Variational Depth from Focus Reconstruction

Michael Moeller*, Martin Benning, Carola Schönlieb, Daniel Cremers

August 4, 2014

Abstract

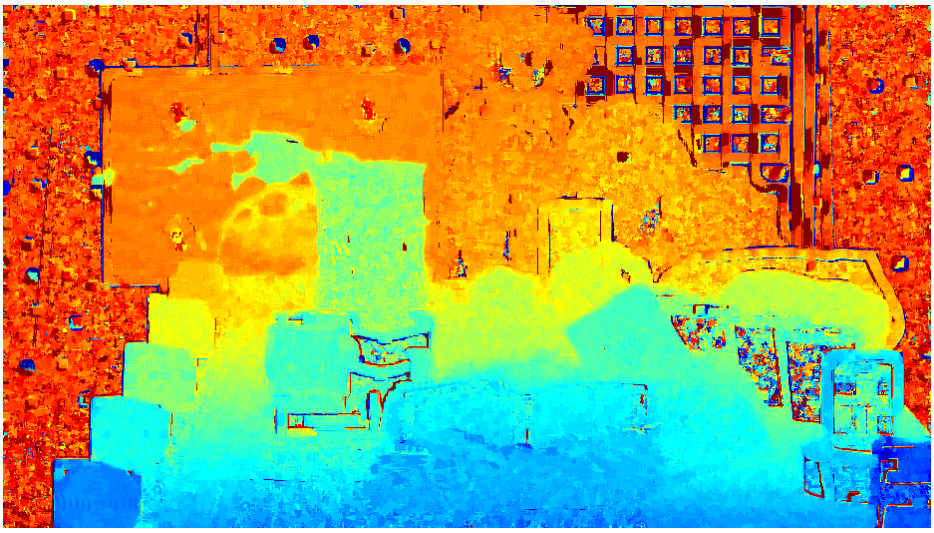
This paper deals with the problem of reconstructing a depth map from a sequence of differently focused images, also known as *depth from focus* or *shape from focus*. We propose to state the depth from focus problem as a variational problem including a smooth but nonconvex data fidelity term, and a convex nonsmooth regularization, which makes the method robust to noise and leads to more realistic depth maps. Additionally, we propose to solve the nonconvex minimization problem with a novel linearized alternating directions method of multipliers (ADMM), allowing to minimize the energy very efficiently. Numerical results are presented on real data.

1 Introduction

The basic idea for *depth from focus* approaches is to assume that the distance of an object to the camera at a certain pixel corresponds to the focal setting at which the pixel is maximally sharp. Thus, for a given data set of differently focused images, one typically first finds a suitable contrast measure for the sharpness of each image at each pixel. Then, the depth at each pixel is determined by finding the focal distance at which the contrast measure is maximal. Figure 1 illustrates this approach. Depth from focus differs from *depth from defocus* (cf. [1–3]) in the sense that many images are given and depth clues are obtained from the sharpness at each pixel. Depth from defocus on the other hand tries to estimate the variance of a spatially varying blur based on a physical model and uses only very few images. Generally, the measurements of differently focused images do not necessarily determine the depth of a scene uniquely, such that the estimation of a depth map is an ill-posed problem. The ill-posedness is particularly strong in textureless areas.

*M.M. (email michael.moeller@in.tum.de) and D.C. are with the Department of Computer Science, Technische Universität München, Boltzmannstrasse 3, 85748 Garching. M.B. is with the Magnetic Resonance Research Centre, Department of Chemical Engineering and Biotechnology, c/o Cavendish Stores, JJ Thomson Avenue, Cambridge, CB3 0HE, United Kingdom. C.S. is with the Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom.

(a) Visualization of the data as a cube.



(b) Depth map reconstruction by finding the maximal contrast.

Figure 1: Example of a simple depth from focus reconstruction: The data set of images can be visualized as a data cube (a), where the z-direction corresponds to a change of focus from the front to the back. In order to determine the focal setting at which the contrast is maximal, one picks a contrast measure, applies a windowed filtering to the contrast coefficients and selects the focal setting for which the coefficients are maximal. By knowing the distance at which a region appears maximally sharp in a focal setting, one reconstructs the depth. Figure (b) shows an example of the depth map, where red values indicate being far away from the camera and blue values correspond to pixels close to the camera. The result was obtained by using the modified Laplacian contrast measure with 9×9 windowed filtering as used in the comparison in [4].

