In A. Bors, E. Hancock, W. Smith, R. Wilson (Eds.): Computer Analysis of Images and Patterns. Lecture Notes in Computer Science, Vol. 8048, pp. 67-74, Springer, Berlin, 2013. The final publication is available at link.springer.com.

Focus Fusion with Anisotropic Depth Map Smoothing

Madina Boshtayeva, David Hafner, and Joachim Weickert

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

Abstract. Focus fusion methods combine a set of images focused at different depths into a single image where all parts are in focus. The quality of the fusion result strongly depends on a decision map that determines the in-focus areas. Most approaches in the literature achieve this by local decisions without explicitly enforcing smoothness of the depth map. The goal of our paper is to introduce a modern regularisation strategy where we assume that neighbouring pixels in the resulting image have a similar depth. To this end, we consider a partial differential equation (PDE) for the depth map. It combines a robustified data fidelity term with an anisotropic diffusion strategy that involves a matrix-valued diffusion tensor. Experiments with synthetic and real-world data show that this depth map regularisation can improve existing fusion methods substantially. Our methodology is general and can be applied to improve many existing fusion methods.

Keywords: focus fusion, depth map, anisotropic diffusion.

1 Introduction

Certain applications such as microscopy and macro photography create images with a very limited depth of field. To overcome this problem, a common approach is to acquire a set of images with focal planes at different depths, and to fuse these data into a single image that is in focus everywhere. This is called *focus fusion*.

We categorise the focus fusion techniques into two main groups: The first group of methods performs a multiresolution decomposition of the input images. This is done by transforming the image set into a particular domain, e.g. a pyramid domain [1, 2] or a wavelet domain [3, 4]. Then they identify the in-focus areas and combine them into a single composite image. In the last step, this composite is transformed back into the original image domain. These methods suffer from the drawback that fusion in the transformed domain may result in undesirable artefacts in the original domain.

The second group of algorithms operates in the image domain directly. In contrast to the decomposition approaches, they rely on the original intensities and, hence, reproduce the image structures without any modification. An intuitive idea is to select for each pixel the input frame that is in focus according to some sharpness criterion [5]. Then these pixels are directly combined within a single composite image. However, this direct pixel fusion is prone to unpleasant seams in the final result.

The following approaches have been proposed to deal with this drawback. Wang *et al.* [6] and Pop *et al.* [7] explicitly model a smoothness assumption of the fused image by formulating an appropriate energy or partial differential equation (PDE), respectively. However, in this case the resulting image may loose important structures due to the inherent smoothing. Another way is to apply the smoothness constraint not on the final image itself, but on the per-pixel decision. In this regard, a common approach is to first construct an initial noisy map using a simple criterion, and segment it afterwards with a segmentation-based algorithm [8–10].

Our Contributions. The goal of our paper is to address these problems of direct pixel fusion methods. We avoid unpleasant seams by introducing a modern regularisation strategy of the depth map such that neighbouring pixels correspond to similar depth values. In contrast to the segmentation approaches in [8–10] that create *piecewise constant* depth maps, our method aims a *piecewise smooth* results. Thus, it is able to handle pixels that are in focus between two frames. This causes not only more realistic depth maps, but also allows smooth transitions in the fused image.
Our method is based on a PDE model that features a robust data term in conjunction with a sophisticated anisotropic diffusion term. It relies on a diffusion tensor that allows *joint image- and depth-driven* regularisation: Image edges determine the direction of depth map discontinuities, while the smoothing across these discontinuities is steered by the magnitude of the depth map gradient. We show that this novel strategy leads to substantially better fusion results than methods that do not incorporate smoothness assumptions on the depth field.
In contrast to decomposition methods that may change the underlying images structures by transforming them into another domain, our method relies on the unmodified intensity values and, hence, is able to reproduce the sharp image structures accurately.

