Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 343

A Focus Fusion Framework with Anisotropic Depth Map Smoothing

Madina Boshtayeva, David Hafner and Joachim Weickert

Saarbrücken 2014

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 343
submitted: July 11, 2014

A Focus Fusion Framework with Anisotropic Depth Map Smoothing

Madina Boshtayeva
Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
boshtayeva@mia.uni-saarland.de

David Hafner
Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
hafner@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Abstract

Focus fusion is the task of combining a set of images focused at different depths into a single image that is entirely in-focus. The crucial point of all focus fusion methods is the decision about the in-focus areas. To this end, we present a general framework for focus fusion that introduces a modern regularisation strategy on these per-pixel decisions. We assume that neighbouring pixels in the fused image belong to similar depth layers. Following this assumption, we smooth the depth map with a sophisticated anisotropic diffusion process combined with a robust data fidelity term. The experiments with synthetic and real-world data demonstrate that our new model yields a better quality than several existing focus fusion methods. Moreover, our methodology is general and can be applied to improve many fusion approaches.

1 Introduction

In applications such as macro photography or optical microscopy, the limited depth of field of standard cameras poses a severe problem: It is not possible to capture an image that is totally in focus. A common remedy is to acquire a set of images while varying the position of the focal plane. In this way, the image stack contains all required information to produce a single image that is sharp everywhere. The task of combining these images of the focus stack into an all-in-focus composite is called *focus fusion*.

1.1 Related Work

We categorise focus fusion techniques into two main groups: The methods in the first group work on multiscale decompositions of the images. In the first step, they apply a multiscale transformation of the complete image stack. Next all images are combined in the transform domain by selecting the coefficients that have the highest probability of belonging to in-focus areas. Finally the composed multiresolution representation is transformed back to the spatial domain. The result is the all-in-focus image. In this class, the pioneering work is the Gaussian and Laplacian pyramid-based method by Ogden et al. [1]. Later Burt and Kolczynski [2] generalised it to alternative pyramid representations, and Petrovic and Xydeas [3] proposed a multiresolution gradient map representation. Also wavelet-based methods belong to this class of algorithms. Here a first approach with application to focus fusion was published by Li et al. [4]. Modifications and extensions have, for instance, been proposed by Forster et al. [5] or Lewis et al. [6]. In [7], Zhang

and Blum present a generic framework for multiscale image fusion and compare different approaches. All these multiresolution-based techniques share the same constitutional drawback: Performing the fusion in the transform domain may change the intensity values and create artificial colours. This produces undesirable artefacts in the fused result.
To overcome this drawback, the algorithms of the second group work in the image domain. Here the basic idea is first to select the regions from all frames that are in focus, and then to combine them to one composite. Recently, many methods for focus fusion have been reported in the literature which employ machine learning techniques to build a sharp image: Wu et al. [8] propose a method using a hidden Markov model, Wan et al. [9] employ principal component analysis for the focus fusion task, and Wang et al. [10] use pulse coupled neural networks to obtain a sharp composite. All of these models work well in the image domain. However, in general, operating in the image domain can cause unpleasant visible seams that appear when simply arranging the identified in-focus areas in a mosaic-like fashion. To tackle these artefacts, Pop et al. [11] as well as Wang et al. [12] explicitly model a smoothness constraint of the resulting composite image by means of a partial differential equation (PDE). Unfortunately this may also cause smoothing of important image structures such that the resulting images appear blurred and not sharp everywhere.
Hence, researchers came up with the idea of not applying the smoothness constraint on the resulting image itself, but on the per-pixel decision of the in-focus areas: In [13, 14, 15, 16, 17, 18] the authors determine an initial decision map by means of a specific sharpness criterion. Subsequently they segment these maps into regions that belong to the same input frames. These segments are then used to fuse the input images to an all-in-focus composite, or even to recover an underlying 3-D surface. Agarwala et al. [13] use graph-cut optimisation to segment different in-focus areas and fuse the input images in the gradient domain. Šroubek et al. [14] propose a level-set segmentation on the decision map solving a suitable PDE. Similarly, in the method of Li and Yang [15] the images are segmented with the normalized cut method; this method is further extended in the work of Liu et al. [19].
There are many other approaches that offer improvements to focus fusion techniques and algorithms: Muhammad and Choi [20] derive the optimal sampling to obtain a reasonable 3-D shape. In [16], Shim and Choi introduce a novel iterative algorithm to reconstruct the 3-D shape. Mahmood et al. [17] propose a combination of different focus measures for constructing the optimal decision map through genetic programming. Mahmood and Choi [18] employ 3-D anisotropic diffusion to enhance the input images, and in turn, to obtain an accurate decision map. Staying with the idea of operating on

decision maps, Bae et al. [21] apply bilateral filtering to this decision map in a related context. However, they do not consider the focus fusion task, but perform a defocus magnification given a single image. While most of the related research aims for piecewise constant solutions, we aim to achieve more realistic piecewise smooth decision maps to the focus fusion problem.

1.2 Contributions

In our work, we follow the idea of processing an initial decision map. However, instead of a segmentation-based approach, we introduce a modern regularisation technique which aims to smooth the initial decision map. Moreover, since each image is sharp at a particular depth value, we interpret the decision maps as depth maps. Consequently, we aim for *piecewise smooth* solutions as opposed to *piecewise constant* ones that are obtained by segmentation-based methods. In this way, we are even able to adequately handle pixels that are never captured totally in-focus since they lie between two focal planes. The explicit modelling of smooth transitions in depth provides not only more accurate depth maps, but also counteracts unpleasant seams in the final image. In our approach, we formulate a similarity to a precomputed depth map or even to a composite of multiple depth maps by a robust data term and combine it with a modern adaptive regularisation technique: Our *joint image- and depth-driven* diffusion is guided by the structures of the evolving all-in-focus image, while the amount of smoothing is determined by the depth map gradients.
In the present paper, we extend our conference publication [22] in several aspects:

(i) In [22], we applied the gradient magnitude as indicator of sharp image regions. However, our method is very general and not limited to this specific choice: It creates a high quality depth map using one or multiple depth maps that can been precomputed with various sharpness measures. Our experiments demonstrate this by means of six measures.

(ii) In the conference paper, we computed the solution of our model via gradient descent, i.e. as the steady state ($t \to \infty$) of a parabolic PDE. Here we present an alternative elliptic formulation and solve the resulting system of equations with a modern well-parallelisable algorithm implemented on GPU. In this way, we reduce the runtime of our approach significantly.

(iii) Last but not least, we conduct a thorough evaluation of our method on synthetic and various real-world focus stacks. We show the perfor-

mance of the proposed nonlinear anisotropic diffusion in comparison to the linear isotropic one. We demonstrate the flexibility and general applicability of our technique, and compare the results with several focus fusion methods from the literature.

1.3 Organisation

Our paper is organised as follows: In Section 2, we introduce our diffusion-based approach and explain its algorithmic realisation in full detail. Section 3 illustrates the performance of our method on synthetic as well as on real-world experiments. This includes a comparison to several focus fusion approaches. Finally, we conclude the paper with a summary and an outlook in Section 4.

2 Our Focus Fusion Framework

Let $f(\boldsymbol{x}, z)$ be a 3D volume where $\boldsymbol{x} := (x, y)^\top$ denotes the location within a rectangular image domain $\Omega \subset \mathbb{R}^2$ and $z \in \mathbb{R}$ the depth. We interpret the K input images $f(\boldsymbol{x}, z_k)$ with $k = 1, \ldots, K$ as equidistant slices of this volume. Our goal is to find a depth map $d(\boldsymbol{x})$ that selects for each location $\boldsymbol{x}$ the frame that is in focus. To this end, our focus fusion framework consists of three main parts: In the first step, we select a single or multiple in-focus measures to identify sharp image regions. This allows to construct the initial depth map or depth maps, respectively. In the core part of our method, we model a similarity to these depth maps and combine it with an anisotropic smoothing technique. Finally, the computed depth map allows to fuse the images to an all-in-focus composite in a straightforward way.

2.1 Initial Depth Map

To determine which image regions are in focus, various sharpness measures have been proposed; see e.g. [23, 24, 25, 26, 27, 28, 29, 30, 31]. All these approaches share the idea that the extrema of a sharpness measure correspond to the image that is in focus. We store this information in the initial depth map d^{init}: It specifies for each pixel the depth value where it is at most in focus.

2.1.1 In-Focus Measures

Our general framework is not limited to any specific choice. To illustrate this, we consider six different in-focus measures that are summarised in Ta-

Table 1: Overview of applied in-focus measures to compute the initial depth map d^{init}.

measure	formula
gradient magnitude	$m_1 = \lvert\boldsymbol{\nabla} f_\sigma\rvert$
norm of the Laplacian	$m_2 = \lvert\Delta f_\sigma\rvert$
Frobenius norm of the Hessian	$m_3 = \lVert\boldsymbol{\mathcal{H}} f_\sigma\rVert_F$
trace of the structure tensor	$m_4 = \operatorname{tr}\left(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma)\right)$
determinant of the structure tensor	$m_5 = \det\left(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma)\right)$
variance	$m_6 = \frac{1}{\lvert\mathcal{N}(\boldsymbol{x})\rvert}\int_{\mathcal{N}(\boldsymbol{x})}\left(f_\sigma(\tilde{\boldsymbol{x}}) - \mu(\boldsymbol{x})\right)^2 \mathrm{d}\tilde{\boldsymbol{x}}$

ble 1. To handle noise, we presmooth the images with a Gaussian of standard deviation σ. Consequently, we compute the following sharpness criteria on smoothed versions $f_\sigma(\boldsymbol{x}, z_k)$ of the input images:

- *Gradient Magnitude*
  The first sharpness criterion is the gradient magnitude

$$m_1(\boldsymbol{x}, z_k) = \left|\boldsymbol{\nabla} f_\sigma(\boldsymbol{x}, z_k)\right|, \tag{1}$$

  where $\boldsymbol{\nabla} := (\partial_x, \partial_y)^\top$ represents the 2D spatial gradient operator, and $|.|$ denotes the Euclidean norm.