The literature dealing with the shape from focus problem has very different contributions. A lot of work has targeted the development of different measures for sharpness and focus (cf. [4] for an overview). Other works deal with different methods of filtering the contrast coefficients before determining the maximum, i.e. the depth. The ideas range from windowed averaging (e.g. [5]), over differently shaped kernels, to nonlinear filtering as proposed in [6]. Ideas for using different focus measures and fusing the resulting depth estimates in a variational approach using the total variation (TV) regularization have been proposed in [7]. However, very little work has been done on finding a noisefree depth map in a single variational framework.

The idea of variational methods is to find an energy E that depends on the (unknown) depth map d such that a low value of $E(d)$ corresponds to a depth map with desired properties. The final estimate for the depth is determined by finding the argument that minimizes the energy. Formulating the shape from focus problem in a variational framework has the advantage that one clearly defines a cost function and tries to find a solution which is optimal with respect to the costs. More importantly, regularity can be imposed on the depth estimate itself, e.g. by favoring piecewise constant solutions. Additionally, our framework is robust to noise and controls the ill-posedness of the shape from focus problem.

While variational methods belong to the most successful class of methods for general image reconstruction tasks, very little work has been done on using them for the shape from focus problem. The only work the authors are aware of is the method proposed in [8], where the framework of Markov randomfields was used to derive an energy minimization method consisting of two truncated quadratic functionals. However, using two nonconvex functions in a setting where the dependency of the depth on the contrast is already nonconvex, results in a great risk of only finding poor local minima. For instance, any initialization with pixels belonging to both truncated parts is a critical point of such an approach.

This paper has two contributions: First, we propose a variational framework for the shape from focus problem that is robust to noise and can easily be adapted to different prior assumptions on the type of depth map we expect. In our experiments we will then focus on the total variation (TV) [9] as a particular prior. Secondly, since modeling the shape from focus problem in a variational framework leads to a nonconvex energy, we discuss the problem of minimizing the resulting energy efficiently. While schemes that can guarantee the convergence to critical points are computationally very expensive, we propose to tackle the minimization problem by an alternating minimization method of multipliers (ADMM) with an additional linearization in the nonconvex part.

2 A Variational Approach

2.1 Proposed Energy

We propose to define a functional $E : \mathbb{R}^{n\times m} \to \mathbb{R}$ that maps a depth map $d \in \mathbb{R}^{n\times m}$ to an energy, where a low energy corresponds to a good depth estimate. The dimension of the depth map, $n \times m$, coincides with the dimension of each image. The energy consists of two

terms, $E = D + \alpha R$. The *data fidelity term* D is taking the dependence on the measured data into account and the *regularization* term R is imposing the desired smoothness. The parameter α determines the trade-off between regularity and fidelity. We find the final depth estimate as the argument that minimizes our energy:

$$\hat{d} = \arg\min_d D(d) + \alpha R(d) \tag{1}$$

Typical approaches from literature find the depth estimate by maximizing some contrast measure and we refer the reader to [4] for an overview of different contrast measures and their performance. Since we want to reconstruct the depth map by an energy minimization problem (rather than maximization) it seems natural to choose the negative contrast at each pixel as the data fidelity term.

$$D(d) = -\sum_i \sum_j c_{i,j}(d_{i,j}), \tag{2}$$

where $c_{i,j}$ denotes the (precomputed) function that maps a depth at pixel (i, j) to the corresponding contrast value. With this choice our method is a generalization of methods that maximize the contrast at each pixel separately, since they are recovered by choosing $\alpha = 0$.

The regularization term R imposes some smoothness on the recovered depth map and should therefore depend on the prior knowledge we have about the expected depth maps. In this paper we use the discrete isotropic total variation (TV), $R(d) = \|Kd\|_{2,1}$, where K is the linear operator such that Ku is a matrix with an approximation to the x-derivative in the first column, to the y-derivative in the second column and $\|g\|_{2,1} := \sum_i \sqrt{\sum_j (g_{i,j})^2}$.