Organisation. Our paper is organised as follows: Sec. 2 introduces our new approach for focus fusion. In Sec. 3, we show experimental results of the method on a synthetic as well as on a real–world image set. The paper is concluded with a summary in Sec. 4.

2 Our Focus Fusion Method

Let $f(\boldsymbol{x}, z)$ be a 3-D volume, where $\boldsymbol{x} := (x, y)^\top \in \Omega$ denotes the location within a rectangular image domain $\Omega \subset \mathbb{R}^2$ and $z \in \mathbb{R}$ denotes the depth. Consequently, the n input images $f(\boldsymbol{x}, z_k)$ with $k = 1, \ldots, n$ represent slices of this volume, where we assume z_k to be distributed equidistantly. To remove noise, we pres-mooth each image via a convolution with a Gaussian of standard deviation σ. This results in f_σ. Our goal is to find a depth map $d : \Omega \to [1, n]$ that selects for each location $\boldsymbol{x}$ the frame that is in focus. To this end, we formulate a PDE that models the similarity to a precomputed depth map d_{init}, combined with an anisotropic smoothness constraint.

2.1 Initial Depth Map

Let us now construct the initial depth map that is later embedded in the similarity assumption of our PDE. We exploit the fact that the sharp regions most probably correspond to the locations where the gradient has the largest magnitude. Since each frame corresponds to a certain depth, selecting a gradient at each location can be understood as deciding for the corresponding depth. Thus, it is possible to estimate a depth map of the resulting image.

Two types of locations cause a problem while estimating the initial depth map: homogeneous regions that hardly have any texture, and regions that are never in focus, such as the background. Here the gradients have approximately the same magnitude in all frames, and thus, the final decision is highly influenced by noise. Accordingly, the initial depth map appears noisy within these locations. An example of such a noisy initial depth map obtained from the real-world image set is shown in Fig. 2d. Here, black pixels correspond to the closest frame and white pixels refer to the farthest frame.

Confidence Function. To separate reliable pixels from noisy ones in the initial depth map, we use a confidence function as proposed in [11]. It is defined as

$$c(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad |\boldsymbol{\nabla} f_\sigma(\boldsymbol{x}, d_{\text{init}}(\boldsymbol{x}))| > T \\ 0 & \text{else} \end{cases} \tag{1}$$

where $\boldsymbol{\nabla} := (\partial_x, \partial_y)^\top$ denotes the spatial gradient operator, and T is a threshold. (In the case of colour images, we define the combined gradient magnitude as the square root of the sum over all squared gradient entries.) An example of the previously shown initial depth map after gradient thresholding ($\sigma = 1.0, T = 40$) is presented in Fig. 2e. Here, red colour denotes unreliable pixels that have been eliminated. The proposed confidence function leaves only reliable locations in the initial depth map and thus allows a better modelling. We see that the depth map may become very sparse, but the filling-in effect of the smoothness term of our new model will allow to reconstruct a dense depth map from these sparse data.

2.2 Our PDE-Based Approach

Let us now discuss our PDE-based approach that allows to find the desired depth map. The underlying idea is that there is a spatial continuity between the parts selected from different frames. This means that for each pixel the neighbouring pixels most probably should be chosen from a similar depth level. Thus, we apply a smoothness constraint on the resulting depth map.

We are searching for the depth map $d(\boldsymbol{x})$ which minimises an energy functional of the form

$$E\,(d) \;:=\; \frac{1}{2}\int_{\Omega}\Big(c\,(\boldsymbol{x})\cdot\, M\,(d\,(\boldsymbol{x}))+\,\alpha\;S\,(\boldsymbol{\nabla} d\,(\boldsymbol{x}))\Big)\,\mathrm{d}\boldsymbol{x}\,, \tag{2}$$

where $c(\boldsymbol{x})$ is the proposed confidence function, and $\alpha > 0$ is the smoothness parameter. The data term $M(d(\boldsymbol{x}))$ assumes that $d(\boldsymbol{x})$ should be similar to the initial depth map $d_{\mathrm{init}}(\boldsymbol{x})$. Furthermore, to reduce the influence of outliers we use the regularised L^1-norm $\Psi(s^2)\!:=\!\sqrt{s^2+\varepsilon^2}$ as a penalisation function, where s denotes the data constraint and $\varepsilon > 0$ is a small constant. The data term is finally given by

$$M(d(\boldsymbol{x})) := \Psi\left(\left(d\,(\boldsymbol{x}) - d_{\mathrm{init}}\,(\boldsymbol{x})\right)^2\right). \tag{3}$$