- *Norm of the Laplacian*
  One measure of the second order variation is the absolute value of the Laplacian $\Delta f_\sigma = \partial_{xx} f_\sigma + \partial_{yy} f_\sigma$:

$$m_2(\boldsymbol{x}, z_k) = \left|\Delta f_\sigma(\boldsymbol{x}, z_k)\right|. \tag{2}$$

- *Frobenius Norm of the Hessian*
  As an additional sharpness criterion, we also investigate the Frobenius norm $\|.\|_F$ of the spatial 2×2 Hessian:

$$m_3(\boldsymbol{x}, z_k) = \lVert\boldsymbol{\mathcal{H}} f_\sigma\rVert_F = \sqrt{(\partial_{xx} f_\sigma)^2 + 2\,(\partial_{xy} f_\sigma)^2 + (\partial_{yy} f_\sigma)^2}\,. \tag{3}$$

- *Trace of the Structure Tensor*
  Following [32, 33] we use the structure tensor [34], namely its trace, to identify sharp image regions. Let us consider the structure tensor

$$\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma) = G_\rho * \left(\boldsymbol{\nabla} f_\sigma(\boldsymbol{x}, z_k)\ \boldsymbol{\nabla} f_\sigma^\top(\boldsymbol{x}, z_k)\right)\ , \tag{4}$$

where $*$ denotes the convolution operator and G_ρ a Gaussian of standard deviation ρ. Then the sharpness criterion is given by

$$m_4(\boldsymbol{x}, z_k) = \operatorname{tr}\left(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma)\right) . \tag{5}$$

We can interpret the trace tr of the structure tensor as a sum of its eigenvalues. Thus, this measure provides useful information about image features such as edges or corners.

- *Determinant of the Structure Tensor*
  Similarly to the previous measure, we consider the determinant of the structure tensor as a sharpness measure:

$$m_5(\boldsymbol{x}, z_k) = \det\left(\boldsymbol{J}_\rho(\boldsymbol{\nabla} f_\sigma)\right) . \tag{6}$$

  This sharpness criterion can be seen as the product of the eigenvalues of the structure tensor. Thus, here both eigenvalues should have non-zero values. In turn, two large eigenvalues of the structure tensor represent corners. Therefore, this in-focus measure can be interpreted as a corner detector.

- *Variance*
  Another frequently applied sharpness criterion is the local variance in some neighbourhood $\mathcal{N}(\boldsymbol{x})$:

$$m_6(\boldsymbol{x}, z_k) = \frac{1}{|\mathcal{N}(\boldsymbol{x})|} \int_{\mathcal{N}(\boldsymbol{x})} \left(f_\sigma(\tilde{\boldsymbol{x}}) - \mu(\boldsymbol{x})\right)^2 \mathrm{d}\tilde{\boldsymbol{x}} , \tag{7}$$

  where $\mu(\boldsymbol{x})$ is the mean value in $\mathcal{N}(\boldsymbol{x})$, i.e. $\mu(\boldsymbol{x}) = \frac{1}{|\mathcal{N}(\boldsymbol{x})|} \int_{\mathcal{N}(\boldsymbol{x})} f_\sigma(\tilde{\boldsymbol{x}}) \mathrm{d}\tilde{\boldsymbol{x}}$.

These six sharpness measures are implemented in a straightforward way using central finite difference approximations. For each pixel, we determine the frame where the in-focus measure takes its maximal value. This yields the initial depth maps d^{init} that allow to reconstruct an all-in-focus image.

2.1.2 Confidence function

Two types of locations cause severe problems while construction this initial depth map: homogeneous regions that hardly have any texture, and regions that are never in focus. Here the sharpness criteria attain small values in all frames. Thus, the decision of the depth value is highly influenced by noise. To tackle this problem, we use a confidence function $c(\boldsymbol{x}) : \Omega \to \mathbb{R}$ that

specifies the quality of the computed depth map. As in [22, 35], we define a binary confidence function by

$$c(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad m(\boldsymbol{x}, d^{\text{init}}(\boldsymbol{x})) > \theta\,, \\ 0 & \text{else}\,. \end{cases} \tag{8}$$

With this formulation, we treat values of a generic measure m that are larger than the threshold parameter $\theta \geq 0$ as reliable. In contrast, smaller values are discarded since we assume they are too much influenced by noise.

2.2 Depth Map Smoothing

The discussed confidence function helps to specify the quality of the initial estimation and in this way, to detect unreliable depth values. However, one still has to specify which depth values should be used at pixels that are identified as unreliable. Moreover, the confidence function is only an indicator of the quality: The depth map may still be erroneous or noisy.
To handle these issues, we propose a diffusion-based approach that allows to compute a dense and smooth depth map, given a combination of the initial maps d_n^{init} and their confidence functions c_n. We pursue the following idea: We assume a spatial continuity between parts selected from different frames, i.e. neighbouring pixels most probably should be chosen from a similar depth level. As a first step to explain our model, we consider the following energy functional:

$$E(d) = \int_\Omega \Big(\sum_{n=1}^{N} c_n \cdot \Psi\big((d - d_n^{\text{init}})^2\big) + \alpha\, |\boldsymbol{\nabla} d|^2 \Big) \mathrm{d}\boldsymbol{x}\,, \tag{9}$$

where the regularisation parameter α allows to steer the impact of the smoothness term $|\boldsymbol{\nabla} d|^2$. The data term $\sum_{n=1}^{N} c_n \cdot \Psi\big((d - d_n^{\text{init}})^2\big)$, where N is the number of focus measures, represents the sum of the similarity assumptions of the solution d and the initial maps d_n^{init}. It allows to use one or multiple focus measure combining them in accordance to user preferences. This is particularly beneficial when the selected focus measures supplement each other, i.e. they perform best on different image regions. The summands of the data term are weighted with the corresponding confidence functions c_n. In addition, to reduce the influence of outliers in the initial depth maps, we apply the sub-quadratic penaliser

$$\Psi(s^2) = \sqrt{s^2 + \varepsilon^2}\,, \tag{10}$$

with the small positive constant ε.

2.2.1 Minimisation

Following the calculus of variations [36], the minimiser of the energy functional (9) must necessarily satisfy the Euler-Lagrange equation. It reads

$$\sum_{n=1}^{N} c_n \cdot \Psi'\big((d - d_n^{\text{init}})^2\big) \cdot (d - d_n^{\text{init}}) \;-\; \alpha\, \Delta d \;=\; 0\,, \tag{11}$$

with homogeneous Neumann boundary condition

$$\boldsymbol{n}^\top \boldsymbol{\nabla} d = 0\,, \tag{12}$$

where $\boldsymbol{n}$ is the outer normal vector to the image boundary $\partial\Omega$.

2.2.2 Anisotropic Modification

The smoothness term $|\boldsymbol{\nabla} d|^2$ leads to the linear isotropic diffusion term Δd in (11). It provides a smoothing that is not space-variant and that is equal in all directions. However, it may be beneficial to adapt the diffusion to the image structures. Thus, inspired by Weickert [37] and Zimmer et al. [38], we replace the discussed diffusion term by a nonlinear anisotropic variant, specifically by a *joint image- and depth-driven* diffusion: First, we determine the smoothing *directions* by the structures of the evolving all-in-focus image $f(\boldsymbol{x}, d(\boldsymbol{x}))$. This image is supposed to provide a richer directional information than the individual blurred input images. Second, we adjust the *amount* of smoothing by the gradients of the computed depth map $d(\boldsymbol{x})$.
Let us consider the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ of the structure tensor [34]

$$\boldsymbol{J}_{\rho_s}(\boldsymbol{\nabla} f_{\sigma_s}) = G_{\rho_s} * \big(\boldsymbol{\nabla} f_{\sigma_s}(\boldsymbol{x}, d(\boldsymbol{x}))\; \boldsymbol{\nabla} f_{\sigma_s}^\top(\boldsymbol{x}, d(\boldsymbol{x}))\big)\;, \tag{13}$$

where $*$ denotes the convolution operator and G_{ρ_s} a Gaussian of standard deviation ρ_s. In our experiments we set the parameters to $\sigma_s = 1.0$ and $\rho_s = 1.3$. In the following, we assume $\boldsymbol{v}_1$ to correspond to the largest eigenvalue, and consequently to point across edges of $f(\boldsymbol{x}, d(\boldsymbol{x}))$. Accordingly, $\boldsymbol{v}_2$ points along edges. These eigenvectors build the basis of the diffusion tensor