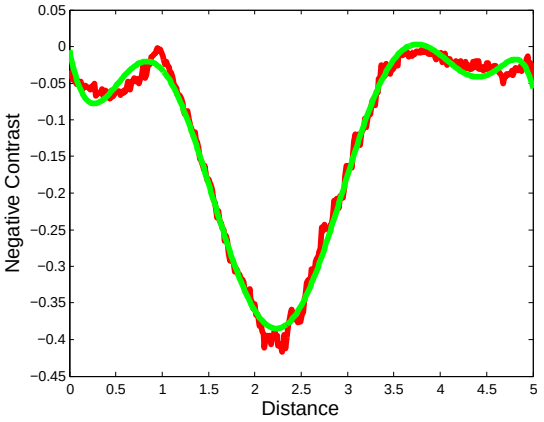
2.2 Contrast measure

Since the main objective of this paper is the introduction of variational methods for the depth from focus problem, we choose the well-known modified Laplacian (ML) function [10] as a measure of contrast. Note that under good imaging conditions the thorough study in [4] found that Laplacian based operators like the modified Laplacian consistently were among the contrast measures yielding the best performances. We determine

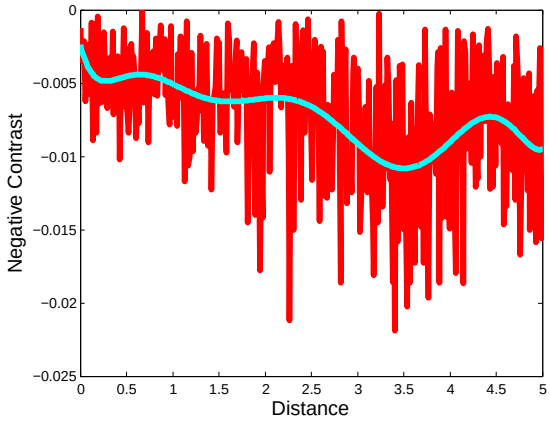
$$\mathrm{ML}(i, j, k) = \sum_l |(\partial_{xx} I)(i, j, l, k)| + |(\partial_{yy} I)(i, j, l, k)|$$

with $(\partial_{xx} I)(i, j, l, k)$ denoting the approximate second derivative of the l-th color channel of the k-th image in the focus sequence in x-direction at pixel (i, j) using a filter kernel $[-1,\ 2,\ -1]$. For each pixel (i, j) we determine the continuous contrast function $c_{i,j}(x)$ depending on a depth x by an eighth order polynomial approximation to the $\mathrm{ML}(i, j, k)$ data. Polynomial approximations for determining the contrast curve have previously been proposed in [11] and offer a good and computationally inexpensive continuous representation of the contrast, which we need for the variational formulation (2). Different from [11] we use a higher order polynomial to approximate a noise free contrast curve using all data

points, see Figure 2 for an example. Obviously, areas with lots of texture are well suited for determining the depth. In areas with no texture (Figure 2 (b)) the contrast measure consists of mostly noise. The magnitude of the contrast values is very low in this case, such that in a variational framework its influence will be small.



(a) Contrast curve at green point



(b) Contrast curve at cyan point

(c) Example image to illustrate the positions of the example points.

Figure 2: Examples for how the contrast curves look at different pixels. While pixels with a lot of surrounding structure (green point) result in a strong and clear contrast curve with an easy to identify minimum, textureless areas (cyan point) cause problems (as expected).

We can see that the contrast values can be well approximated by smooth curves such as splines or - for the sake of simplicity in the minimization method - by a higher order polynomial. However, we can also see that even the smoothed contrast curves will not be convex. Thus, our data fidelity does not couple the pixels, is smooth, but nonconvex. The regularization term on the other hand is convex, but couples the pixels and is nonsmooth, i.e. not differentiable in the classical sense. In the next section we will propose an efficient algorithm that exploits the structure of the problem to quickly determine depth maps with low energies.

3 Numerical Minimization

In the literature of nonconvex optimization, particularly the one related to image processing, there exist approaches for problems of the form smooth nonconvex plus nonsmooth convex that provable convergence to critical points under mild conditions. Such approaches include for instance forward-backward splittings (FBS), see [12] and the references therein, and methods based on the difference of convex functions (DC), e.g. [13]. The drawback of these methods is that they require to solve a problem like TV denoising at each iteration which makes these schemes computationally expensive. A simple FBS approach that provably converges to a critical point would be

$$d^{k+1} = \arg\min_d \frac{1}{2}\|d - d^k + \tau \nabla D(d^k)\|^2 + \tau\alpha R(d).$$