A simple smoothness term enforcing $d(\boldsymbol{x})$ to vary smoothly in space is given by

$$S(\boldsymbol{\nabla} d(\boldsymbol{x})) := |\boldsymbol{\nabla} d(\boldsymbol{x})|^2 \;\;. \tag{4}$$

Minimisation by Gradient Descent. Applying gradient descent to minimise the energy (2) with data term (3) and smoothness term (4) yields the following evolution for $d(\boldsymbol{x}, t)$:

$$\partial_t d \;=\; \alpha\,\Delta d \;-\; c\cdot\Psi'\left((d-d_{\mathrm{init}})^2\right)\cdot(d-d_{\mathrm{init}}). \tag{5}$$

The desired minimiser is obtained as the steady state for $t \to \infty$.

Anisotropic Modification. The smoothness term (4) leads to the homogeneous diffusion operator Δd in (5). It smooths in all directions without respecting image structures. To overcome this problem, we replace it by an anisotropic diffusion term that is inspired by the optic flow approach of Zimmer *et al.* [12]. It allows an adaptation of the diffusion process to the image edges, which are characterised by the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ of the structure tensor [13]

$$\boldsymbol{J}_{\rho,\sigma} := K_\rho * \left(\boldsymbol{\nabla} f_\sigma(\boldsymbol{x}, d)\;\boldsymbol{\nabla} f_\sigma^\top(\boldsymbol{x}, d)\right). \tag{6}$$

Here K_ρ is a Gaussian of standard deviation ρ, and $*$ denotes the convolution operator. We assume that $\boldsymbol{v}_1$ belongs to the larger eigenvalue of $\boldsymbol{J}_{\rho,\sigma}$.
To steer the diffusion process by the directions $\boldsymbol{v}_1, \boldsymbol{v}_2$ that point across and along the image edges respectively, we construct a diffusion tensor

$$\boldsymbol{D} := (\boldsymbol{v}_1 \boldsymbol{v}_2)\begin{pmatrix} g\left((\boldsymbol{v}_1^\top \boldsymbol{\nabla} d)^2\right) & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \boldsymbol{v}_1^\top \\ \boldsymbol{v}_2^\top \end{pmatrix} \;, \tag{7}$$

where g is the Perona-Malik diffusivity [14]

$$g(s^2) := \frac{1}{1 + s^2/\lambda^2} \tag{8}$$

with contrast parameter $\lambda > 0$. It allows to smooth strongly along edges, while reducing the diffusion across edges. We obtain the desired anisotropic evolution by replacing Δd in (5) by its anisotropic counterpart $\operatorname{div}(\boldsymbol{D}\,\boldsymbol{\nabla} d)$:

$$\partial_t d \;=\; \alpha \operatorname{div}\left(\boldsymbol{D}\left(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{\nabla} d\right)\;\boldsymbol{\nabla} d\right) \;-\; c\cdot\Psi'\left((d-d_{\mathrm{init}})^2\right)\cdot(d-d_{\mathrm{init}})\,. \tag{9}$$

The steady state of this evolution depends on the initialisation. Based on our experiments we recommend to initialise it with the depth of the middle frame.

This is our final model for anisotropic depth map smoothing with a robustified fidelity term. It performs a joint image- and depth-driven diffusion in adaptive directions.