$$\boldsymbol{D} = (\boldsymbol{v}_1 \;\; \boldsymbol{v}_2) \begin{pmatrix} g\left((\boldsymbol{v}_1^\top \boldsymbol{\nabla} d)^2\right) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^\top \\ \boldsymbol{v}_2^\top \end{pmatrix}\,, \tag{14}$$

where g is the Charbonnier diffusivity [39]

$$g(s^2) = \frac{1}{\sqrt{1 + s^2/\lambda^2}} \tag{15}$$

with some contrast parameter $\lambda > 0$. The diffusion tensor (14) creates a reduced smoothing across image structures (along $\boldsymbol{v}_1$) and a strong homogeneous smoothing along them (along $\boldsymbol{v}_2$). By replacing Δd in (11) by its anisotropic counterpart $\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d)$ we obtain our final model

$$\sum_{n=1}^{N} c_n \cdot \Psi'\big((d - d_n^{\text{init}})^2\big) \cdot (d - d_n^{\text{init}}) \;-\; \alpha \operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d) \;=\; 0 \tag{16}$$

with the boundary condition

$$\boldsymbol{n}^\top \boldsymbol{D}\,\boldsymbol{\nabla} d = 0\,. \tag{17}$$

2.2.3 Implementation

We discretise Equation 16 on a rectangular pixel grid with uniform grid sizes. Accordingly, we apply a finite difference scheme for the space discretisation of the divergence expression. In particular, we compute for a pixel $(i,j)^\top$ eight neighbouring weights $w_{p,q}$ with the anisotropic stencil proposed by Weickert et al. [40] (with $\alpha_{\text{stencil}} = 0.4$ and $\gamma_{\text{stencil}} = 1.0$). This results in

$$\sum_{n=1}^{N} c_{n,i,j} \cdot \Psi'_{n,i,j} \cdot (d_{i,j} - d_{n,i,j}^{\text{init}}) \;-\; \alpha \sum_{(p,q)\in\mathcal{N}(i,j)} w_{p,q}\,(d_{p,q} - d_{i,j}) \;=\; 0\,, \tag{18}$$

where $\mathcal{N}(i,j)$ is the set of direct neighbours, and $\Psi'_{n,i,j}$ denotes the discrete version of the expression $\Psi'\big((d - d_n^{\text{init}})^2\big)$.
The application of Equation 18 to each pixel constitutes a nonlinear system of equations. We solve it for the unknowns $d_{i,j}$ with a so-called *lagged nonlinearity* method, which basically consists of two nested loops: In an outer loop, we update the nonlinearity terms $\Psi'_{n,i,j}$ and $w_{p,q}$. In the inner loop, we keep these expressions fixed and thus, only have to solve a sparse linear system of equations. For this purpose, we apply the *Fast Jacobi* algorithm of Grewenig et al. [41]. Basically, it resembles a standard Jacobi solver. However, varying over-relaxations where half of them may exceed the stability limit allow an enormous speed-up. This method is easily parallelisable and hence well-suited for a fast GPU implementation. More precisely, one step of this Fast Jacobi solver with iteration index k and varying over-relaxation parameter ω^k reads

$$d_{i,j}^{k+1} = (1-\omega^k)\cdot d_{i,j}^k + \omega^k \cdot \frac{\sum_{n=1}^{N} c_{n,i,j} \cdot \Psi'_{n,i,j} \cdot d_{n,i,j}^{\text{init}} + \alpha \sum_{\mathcal{N}(i,j)} w_{p,q} d_{p,q}^k}{\sum_{n=1}^{N} c_{n,i,j} \cdot \Psi'_{n,i,j} + \alpha \sum_{\mathcal{N}(i,j)} w_{p,q}}\,. \tag{19}$$

This elliptic approach with the Fast Jacobi solver allows a significant speed-up. Furthermore, the runtime of the core part of our algorithm, the anisotropic depth map smoothing, is independent of the number of images in the focus stack.

2.2.4 Colour Images

For didactic reasons, we have restricted ourselves to grey-valued images so far. However, the extension of our model to colour images is straightforward. Instead of using the structure tensor (13) for a single colour channel, we apply the combined structure tensor [42, 43]

$$G_{\rho_s} * \sum_{\ell=1}^{3} \boldsymbol{\nabla} f_{\sigma_s}^{\ell}(\boldsymbol{x}, d)\ \boldsymbol{\nabla} f_{\sigma_s}^{\ell\,\top}(\boldsymbol{x}, d) \tag{20}$$

to determine the eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Here, $f_{\sigma_s}^{\ell}(\boldsymbol{x}, d)$ represents the colour channel ℓ of the RGB image $\boldsymbol{f}_{\sigma_s}$. Overall, we still solely have to determine *one* joint depth map for all channels. Hence, the increase in runtime is negligible.

2.3 Fusion

With the computed accurate depth map we can directly fuse the colour values from the source images to the final composite. We apply linear interpolation to determine the colour values at non-integer depth values:

$$f_{i,j,d_{i,j}} = (1 - \delta) \cdot f_{i,j,\lfloor d \rfloor} + \delta \cdot f_{i,j,\lfloor d \rfloor + 1} \tag{21}$$

where $\lfloor d \rfloor$ denotes the largest integer smaller or equal to $d_{i,j}$, and $\delta := d_{i,j} - \lfloor d \rfloor$. This procedure allows to mimic colour values of pixels that are in focus *between* two frames. Obviously, this cannot be achieved with segmentation-based methods.

3 Experiments

The evaluation of our method consists of the following five parts: First, we test the discussed in-focus measures and demonstrate the improvements provided by the proposed depth map smoothing technique. Second, we show that an anisotropic modification further enhances the quality of the fused image. Next we demonstrate that our method can combine the information from multiple in-focus measures, and thus, yields even more accurate results. In our fourth experiment, we compare the performance of our approach to three focus fusion methods from the literature. Finally, we evaluate the runtime of our algorithm.

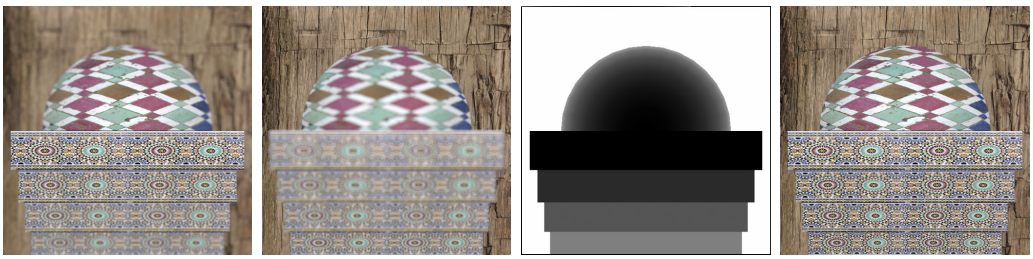


Figure 1: Synthetic data set. **From left to right:** (a) Frame 1 with the shortest focal distance. (b) Frame 13 with the largest focal distance. (c) Ground truth depth map: Brighter grey tones describe larger depth values. (d) Ground truth image (all-in-focus).

3.1 Comparison of In-Focus Measures

To evaluate the different sharpness criteria in terms of an error measure, we generated a synthetic data set with *Blender*[1] (cf. Figure 1). It consists of 13 images of size 512×512: The images are created with the increasing focal plane. In addition, we generated the ground truth depth map (c) and the ground truth all-in-focus image (d). Thus, it is possible to measure the difference between the resulting fused image and its ground truth.

Results with Initial Depth Maps First, we compare the quality of the initial depth maps resulting from the discussed in-focus measures (Table 1). These depth maps and the fused images are shown in Figure 2. In addition, Table 2 lists the quantitative results corresponding to the in-focus measures: The mean squared errors (MSEs) between the fused image and its ground truth, as well as the parameters optimised w.r.t. these MSEs. We observe that the gradient performs worst, while the trace and the determinant of the structure tensor produce the most accurate depth maps.

Results with Final Depth Maps Figure 3 depicts the final depth maps and the fused images after smoothing the depth maps with homogeneous diffusion (Equation 11). It is obvious that all depth maps improved substantially: The final depth maps are much more reliable. This observation is underlined by the error measures in Table 3 (MSE Isotropic): The MSEs of the fused images are for all depth maps significantly reduced. This clearly illustrates the usefulness of depth map regularisation in focus fusion applications.

[1] http://www.blender.org/

Table 2: Results with initial depth maps d^{init}: The pre-smoothing parameter σ is optimised w.r.t. the mean square error (MSE) between the fused image and its ground truth.

measure	parameters	MSE
gradient magnitude	$\sigma = 0.7$	95.35
Laplacian	$\sigma = 0.5$	23.08
Frobenius norm of the Hessian	$\sigma = 0.5$	10.55
trace of the structure tensor	$\sigma = 0.2$, $\rho = 1.7$	3.59
determinant of the structure tensor	$\sigma = 0.6$, $\rho = 1.8$	4.74
variance	9×9 stencil	4.74

3.2 Isotropic vs. Anisotropic Smoothing

In this experiment we compare the linear isotropic and the nonlinear anisotropic diffusion process. As before we test the performance on the synthetic image set. Figure 4 shows the results with the different in-focus measures using an anisotropic diffusion from Equation 16. The quantitative results with their MSEs are provided in Table 3 (MSE Anisotropic). We see that the anisotropic diffusion performs always better than the isotropic one, yielding an improvement up to 24%. The anisotropic depth map smoothing is a more complex method, however, it allows a very accurate depth map estimation.