While one avoids the difficulty of dealing with the nonconvex term by linearizing it, it can be seen that one still has to minimize the sum of a quadratic fidelity and a nonsmooth regularization term, which can be costly. Therefore, we propose to apply the alternating directions method of multipliers (ADMM) (cf. [14]) in the same fashion as if the energy was convex. We introduce a new variable g along with the constraint $g = Kd$ and iteratively enforce the constraint by using the augmented Lagrangian method. The minimization for g and d is done in an alternating fashion. Additionally, we incorporate the FBS idea of linearizing the nonconvex term, such that the final computational scheme becomes

$$\begin{aligned} d^{k+1} =& \arg\min_d \frac{\lambda}{2}\|Kd - g^k + b^k\|_2^2 \\ &\qquad + \frac{1}{2}\|d - d^k + \tau\nabla D(d^k)\|^2, \\ g^{k+1} =& \arg\min_g \frac{\lambda}{2}\|g - Kd^{k+1} + b^k\|_2^2 + \alpha\tau\|g\|_{2,1}, \\ b^{k+1} =& b^k + (Kd^{k+1} - d^{k+1}). \end{aligned} \tag{3}$$

Note that each subproblem can be solved very efficiently. The update for d involves the inversion of the operator $(\lambda K^T K + I)$ which can be done using a discrete cosine transform, while the update for g has a closed form solution known as soft shrinkage of $z = Kd^{k+1} - b^k$,

$$g_{i,j} = \operatorname{sign}(z_{i,j}) \cdot \max(\|z_{i,:}\|_2 - \alpha\tau/\lambda, 0).$$

In this sense, the proposed algorithm can also be interpreted as solving each FBS subproblem with a single ADMM iteration. We simply omit converging to the minimum of the convex energy but compute a new linearization of the data fidelity term at each iteration. Although the techniques contributing to the algorithm are all known, the authors are not aware of any literature using the method as stated above - even in the convex setting. Hence, let us briefly state a convergence result in the convex setting:

Theorem 3.1. *If D and R are convex and the symmetric Bregman distance with respect to D, $S_D(d, v) = \langle d - v, p_d - p_v \rangle$ with $p_d \in \partial D(d)$, $p_v \in \partial D(v)$, can be bounded by $\frac{1}{2\tau}\|d - v\|^2$ for all d and v, then the iterates d^k produced by the linearized ADMM scheme* (3) *converge to a solution $\hat{d}$ of the minimization problem* (1) *with*

$$\tau \left(S_D(d^k, \hat{d}) + \alpha S_R(g^k, K\hat{d}) \right) \leq C/k,$$

and $\lambda\|Kd^{k+1} - g^k\|^2 \leq C/k$ for some constant C.

The proof follows the arguments of Cai et al. in [15], and is given in the appendix.

While the idea of applying splitting methods to nonconvex problems as if they were convex is not new, a convergence theory is still an important open problem. Recently, progress has been made for particular types of nonconvexities, e.g. [16], which however do not apply to our case. However, as shown in the next section the proposed method numerically results in a stable optimization scheme that is much more efficient than methods like FBS which have to solve a full convex minimization problem at each iteration.

4 Numerical Results

We test the proposed scheme on a data set taken with the ARRI Alexa camera[1]. The data set consists of 373 images of a tabletop scene. The focus was continuously and evenly changed from the front to the back of the scene using a wireless compact unit (WCU-4) connected to a motor for focus adjustments. The images ran through the usual ARRI processing chain, which drastically influences the data characteristics as observed in [17]. Example images from the data set were already used for Figures 1 and 2. Note that some laser dots were used to investigate whether they can improve the depth estimates in textureless areas, which, as we can see in Figure 1 (b), unfortunately is not the case.

We set up the variational model as proposed in Section 2, and minimize it computationally as described in Section 3. Figure 3 shows the resulting depth maps after 2000 iterations, using $\tau = 40$, $\alpha = 1/40$, $\lambda = 150$ (upper image) and $\tau = 4$, $\alpha = 1/640$, $\lambda = 80$ (lower image), initializing with the image from Figure 1 (b).