Implementation. We implement the anisotropic evolution equation (9) with an explicit finite difference scheme, where the space discretisation of the divergence expression uses the stencil from [15]. In order to avoid any stability deteriorations by the data term, we approximate the expression $(d-d_{\mathrm{init}})$ outside the argument of Ψ' in an implicit way. This still allows an explicit update of d without any need to solve linear or nonlinear systems of equations.

Colour Images. It is easy to extend our model to colour images. We exchange the structure tensor $\boldsymbol{J}_{\rho,\sigma}$ in (6) by the combined structure tensor [16, 17]

$$K_\rho * \sum_i \boldsymbol{\nabla} f^i_\sigma(\boldsymbol{x},d)\;\boldsymbol{\nabla} {f^i_\sigma}^\top(\boldsymbol{x},d)\,, \tag{10}$$

where $f^i\,(\boldsymbol{x},d)$ represents colour channel i.

2.3 Image Fusion

After computing the optimal depth map by means of our PDE, we can easily construct the fused image by combining the colour values directly from the source images. However, our approach computes the continuous depth map that represents the corresponding depth for every pixel, while we have a discrete number of images. Therefore, results at non-integer depth values are obtained by linear interpolation. For colour images, the fusion is performed channelwise.

3 Experimental Results

We test our model with two image sets. For the first experiment we use two synthetically generated images where the ground truth is known (Fig. 1a, 1b). For the second experiment we used a commonly available[1] real-world image set Fig. 2a, 2b).

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

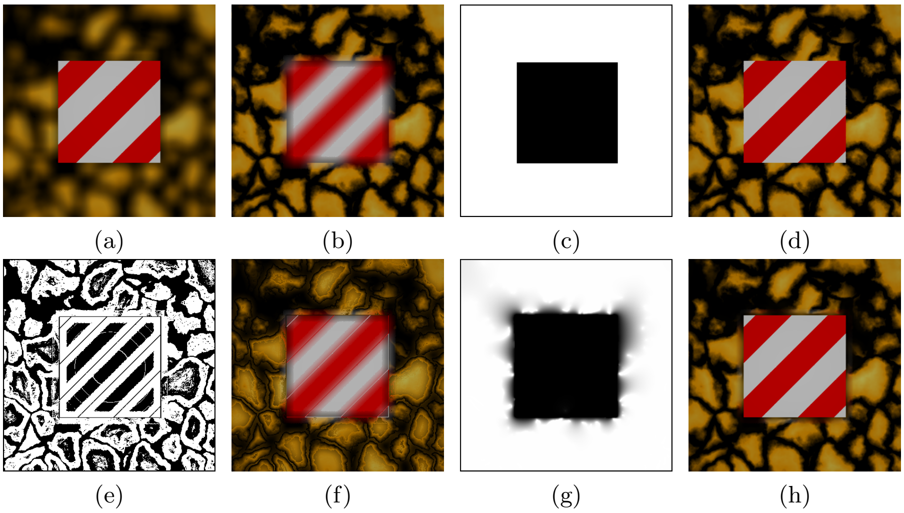

Fig. 1. Results for the synthetic data set. **(a)** Frame 1 with small focal length. **(b)** Frame 2 with large focal length. **(c)** Ground truth depth map. Brighter grey tones describe larger depth values. **(d)** Ground truth image (all-in-focus). **(e)** Initial depth map. **(f)** Fused image obtained with initial depth map. **(g)** Depth map of our approach. **(h)** Our fused image.

Synthetic Data Set. The image set consists of two frames of size 400×400. In the first frame the foreground is in focus, in the second frame the background. For this synthetic images *ground truth* data is available, thus it is possible to rate the quality of our result in terms of an error measure. Figure 1 depicts the ground truth depth map and the *all-in-focus* image (Fig. 1c, 1d), the initial depth map and the corresponding image (Fig. 1e, 1f), as well as the final smoothed depth map and the resulting image (Fig. 1g, 1h). The initial depth map is not reliable, thus it produces an unsharp result with a mean squared error (MSE) 535.42. The final depth map obtained with our approach is very close to the ground truth depth map, and the fused image is sharp with an MSE of 45.88.