3.3 Multiple In-Focus Measures

As described in Section 2, our general approach is not restricted to a single in-focus measure. Here, we provide an example that shows an improvement by means of multiple focus measures. We apply a combination of three measures combined with an anisotropic smoothness term on the synthetic data. For the in-focus measures we used the Laplacian, the Frobenius norm of the Hessian and the trace of the structure tensor. The result is shown in Figure 5 (*last column*), and the corresponding error measure is given in Table 4. As we can see the error measure is further decreased resulting in a smaller value than any of the single sharpness criteria. Moreover, this result may be further improved by introducing the weights and optimising them.

Table 3: Results with final depth maps d (isotropic and anisotropic): The threshold parameter θ and the smoothing parameter α are optimised w.r.t. the mean square error (MSE) between the fused image and its ground truth.

measure	MSE Isotropic	MSE Anisotropic
gradient magnitude	8.66 ($\theta = 30$, $\alpha = 2$)	8.39 ($\theta = 37$, $\alpha = 6$)
Laplacian	4.61 ($\theta = 48$, $\alpha = 1$)	3.53 ($\theta = 60$, $\alpha = 2$)
Frobenius norm of the Hessian	4.19 ($\theta = 37$, $\alpha = 1$)	3.47 ($\theta = 41$, $\alpha = 3$)
trace of the structure tensor	3.57 ($\theta = 0$, $\alpha = 1$)	3.48 ($\theta = 0$, $\alpha = 4$)
determinant of the structure tensor	4.45 ($\theta = 0$, $\alpha = 0.5$)	4.17 ($\theta = 0$, $\alpha = 8$)
variance	4.00 ($\theta = 0$, $\alpha = 1$)	3.98 ($\theta = 0$, $\alpha = 9$)

Table 4: Results with multiple in-focus measures (Laplacian, Hessian, trace of the structure tensor). The threshold parameters θ can be found in Table 3. The smoothing parameter α is optimised w.r.t. the mean square error (MSE) between the fused image and its ground truth.

measure	MSE
multiple in-focus measure	3.25 ($\alpha = 2$)

3.4 Comparison to Other Methods

Let us now compare our method with three publicly available approaches that represent different groups of techniques: As multiscale transformation-based method, we select the approach of Forster et al. [5] that performs a complex wavelet decomposition. A representative of segmentation-based techniques is the method of Agarwala et al. [13]. In particular to evaluate our depth map estimation, we additionaly compare our results to the *depth-from-defocus* method of Aguet et al. [44]. Based on deconvolution, they explicitly model the physical image acquisition process and jointly estimate the sharp image and the depth map.
We obtained the results of the methods of Forster et al. and Aguet et al. using

Table 5: Comparison to the other methods in terms of the mean square error (MSE), the correlation and the structural similarity index (SSIM) between the fused image and its ground truth.

method	MSE	Correlation	SSIM
Forster et al. [5]	152.12	0.95	0.98
Agarwala et al. [13]	135.97	0.98	0.98
Aguet et al. [44]	113.73	0.98	0.98
proposed	**3.25**	**1.00**	**1.00**

plug-ins for the ImageJ public domain program provided by the authors[2] with optimised parameters. For the method of Agarwala et al. we used the provided executable[3]. Here we obtained all results with "maximum contrast image objective" and gradient-domain fusion.

We start with our synthetic data set for which the results are shown in Figure 5. We observe that the most accurate depth map is achieved with our approach. To objectively evaluate the methods, we use three measures: the MSE, the correlation, the structural similarity (SSIM) index proposed by Wang et al. [45]. The SSIM index is computed with the MATLAB code provided by the authors[4] using the default parameters. While for the MSE smaller values describe a better performance, larger values of the correlation and the SSIM show that the images are more similar. Table 5 contains the computed quantitative measures between the fused images and the ground truth image. As we see the best quantitative results for all three measures are achieved with our method.

Besides the synthetic data set, we used two commonly available real–world data sets: *Insect*[5] and *Clock*[6]. The *Insect* data set consists of 13 frames of size 1344×1021 with increasing focal plane distance. The corresponding results are depicted in Figure 6. Our computed depth map is less noisy and more reliable than the other depth maps (*second row*). Zooms into the fused images (*fourth row*) help to provide more details. The quality of our approach evidently outperforms the methods of Forster et al. and Aguet et al. Solely the method of Agarwala et al. is comparable to our method: Both methods

[2] `http://bigwww.epfl.ch/demo/edf/`

[3] `http://grail.cs.washington.edu/projects/photomontage/`

[4] `http://www.ece.uwaterloo.ca/z70wang/research/ssim/`

[5] available from `http://grail.cs.washington.edu/projects/photomontage/`

[6] available from [35]

Table 6: Comparison to the method of Mahmood and Choi [18] in terms of the root mean square error.

data	Mahmood and Choi [18]	proposed
slope	0.67	0.12
sinusoidal	3.61	0.46
wave	5.32	0.58

preserve fine details well.
The *Clock* data set consists of 2 frames of size 480×480. Figure 7 demonstrates the results for this data set. Again we see that the proposed method creates a better depth map than the other techniques (*second row*). From a zoom into the fused images (*fourth row*) we can observe the following: In the result with the method of Forster et al. there are blurring artefacts between the two clocks. The result with the method of Agarwala et al. contains an artefact near to the number "8", and the result with the method of Aguet et al. is less sharp than our image. Thus, relying on the visual quality of the fused image the proposed method is ahead of the other methods.
In addition, we evaluate the performance of our method for a number of other data sets from the *Focal Stack Photography* project [35], which are illustrated in Figure 8. One observes that thanks to the edge-preserving smoothing the depth maps are accurate. Accordingly, the fusion results are of high quality as well. Furthermore, we demonstrate the results of our method for data sets from Mahmood and Choi [18] shown in Figures 9 and 10. For the synthetic data we compare the quantitative results in terms of the root mean square error to the method [18] (Table 6). As their result we show their best performance with the modified Laplacian as a focus measure. We observe that our method performs favourably for all data sets. Once again, our proposed model is able to precisely estimate the depth maps as well as the fused images: The depth maps are piecewise smooth while preserving important image structures, and the fused images are sharp everywhere.

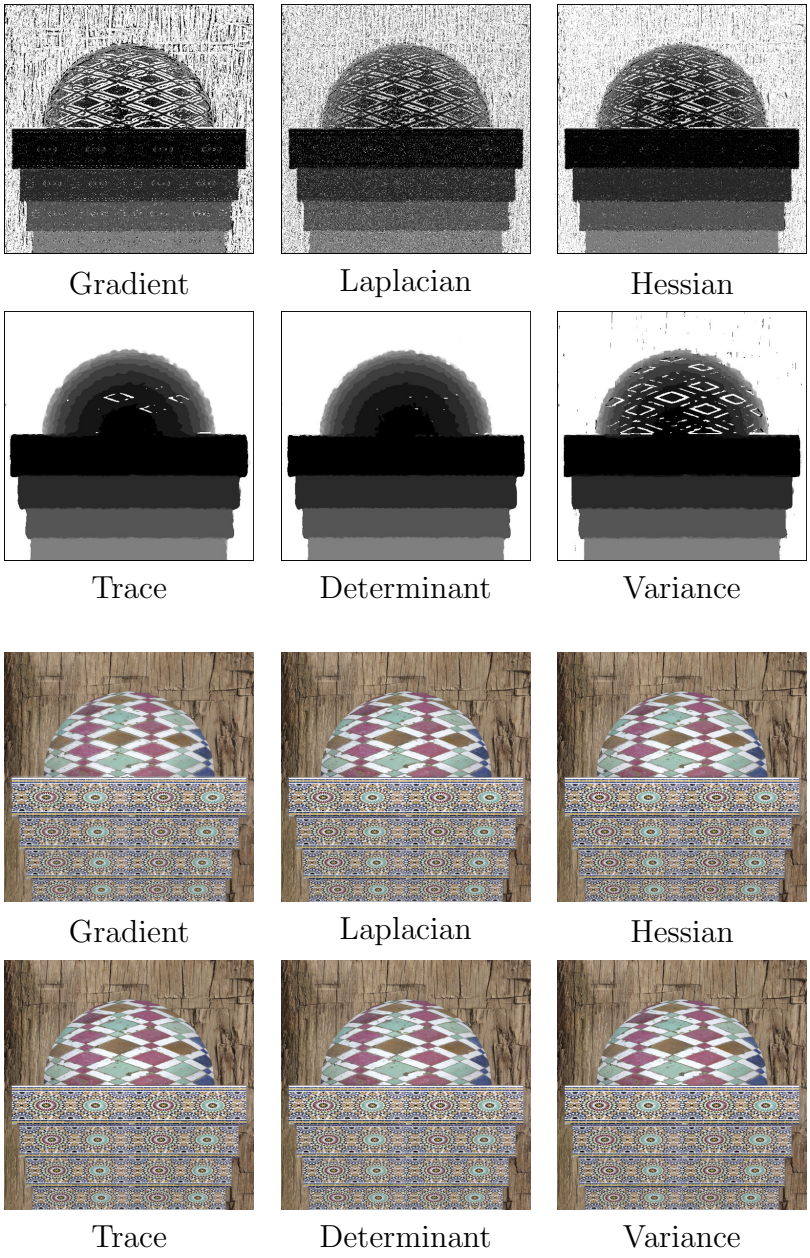


Figure 2: Comparison of different in-focus measures by means of the initial depth maps d^{init} and the fused results. Parameters and error measures can be found in Table 2.

