latter two are either initialised with a constant depth or an estimation of the VM.

Method		VM		$\mathcal{F}_\text{U}$		Ours	
		$\sigma = 0$	$\sigma = 4$	Constant	VM	Constant	VM
Depth	MSE	2.83	1.10	21.66	2.26	0.77	0.58
	SSIM	0.98	0.93	0.94	0.99	1.00	1.00
Image	MSE	48.33	46.38	124.52	52.17	49.09	45.17
	SSIM	0.87	0.87	0.67	0.87	0.90	0.92

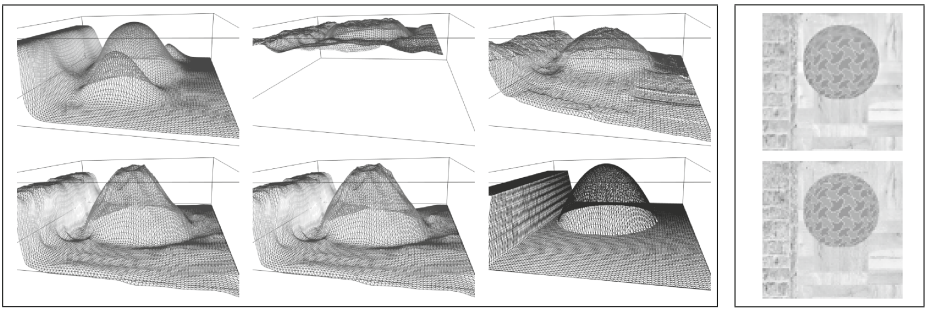

Fig. 3. Visual comparison. **Left box:** *In reading order* **(a)** Variance Method (VM) with Gaussian post–smoothing (patch-size = 6, $\sigma = 4.0$). **(b)** $\mathcal{L}_{\mathrm{U}}$ ignoring normalisation, initialised with constant depth ($\alpha = 45$). **(c)** Dito initialised with VM. **(d)** Ours $\mathcal{L}_{\mathrm{N}}$ initialised with constant depth ($\alpha = 150$). **(e)** Dito with VM initialisation. **(f)** Ground–truth of the topography. **Right box: (g) Top:** Estimated sharp image. **(h) Bottom:** Ground–truth of the sharp image.

of the model assumptions at large depth changes (Fig. 3(b,c)). It implicitly introduces a regularisation that erroneously prefers only smooth changes of depth. While Fig. 3(b) has been initialised with a constant depth map, Fig. 3(c) has been initialised with the result of the variance method. Apparently the initialisation strongly affects the outcome. Due to wrong assumptions at depth changes, it is not possible to converge to a reasonable solution with a constant initialisation. In contrast to that, the reconstructions with our normalised forward operator $\mathcal{F}_N$ shown in the Fig. 3(d, e) do not suffer from this effect. Our approach is less sensitive to the initialisation, yielding similar error values and reconstructions in both cases. In fact, our model entails a physically plausible behaviour also at depth changes. We obtain reconstructions that match the ground truth in a better way, concerning both the depth and the sharp image; cf. Table 1. Figure 3(g) also shows the estimated sharp image with our approach. It closely resembles the ground truth.

Real–World Data. In our second experiment, we test our approach on a real world focal stack showing a house fly eye (see Fig. 4(a)). We employ a coarse–to–fine strategy with the variance method as initialisation on the coarsest grid. To demonstrate the performance of our approach on this real word data set, we present the reconstructed sharp image along with the estimated depth profile (Fig. 4(b)). For the illustration of the depth profile a grey value coding is used: The brighter the grey value the larger the distance of the object to the lens. In our result, fine structures are clearly visible in the sharp image and the depth profile. This can be seen in Fig. 4(b) when considering the small hair, for example.

In our third experiment, we use an image stack consisting of 22 frames that depict a coffee bean. Figure 5(a) shows 3 different slices of this focal stack. In this experiment we employ a coarse–to–fine strategy again but this time with

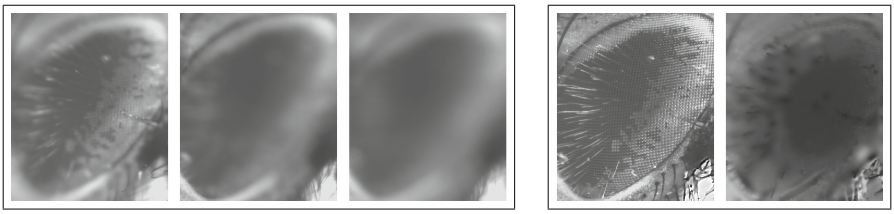

Fig. 4. House fly eye (provided by the Biomedical Imaging Group EPFL, Lausanne, Switzerland). **(a) Left box:** 3 out of 21 images of a focal stack of a house fly eye. **(b) Right box:** Estimated sharp image and topography for $\alpha = 25$.

Fig. 5. Coffee bean (provided by the Computer Graphics Group, MPI for Informatics, Saarbrücken, Germany). **(a) Left box:** 3 out of 22 images of the focal stack of a coffee bean. **(b) Right box:** Reconstructed sharp image along with the estimated depth ($\alpha = 160$).

a constant depth value as initialisation on the coarsest grid. Even with such a crude initialisation, accurate results are possible.

5 Conclusions

We have shown the benefits of introducing two physical constraints into variational depth–from–defocus models: a maximum–minimum principle w.r.t. the unknown sharp image, and the positivity of the sought image intensities and the depth values. Our resulting model offers clear advantages especially in the presence of strong depth variations. Moreover, we advocate to replace the traditional Euler–Lagrange formalism by a multiplicative variant, whenever positivity is to be preserved. It is our hope that both physically refined modelling and multiplicative calculi will receive more popularity in future computer vision models.