Real–World Data Set. The image set consists of 13 images of size 1344×1021 pixels with increasing focal length. Figure 2 demonstrates the results: In the middle row we observe the depth maps corresponding to the different stages of our algorithm (the initial, the thresholded, and the final depth map). We can see that the final depth map is nicely segmented, the noise is removed, and we obtain the desired piecewise smooth result. In the bottom row we observe the magnified details of the fused image obtained with the initial depth map (Fig. 2g), the details of the fused image obtained with our approach (Fig. 2h) as well as the fused image itself (Fig. 2i). Comparing the details we see that the result of our approach is sharper, the fine details are well-preserved, and it contains much less noise.

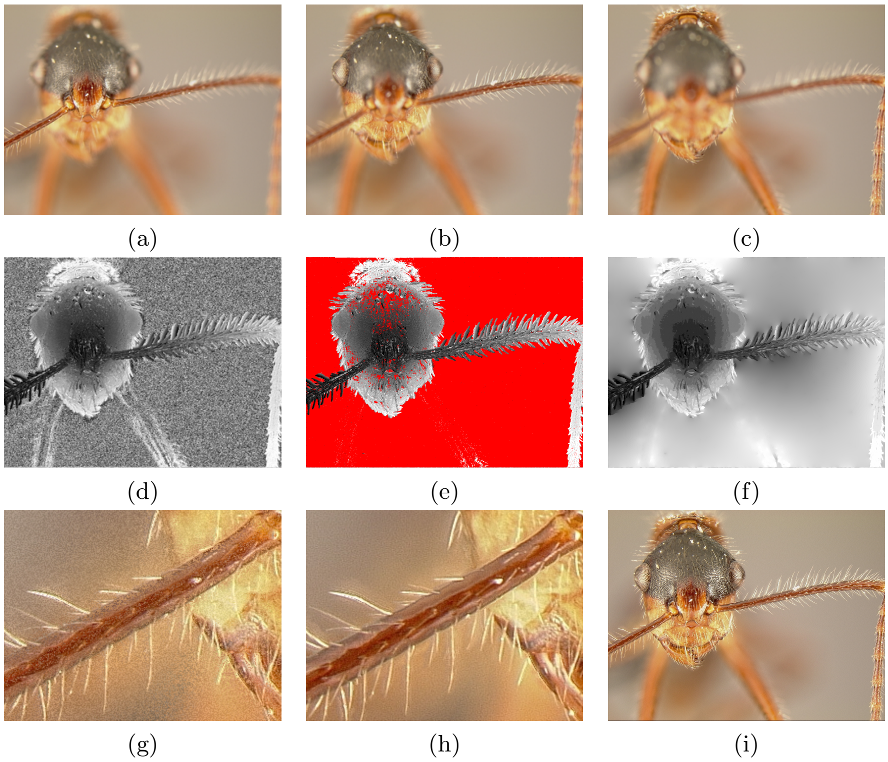

Fig. 2. Results for the real-world data set. **(a)-(c)** Input frames 3, 7, and 13. **(d)** Initial depth map. Brighter pixels correspond to farther frames. **(e)** Thresholded depth map. Red colour denotes unreliable pixels that have been removed. **(f)** Final depth map. **(g)** Zoom into the fused images obtained with initial depth map. **(h)** Zoom with final depth map. **(i)** Fused image.

4 Conclusions and Outlook

We have identified depth map regularisation as an important aspect of focus fusion that has hardly been explored in the literature so far. Applying concepts of modern PDE-based smoothing methods such as robustified fidelity terms and anisotropic smoothness terms, it was possible to improve the quality of the fusion result substantially.

This approach is very general since it can be combined with many fusion criteria. The fact that we have chosen a gradient-based depth map initialisation and direct pixel fusion was for didactic reasons only, since we wanted to keep the method as simple as possible.