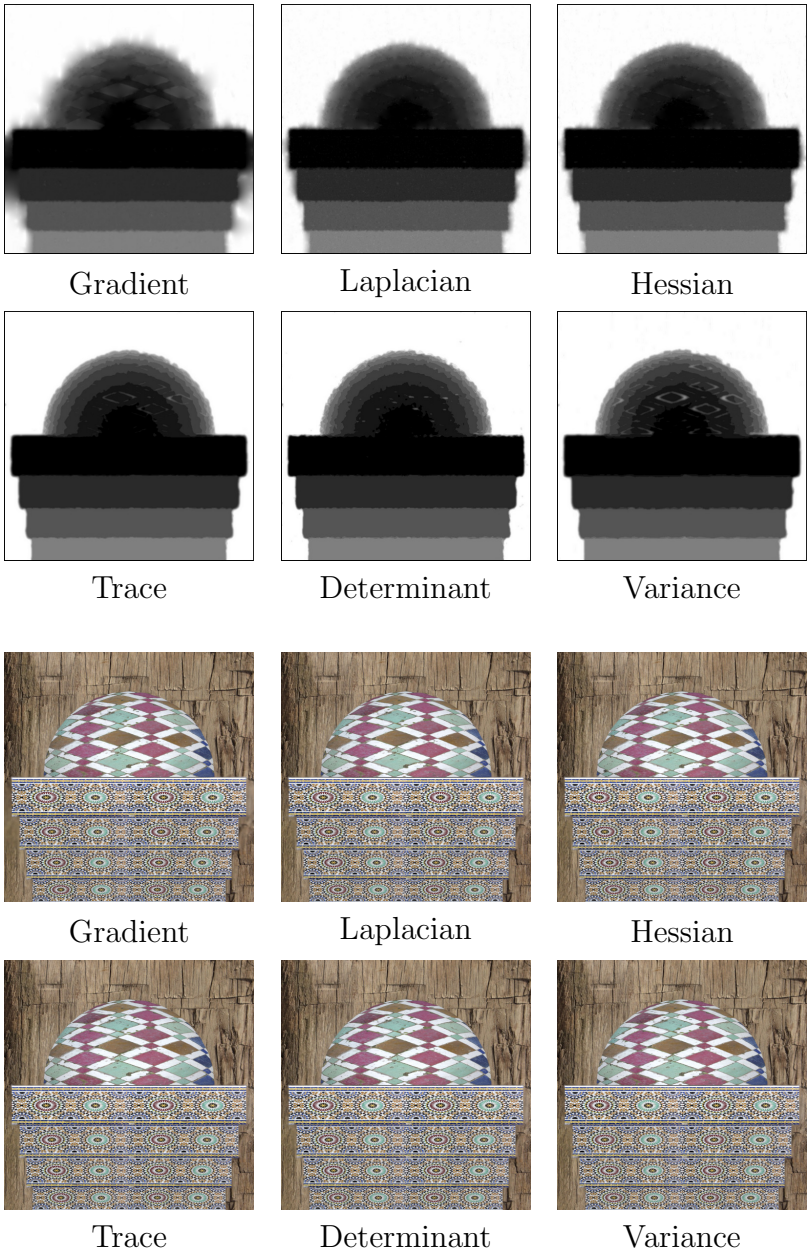


Figure 3: Comparison of different in-focus measures by means of the final depth maps d and the fused results with *linear isotropic diffusion*. Parameters and error measures can be found in Table 3.

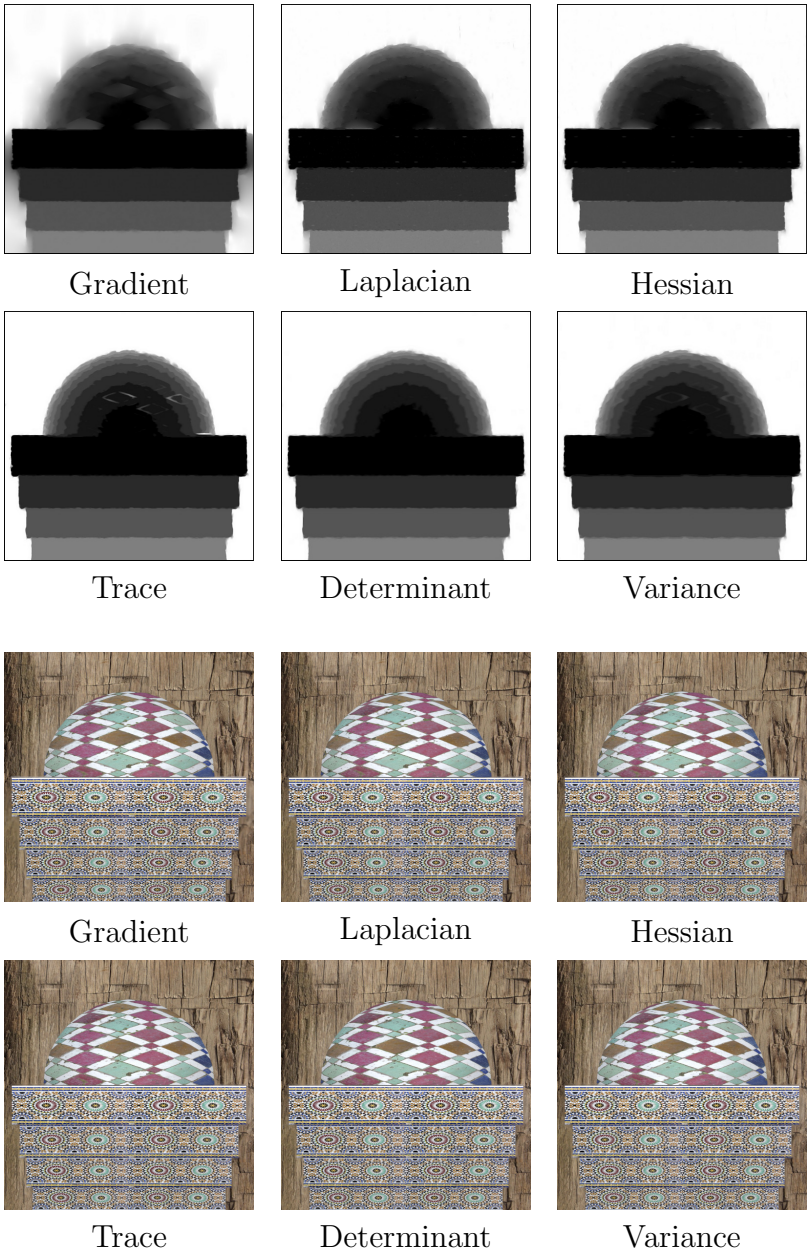
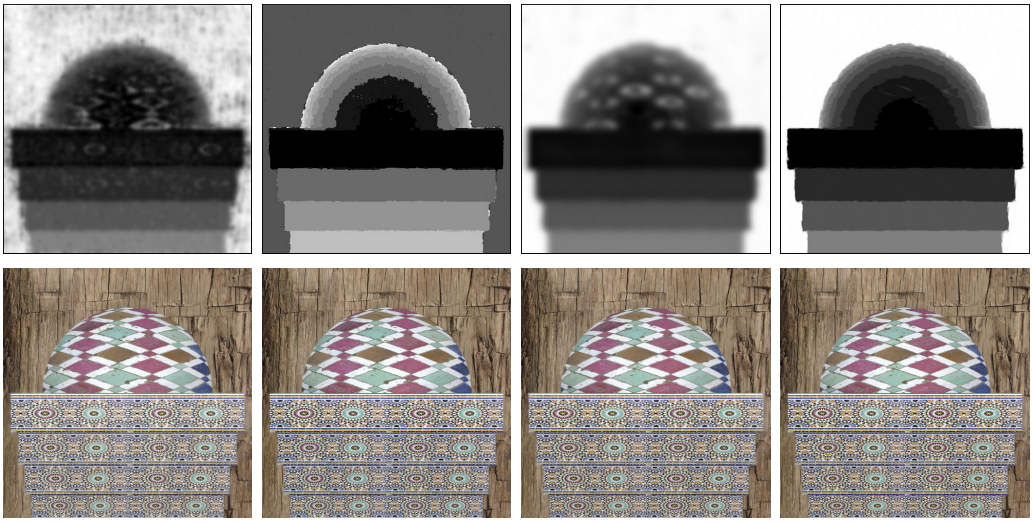


Figure 4: Comparison of different in-focus measures by means of the final depth maps d and the fused results with the *nonlinear anisotropic diffusion*. Parameters and error measures can be found in Table 3.



Forster et al. Agarwala et al. Aguet et al. proposed

Figure 5: Visual comparison for synthetic data set. **Top:** Depth maps. **Bottom:** Fused images. Here we applied a multiple in-focus measure: the Laplacian, the Frobenius norm of the Hessian, the trace of the structure tensor. The corresponding parameters are listed in Table 4.