Acknowledgements. Our research has been partly funded by the Deutsche Forschungsgemeinschaft (DFG) through a Gottfried Wilhelm Leibniz prize for Joachim Weickert and the Cluster of Excellence *Multimodal Computing and Interaction*.

References

1. Aguet, F., Van De Ville, D., Unser, M.: Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy. IEEE Trans. Image Process. **17**(7), 1144–1153 (2008)

2. Barsky, B.A., Kosloff, T.J.: Algorithms for rendering depth of field effects in computer graphics. In: Proceedings of WSEAS International Conference on Computers, pp. 999–1010. World Scientific and Engineering Academy and Society, Heraklion, July 2008
3. Bhasin, S., Chaudhuri, S.: Depth from defocus in presence of partial self occlusion. In: Proceedings of IEEE International Conference on Computer Vision, vol. 1, pp. 488–493. Vancouver, Canada, July 2001
4. Born, M., Wolf, E.: Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, 4th edn. Pergamon Press, Oxford (1970)
5. Cant, R., Langensieoen, C.: Creating depth of field effects without multiple samples. In: Proceeding of IEEE International Conference on Computer Modelling and Simulation, pp. 159–164. Cambridge, UK, Mar 2012
6. Chan, T.F., Wong, C.K.: Total variation blind deconvolution. IEEE Trans. Image Process. **7**, 370–375 (1998)
7. Chaudhuri, S., Rajagopalan, A.: Depth from Defocus: A Real Aperture Imaging Approach. Springer, Berlin (1999)
8. Cook, R.L., Porter, T., Carpenter, L.: Distributed ray tracing. In: Computer Graphics, SIGGRAPH '84, pp. 137–145. ACM, Minneapolis, Jul 1984
9. Favaro, P., Osher, S., Soatto, S., Vese, L.: 3D shape from anisotropic diffusion. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Madison, USA, Jun 2003
10. Favaro, P., Soatto, S.: Shape and radiance estimation from the information divergence of blurred images. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1842, pp. 755–768. Springer, Heidelberg (2000)
11. Favaro, P., Soatto, S., Burger, M., Osher, S.: Shape from defocus via diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 518–531 (2008)
12. Gelfand, I.M., Fomin, S.V.: Calculus of Variations. Dover, New York (2000)
13. Hong, L., Yu, J., Hong, C., Sui, W.: Depth estimation from defocus images based on oriented heat-flows. In: Proceedings of IEEE International Conference on Machine Vision, pp. 212–215. Dubai, UAE (2009)
14. Jin, H., Favaro, P.: A variational approach to shape from defocus. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part II. LNCS, vol. 2351, pp. 18–30. Springer, Heidelberg (2002)
15. Luenberger, D., Ye, Y.: Linear and Nonlinear Programming, 3rd edn. Springer, New York (2008)
16. Namboodiri, V.P., Chaudhuri, S.: Use of linear diffusion in depth estimation based on defocus cue. In: Chanda, B., Chandran, S., Davis, L.S. (eds.) Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing, pp. 133–138. Allied Publishers Private Limited, Kolkata (2004)
17. Namboodiri, V.P., Chaudhuri, S.: On defocus, diffusion and depth estimation. Pattern Recogn. Lett. **28**(3), 311–319 (2007)
18. Namboodiri, V., Chaudhuri, S., Hadap, S.: Regularized depth from defocus. In: Proceedings of IEEE International Conference on Image Processing, San Diego, USA, pp. 1520–1523, Oct 2008
19. Pentland, A.P.: A new sense for depth of field. IEEE Trans. Pattern Anal. Mach. Intell. **9**(4), 523–531 (1987)
20. Pharr, M., Humphreys, G.: Physically Based Rendering: From Theory to Implementation. Morgan Kaufmann, San Francisco (2004)
21. Rokita, P.: Fast generation of depth of field effects in computer graphics. Comput. Graph. **17**(5), 593–595 (1993)

22. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
23. Subbarao, M.: Parallel depth recovery by changing camera parameters. In: Proceedings of IEEE International Conference on Computer Vision, Washington, USA, pp. 149–155, Dec 1988
24. Sugimoto, S.A., Ichioka, Y.: Digital composition of images with increased depth of focus considering depth information. Appl. Optics **24**(14), 2076–2080 (1985)
25. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Sov. Math. Doklady **4**, 1035–1038 (1963)
26. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
27. Wei, Y., Dong, Z., Wu, C.: Global depth from defocus with fixed camera parameters. In: Proceedings of IEEE International Conference on Mechatronics and Automation, Changchun, China, pp. 1887–1892, Aug 2009
28. Welk, M., Nagy, J.G.: Variational deconvolution of multi-channel images with inequality constraints. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (eds.) IbPRIA 2007. LNCS, vol. 4477, pp. 386–393. Springer, Heidelberg (2007)
29. Whittaker, E.T.: A new method of graduation. Proc. Edinburgh Math. Soc. **41**, 65–75 (1923)

http://www.springer.com/978-3-319-11751-5

Pattern Recognition
36th German Conference, GCPR 2014, Münster, Germany,
September 2-5, 2014, Proceedings
Jiang, X.; Hornegger, J.; Koch, R. (Eds.)
2014, XVIII, 775 p. 310 illus., Softcover
ISBN: 978-3-319-11751-5