[1]Provided by the Arnold and Richter Cine Technik, www.arri.com

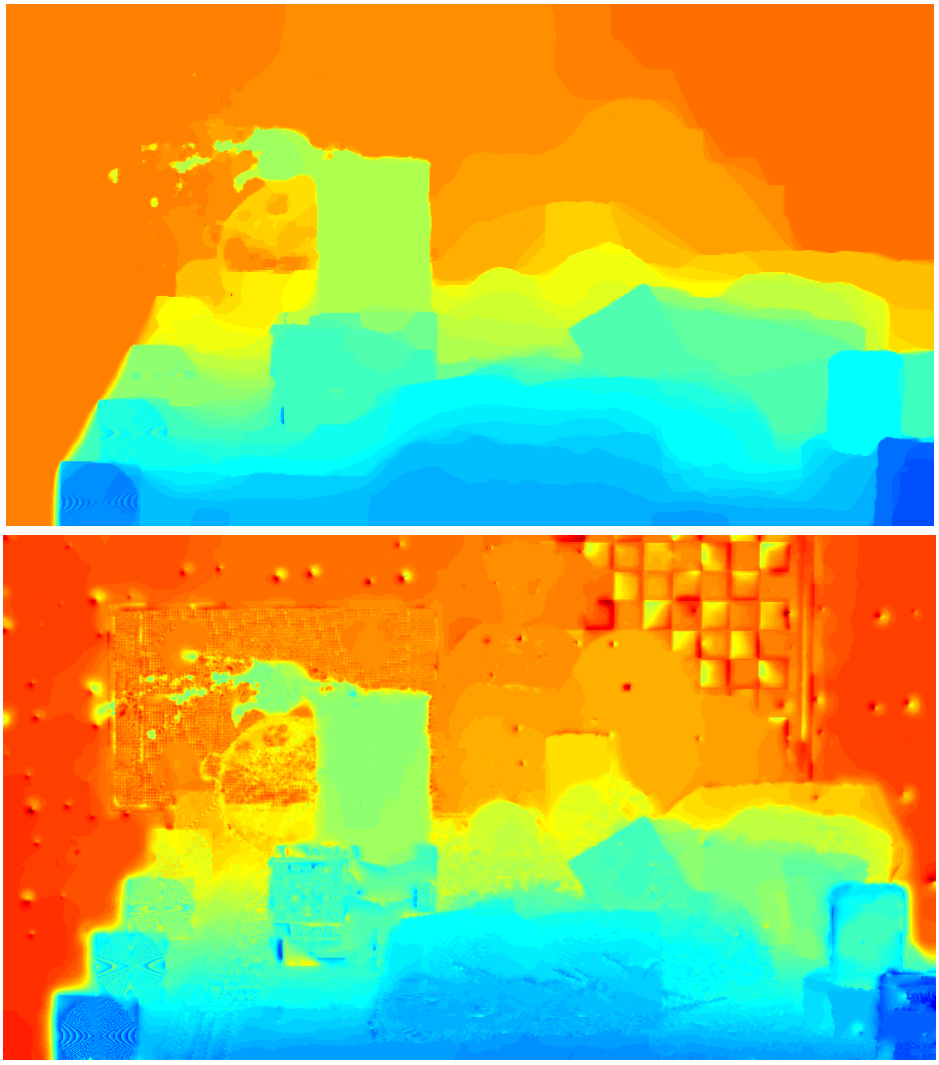
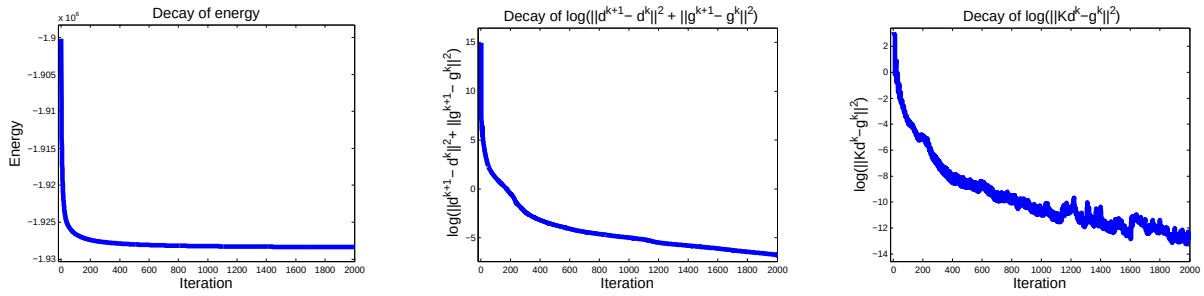


Figure 3: Examples for variational depth reconstructions with high and low regularization parameters (upper and lower image respectively).