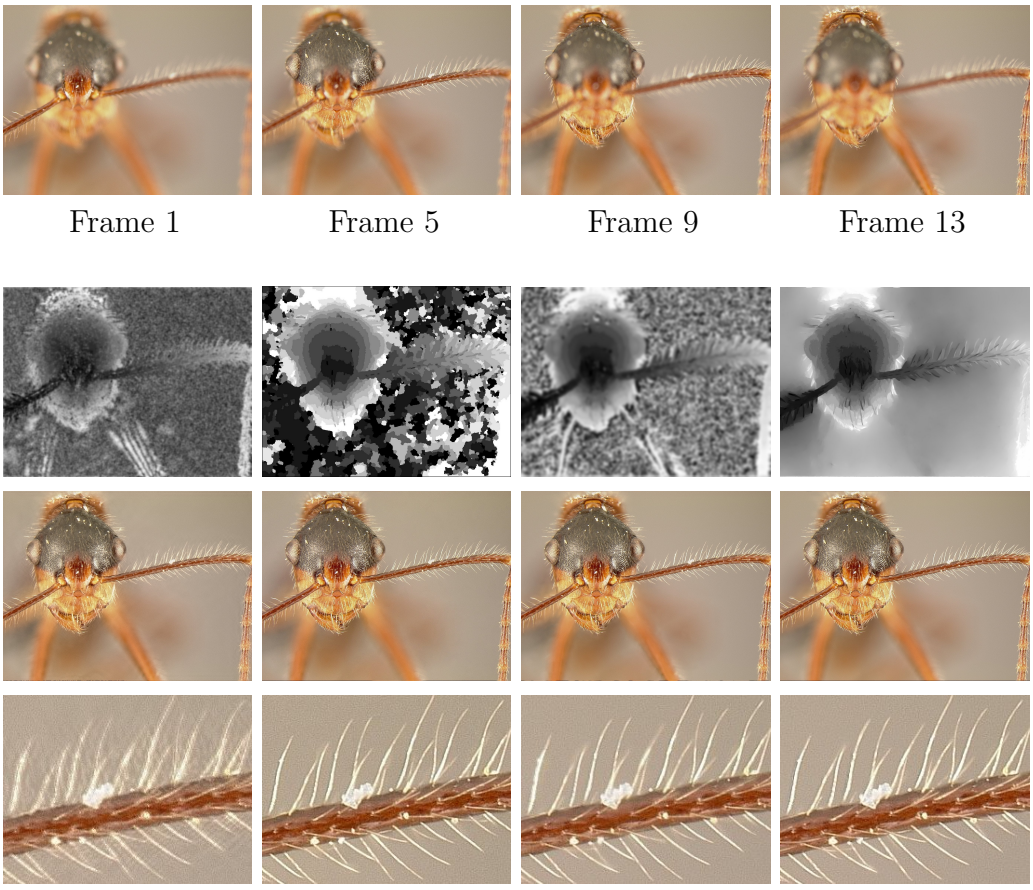
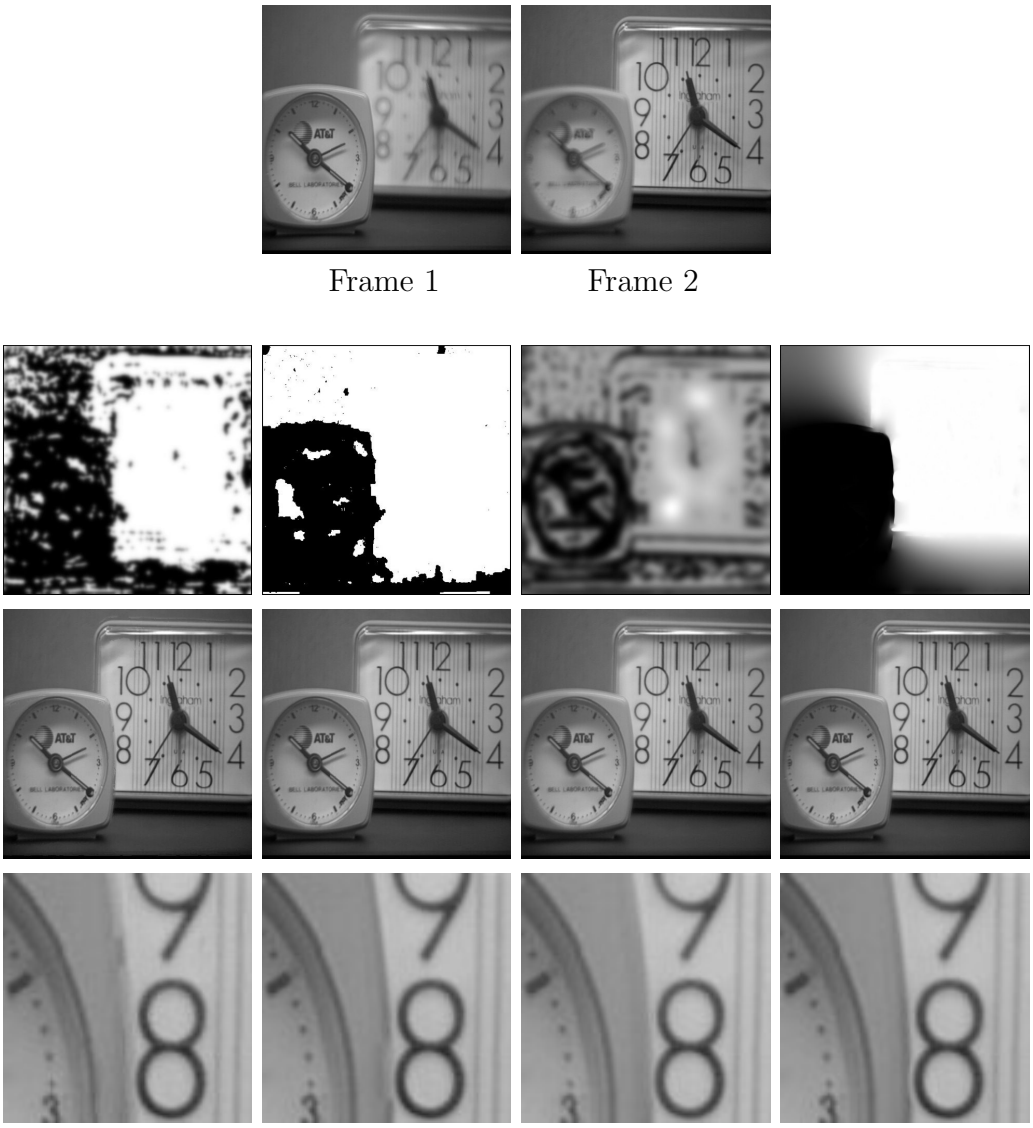


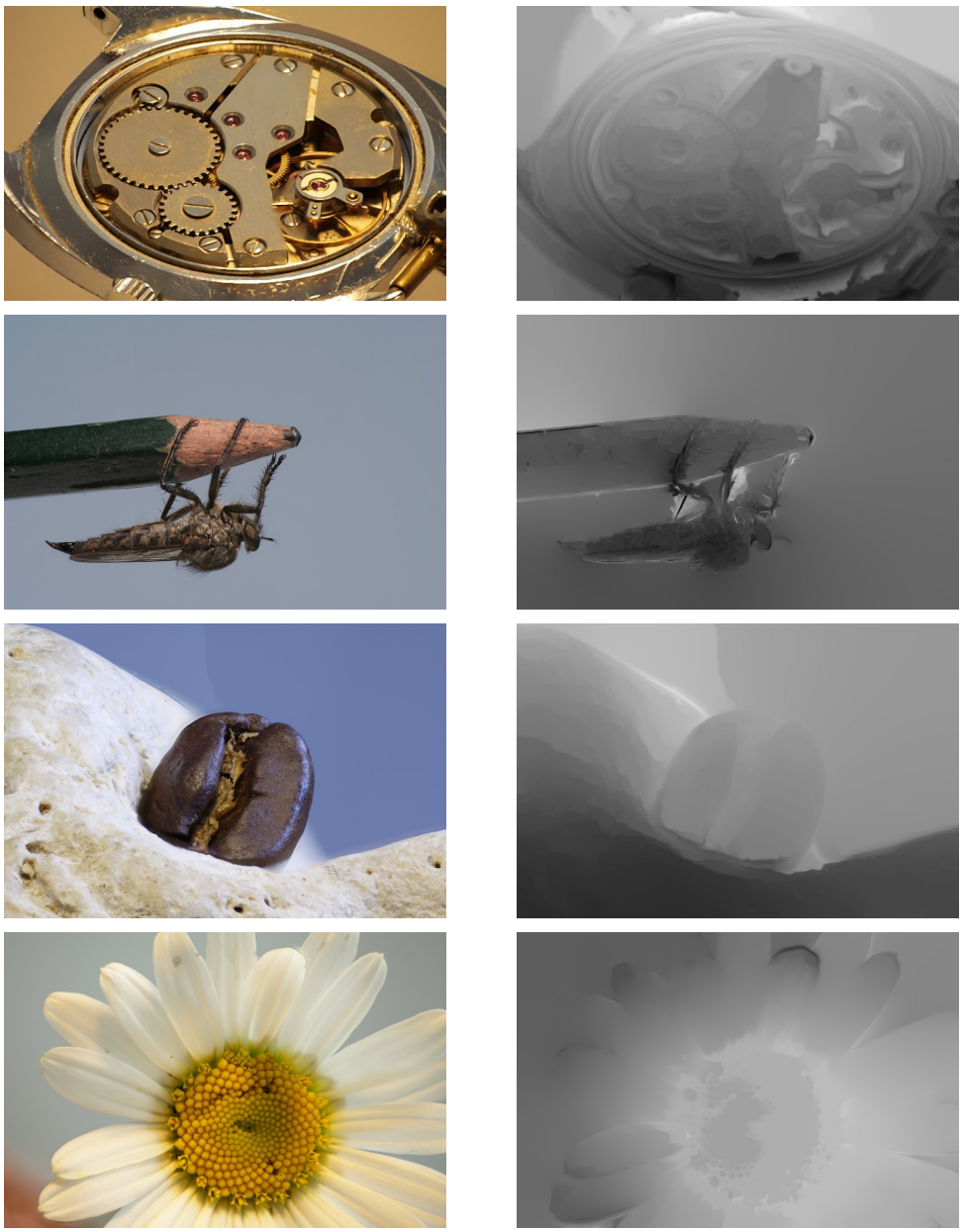
Figure 6: Results for the *Insect* data set. **First row:** Input frames with different focal planes. **Second row:** Depth maps. **Third row:** Fused images. **Fourth row:** Zoom-ins. Here we use the Frobenius norm of the Hessian as in-focus measure and the parameters $\sigma = 0.5$, $\theta = 60$, and $\alpha = 5$.