Last but not least, focus fusion is only one special application of image fusion. Our approach can also be extended to other fusion tasks such as exposure fusion or superresolution. This is part of our ongoing research.

References

1. Ogden, J.M., Adelson, E.H., Bergen, J.R., Burt, P.J.: Pyramid-based computer graphics. RCA Engineer 30(5), 4–15 (1985)
2. Petrovic, V.S., Xydeas, C.S.: Gradient-based multiresolution image fusion. IEEE Transactions on Image Processing 13(2), 228–237 (2004)
3. Li, H., Manjunath, B.S., Mitra, S.K.: Multisensor image fusion using the wavelet transform. Computer Vision, Graphics, and Image Processing 57(3), 235–245 (1995).
4. Valdecasas, A.G., Marshall, D., Becerra, J.M., Terrero, J.J.: On the extended depth of focus algorithms for bright field microscopy. Micron 32(6), 559–69 (2001)
5. Aslantas, V., Kurban, R.: A comparison of criterion functions for fusion of multi-focus noisy images. Optics Communications 282(16), 3231–3242 (2009)
6. Wang, W.-W., Shui, P.-L., Feng, X.-C.: Variational models for fusion and denoising of multifocus images. IEEE Signal Processing Letters 15, 65–68 (2008)
7. Pop, S., Lavialle, O., Romulus, T., Borda, M.: A PDE-based approach for image fusion. Blanc–Talon, Philips, W., Popescu, D.C., Scheunders, P. (Eds.): Advanced Concepts for Intelligent Vision Systems. Lecture Notes in Computer Science 4678, 121–131 (2007)
8. Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S.M., Colburn, A., Curless, B., Salesin, D., Cohen, M.F.: Interactive digital photomontage. ACM Transactions on Graphics 23(3), 294–302 (2004)
9. Šroubek, F., Cristóbal, G., Flusser, J.: Image fusion based on level set segmentation. In Proc. 14th European Signal Processing Conference, 1–5. Florence, Italy (Sep 2006)
10. Li, S., Bin, Y.: Multifocus image fusion using region segmentation and spatial frequency. Image and Vision Computing 26(7), 971–979 (2008)
11. Barth, O., Osthof, M.: "Focal stack photography" project. "Computational photography" lecture. Saarland University, Germany. http://www.mpi-inf.mpg.de/departments/d4/teaching/ss2010/ComputationalPhotography/projects/focalstack/, (2010)
12. Zimmer, H., Bruhn, A., Weickert, J.: Optic flow in harmony. International Journal of Computer Vision 93(3), 368–388 (July 2011).
13. Förstner, W., Gülch, E.: 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. 281–305 (1987)
14. Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 629–639 (1990)
15. Weickert, J., Welk, M., Wickert, M.: L^2-stable nonstandard finite differences for anisotropic diffusion. In Kuijper, A., Pock, T., Bredies, K., Bischof, H. (Eds.): Scale-Space and Variational Methods in Computer Vision. Lecture Notes in Computer Science, in press (2013)
16. Di Zenzo, S.: A note on the gradient of multi-image. Computer Vision, Graphics, and Image Processing 33, 116–125 (1986)
17. Weickert, J.: Coherence-enhancing diffusion of colour images. Image and Vision Computing 17, 201–212 (1999)

Update for the paper

Focus Fusion with Anisotropic Depth Map Smoothing

On page 3, Subsec. Confidence Function: An example of the previously shown initial depth map after gradient tresholding ($\sigma = 1.0$, $T = 40$) is presented in Fig. 2e.

On page 7, Fig. 2: Fig. 2e changed according to the parameters above, Fig. 2f obtained with the time step size $\tau = 0.25$ and 2500000 iterations, Fig. 2h changed according to the final fused image, Fig. 2i changed according to the final depth map.

The small time step size and a large number of iterations ensure that the algorithm has converged and a steady state of the evolution equation Eq. 9 has been reached.