Frame 1 Frame 2

Forster et al. Agarwala et al. Aguet et al. Proposed

Figure 7: Results for the *Clock* data set. **First row:** Input frames with different focal planes. **Second row:** Depth maps. **Third row:** Fused images. **Fourth row:** Zooms. Here we use the Frobenius norm of the Hessian as in-focus measure and the parameters $\sigma = 1.0$, $\theta = 6$, and $\alpha = 100$.



Fused image | Final depth map

Figure 8: Results for data sets from *Focal Stack Photography* project [35]. **From top to bottom:** *Watch* data set: 395 frames of size 640 × 427 ($\theta = 10$, $\alpha = 5$). *Fly* data set: 20 frames of size 1024 × 682 ($\theta = 10$, $\alpha = 5$). *Coffee* data set: 91 frames of size 1024 × 683 ($\theta = 10$, $\alpha = 25$). *Flower* data set: 20 frames of size 1024 × 683 ($\theta = 5$, $\alpha = 30$). Here we use the Frobenius norm of the Hessian as in-focus measure. The pre-smoothing parameter σ is always fixed to 1.0.

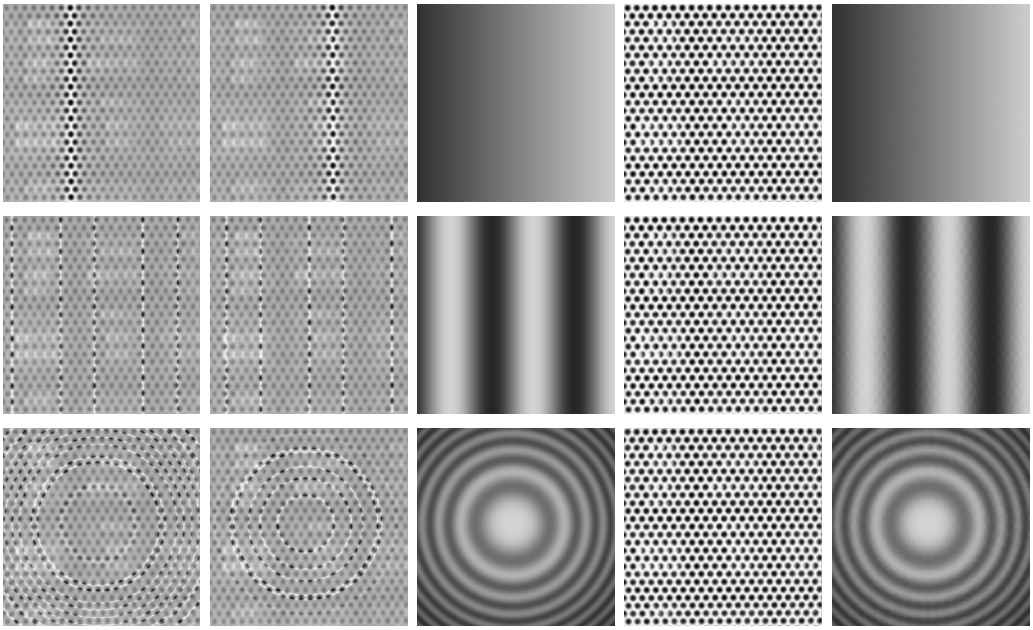
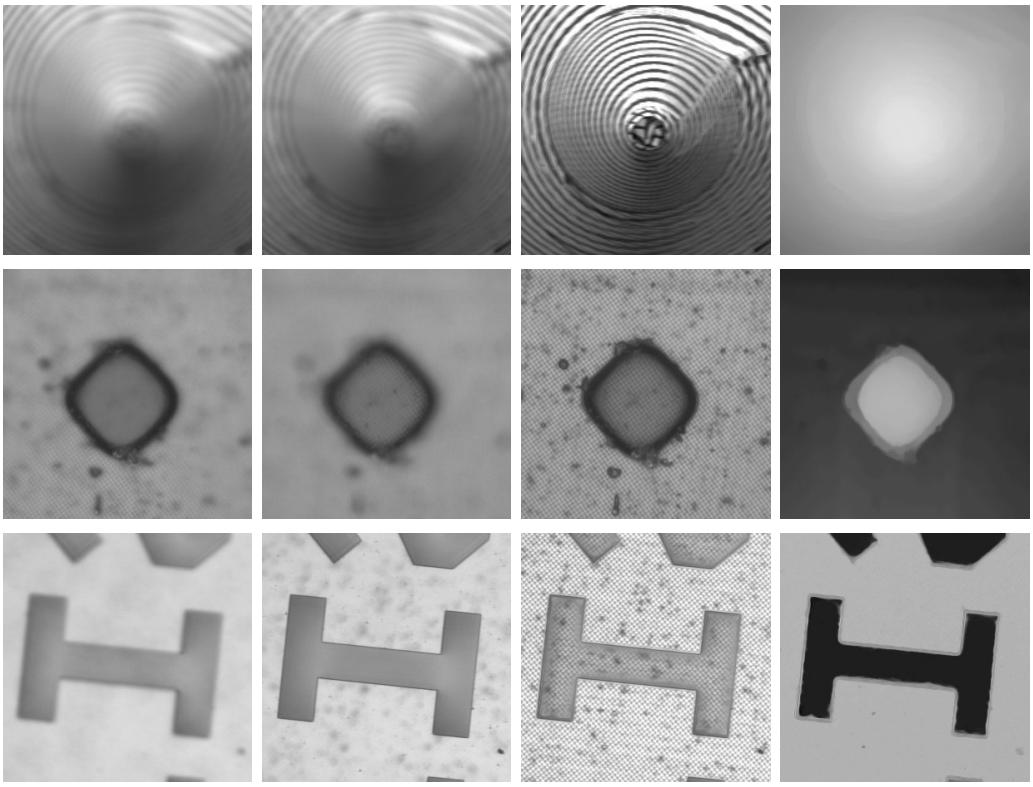


Figure 9: Results for synthetic data from Mahmood and Choi [18]. **First row:** *Slope* data set: 60 frames of size 300 × 300 ($\sigma = 2.5$, $\theta = 0$, $\alpha = 42$). **Second row:** *Sinusoidal* data set: 60 frames of size 300 × 300 ($\sigma = 0.5$, $\theta = 46$, $\alpha = 2$). **Third row:** *Wave* data set: 60 frames of size 300 × 300 ($\sigma = 0.4$, $\theta = 31$, $\alpha = 1$). Here we use the Frobenius norm of the Hessian as in-focus measure.



Frame 25 Frame 35 Fused image Final depth map

Figure 10: Results for real-world data from Mahmood and Choi [18]. **First row:** *Cone* data set: 97 frames of size 200 × 200 ($\sigma = 1.0$, $\theta = 0$, $\alpha = 10$). **Second row:** *TFT-LCD color filter* data set: 60 frames of size 300 × 300 ($\sigma = 0.5$, $\theta = 0$, $\alpha = 7$). **Third row:** *Letter I* data set: 60 frames of size 300 × 300 ($\sigma = 0.2$, $\theta = 15$, $\alpha = 1$). Here we use the Frobenius norm of the Hessian as in-focus measure.

Table 7: Comparison of the runtime (in seconds) of the gradient descent method from [22] and our proposed Fast Jacobi algorithm.

number of images	resolution	gradient descent	Fast Jacobi
2	256×256	32.9	0.5
2	512×512	139.6	1.2
2	1024×1024	644.2	4.1
13	256×256	33.7	0.5
13	512×512	145.5	1.3
13	1024×1024	660.6	4.2

3.5 Runtime

Finally, we evaluate the efficiency of our method. We compare the minimisation via the gradient descent used in our conference paper to the Fast Jacobi algorithm described in Section 2. We tested the gradient descent on the Intel(R) Xeon(R) W3565 @ 3.20GHz processor. Our implementation of Fast Jacobi is written in CUDA and runs on an NVidia Geforce GTX 460 graphics card. Table 7 lists the runtime for different data sets, i.e. with varying number of images and resolution. The advantage is that our algorithm operates on the depth map. Thus, changing the number of images of the focal stack does not significantly increase the runtime. Furthermore, we obviously achieve a considerable speed-up using the parallel GPU implementation.

4 Conclusions and Outlook

We have shown that a regularisation of the depth map is generally a powerful tool for enhancing the quality of focus fusion. To this end, our framework implements modern techniques such as robustified data fidelity terms and an anisotropic smoothness term. Moreover, our flexibile approach is able to combine the information from multiple in-focus measures. In this way, we are able to improve the focus fusion substantially. We demonstrate this with several in-focus measures and evaluate our final results w.r.t. other approaches from the literature. Not only the all-in-focus composites benefit from the proposed approach, but also the computed depth maps provide valuable information that may serve as input for further computer vision tasks. Furthermore, the proposed algorithmic realisation leads to a very

substantial speed-up that allows an interactive usage and tuning of the fusion results.
In future work, we intend to test and modify our focus fusion framework for other image fusion tasks such as exposure fusion or multimodal image fusion.

Acknowledgements.

Our research has been partly funded by the Deutsche Forschungsgemeinschaft through a Gottfried Wilhelm Leibniz prize for Joachim Weickert. This is gratefully acknowledged. We further would like to thank Dr. Muhammad Tariq Mohammed from Korea University of Technology and Education (Korea) and Prof. Tae-Sun Choi from Gwangju Institute of Science and Technolgy (Korea) for providing the data used in Figures 9 and 10.

References

[1] J. Ogden, E. Adelson, J. Bergen, P. Burt, Pyramid-based computer graphics, RCA Engineer 30 (5) (1985) 4–15.

[2] P. Burt, R. Kolczynski, Enhanced image capture through fusion, in: Proc. Fourth IEEE International Conference on Computer Vision, 1993, pp. 173–182.

[3] V. Petrovic, C. Xydeas, Gradient-based multiresolution image fusion, IEEE Transactions on Image Processing 13 (2) (2004) 228–237.

[4] H. Li, B. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing 57 (3) (1995) 235–245.

[5] B. Forster, D. Van De Ville, J. Berent, D. Sage, M. Unser, Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images, Microscopy Research and Technique 65 (1-2) (2004) 33–42.

[6] J. Lewis, R. O'Callaghan, S. Nikolov, D. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, Information Fusion 8 (2) (2007) 119–130.

[7] Z. Zhang, R. Blum, A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application, Proceedings of the IEEE 87 (8) (1999) 1315–1326.

[8] W. Wu, X. Yang, Y. Pang, J. Peng, G. Jeon, A multifocus image fusion method by using Hidden Markov Model, Optics Communications 287 (2013) 63–72.

[9] T. Wan, C. Zhu, Z. Qin, Multifocus image fusion based on robust principal component analysis, Pattern Recognition Letters 34 (9) (2013) 1001–1008.

[10] Z. Wang, Y. Ma, J. Gu, Multi-focus image fusion using PCNN, Pattern Recognition 43 (6) (2010) 2003–2016.

[11] S. Pop, O. Lavialle, R. Terebes, M. Borda, A PDE-based approach for image fusion, in: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems, Vol. 4678 of Lecture Notes in Computer Science, Springer, Heidelberg, 2007, pp. 121–131.

[12] W. Wang, P. Shui, X. Feng, Variational models for fusion and denoising of multifocus images, IEEE Signal Processing Letters 15 (2008) 65–68.

[13] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, M. Cohen, Interactive digital photomontage, ACM Transactions on Graphics 23 (3) (2004) 294–302.

[14] F. Šroubek, G. Cristóbal, J. Flusser, Image fusion based on level set segmentation, in: Proc. 14th European Signal Processing Conference (EUSIPCO), Florence, Italy, 2006, pp. 1–5.

[15] S. Li, B. Yang, Multifocus image fusion using region segmentation and spatial frequency, Image and Vision Computing 26 (7) (2008) 971–979.

[16] S. Shim, T. Choi, A novel iterative shape from focus algorithm based on combinatorial optimization, Pattern Recognition 43 (10) (2010) 3338–3347.

[17] M. Mahmood, A. Majid, T. Choi, Optimal depth estimation by combining focus measures using genetic programming, Information Sciences 181 (7) (2011) 1249–1263.

[18] M. Mahmood, T. Choi, Nonlinear approach for enhancement of image focus volume in shape from focus, IEEE Transactions on Image Processing 21 (5) (2012) 2866–2873.

[19] Y. Liu, J. Jin, Q. Wang, Y. Shen, X. Dong, Region level based multi-focus image fusion using quaternion wavelet and normalized cut, Signal Processing 97 (2014) 9–30.

[20] M. Muhammad, T. Choi, Sampling for shape from focus in optical microscopy, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (3) (2012) 564–573.

[21] S. Bae, F. Durand, Defocus magnification, Computer Graphics Forum 26 (3) (2007) 571–579.

[22] M. Boshtayeva, D. Hafner, J. Weickert, Focus fusion with anisotropic depth map smoothing, in: A. Bors, E. Hancock, W. Smith, R. Wilson (Eds.), Computer Analysis of Images and Patterns, Vol. 8048 of Lecture Notes in Computer Science, Springer, Berlin, 2013, pp. 67–74.

[23] J. M. Tenenbaum, Accommodation in computer vision, Ph.D. thesis, Department of Electrical Engineering, Stanford University, California, U.S.A. (Nov 1970).

[24] J. F. Schlag, A. C. Sanderson, C. P. Neuman, F. C. Wimberly, Implementation of automatic focusing algorithms for a computer vision system with camera control, Tech. Rep. CMU-RI-TR-83-14, Robotics Institute, Carnegic-Mellon University, Pittsburgh, PA, U.S.A. (Aug. 1983).

[25] E. Krotkov, Focusing, International Journal of Computer Vision 1 (3) (1987) 223–237.

[26] S. Nayar, Y. Nakagawa, Shape from focus, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (8) (1994) 824–831.

[27] N. Xu, K. Tan, H. Arora, N. Ahuja, Generating omnifocus images using graph cuts and a new focus measure, in: Proc. 17th International Conference on Pattern Recognition, Vol. 4, IEEE Computer Society Press, Washington, DC, 2004, pp. 697–700.

[28] M. Subbarao, T. Choi, A. Nikzad, Focusing techniques, Journal of Optical Engineering 32 (1993) 2824–2836.

[29] J. Tian, L. Chen, L. Ma, W. Yu, Multi-focus image fusion using a bilateral gradient-based sharpness criterion, Optics Communications 284 (1) (2011) 80–87.

[30] W. Huang, Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognition Letters 28 (4) (2007) 493–500.

[31] V. Aslantas, R. Kurban, A comparison of criterion functions for fusion of multi-focus noisy images, Optics Communications 282 (16) (2009) 3231–3242.

[32] B. Lu, C. Miao, Structure tensor based image fusion, in: Proc. Third International Symposium on Electronic Commerce and Security Workshops, Academy Publisher, Guangzhou, P.R. China, 2010, pp. 343–346.

[33] F. Luo, B. Lu, C. Miao, Multifocus image fusion with trace-based structure tensor, in: Proc. SPIE, Vol. 8200, 2011, p. 82001G.

[34] W. Förstner, E. Gülch, A fast operator for detection and precise location of distinct points, corners and centres of circular features, in: Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, 1987, pp. 281–305.

[35] O. Barth, M. Osthof, Focal stack photography project, Lecture on Computational Photography, Saarland University, Germany (2010). URL `http://www.mpi-inf.mpg.de/departments/d4/teaching/ss2010/ComputationalPhotography/projects/focalstack/`

[36] I. M. Gelfand, S. V. Fomin, Calculus of Variations, Dover, New York, 2000.

[37] J. Weickert, Theoretical foundations of anisotropic diffusion in image processing, Computing Supplement 11 (1996) 221–236.

[38] H. Zimmer, A. Bruhn, J. Weickert, Optic flow in harmony, International Journal of Computer Vision 93 (3) (2011) 368–388.

[39] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proc. 1994 IEEE International Conference on Image Processing, Vol. 2, IEEE Computer Society Press, Austin, TX, 1994, pp. 168–172.

[40] J. Weickert, M. Welk, M. Wickert, L^2-stable nonstandard finite differences for anisotropic diffusion, in: A. Kuijper, T. Pock, K. Bredies, H. Bischof (Eds.), Scale-Space and Variational Methods in Computer Vision, Vol. 7893 of Lecture Notes in Computer Science, Springer, Berlin, 2013, pp. 380–391.

[41] S. Grewenig, J. Weickert, C. Schroers, A. Bruhn, Cyclic schemes for PDE-based image analysis, Tech. Rep. 327, Department of Mathematics, Saarland University, Saarbrücken, Germany (2013).

[42] S. Di Zenzo, A note on the gradient of a multi-image, Computer Vision, Graphics and Image Processing 33 (1986) 116–125.

[43] J. Weickert, Coherence-enhancing diffusion of colour images, Image and Vision Computing 17 (1999) 201–212.

[44] F. Aguet, D. Van De Ville, M. Unser, Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy, IEEE Transactions on Image Processing 17 (7) (2008) 1144–1153.

[45] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.