As we can see in the upper image of Figure 3 the variational approach succeeded in finding a much smoother solution than Figure 1 eliminating all noise in the image. Particularly impressive is the fact that it even restored the wall in the background, the checkerboard and the plastic cups in the foreground. All these things were not reconstructed correctly in the windowed approach as we can see in Figure 1. Due to the piecewise smooth structure, the depth map in Figure 3 is much better suited for applications like 3d image warping. Note that to a human observer Figure 1 might partly look sharper at first glance, however, distinguishing neighboring objects should not be possible in a realistic depth map as long as the objects are at the same distance to the camera. Particularly, some object edges in Figure 1 can be seen as red or blue lines, which are due to erroneous depth estimates. The lower image of Figure 3 shows a depth map obtained with parameters that force the reconstruction



(a) Decay of energy vs. iterations (b) Decay of $\|g^{k+1} - g^k\|^2 + \|d^{k+1} - d^k\|^2$ (c) Decay of $\|Kd^k - g^k\|^2$

Figure 4: Convergence plots for linearized ADMM algorithm applied to the variational depth from focus problem. As we can see the total energy as well as the difference of successive iterates show a nice and monotonic decay. The decay of $\|Kd^k - g^k\|$ shows some oscillations but only for cases where $\|Kd^k - g^k\|^2 \leq 10^{-9}$.

to stay much closer to a depth map that maximizes the contrast. Although there are some erroneous parts (particularly in the checkerboard), the variational approach got rid of large parts of the noise. Comparing the upper and lower image, one can see the drawback of a large regularization parameter leading to a loss of contrast, i.e. the range of depth values seems compressed. This phenomenon is well known in convex total variation denoising. In future work we will incorporate Bregman iteration for contrast correction into the nonlinear variational model.

Regarding the convergence of the algorithm, we plotted the decay of energy in Figure 4 on the left. As we can see the algorithm shows nice decay properties. Also included in Figure 4 are the decays of $\|g^{k+1} - g^k\|^2 + \|d^{k+1} - d^k\|^2$ (middle) and $\|Kd^k - g^k\|^2$ (right), both plotted in a logarithmic fashion. Although $\|Kd^k - g^k\|^2$ shows some oscillations in the logarithmic plot, it decayed nicely to about 10^{-12}. Since both types of residuals approach zero, one can verify from the optimality conditions arising from the updates in (3) that this provably leads to convergence to a critical point of our nonconvex energy (along subsequences).

5 Conclusions & Outlook

In this paper we proposed a variational approach to the shape from focus problem, which can be seen as a generalization of current common shape from focus approaches. It uses an efficient nonconvex minimization scheme to determine depth maps which are based on prior knowledge such as a realistic depthmap often being piecewise constant. In future work we will incorporate more sophisticated regularizations such as the total generalized variation in our studies. Additionally, we'll work on correcting the inherent loss of contrast caused by the total variation regularization by applying nonlinear Bregman iterations in the fashion of [18] to our nonconvex optimization problem.

Acknowledgements

M.B. and C.S. were supported by EPSRC GrantEP/F047991/1, and the Microsoft Research Connections. C.S. additionally acknowledges the support of the KAUST Award No. KUK-I1-007-43. M.M. and D.C. were supported by the ERC Starting Grant "ConvexVision".

References

[1] P. Favaro and S. Soatto, *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-blur*. Secaucus, NJ, USA: Springer New York, Inc., 2006.

[2] P. Favaro, S. Soatto, M. Burger, and S. Osher, "Shape from defocus via diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 518–531, 2008.

[3] X. Lin, J. Suo, X. Cao, and Q. Dai, "Iterative feedback estimation of depth and radiance from defocused images," in *Computer Vision ACCV 2012*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7727, pp. 95–109.

[4] S. Pertuz, D. Puig, and M. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, vol. 46, pp. 1415–1432, 2013.

[5] A. Thelen, S. Frey, S. Hirsch, and P. Hering, "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation," *IEEE Trans. on I. Proc.*, vol. 18, no. 1, pp. 151–157, 2009.

[6] M. Mahmood and T.-S. Choi, "Nonlinear approach for enhancement of image focus volume in shape from focus," *IEEE Trans. on Image Proc.*, vol. 21, no. 5, pp. 2866–2873, 2012.

[7] M. Mahmood, "Shape from focus by total variation," in *2013 IEEE 11th IVMSP Workshop*, 2013, pp. 1–4.

[8] V. Gaganov and A. Ignatenko, "Robust shape from focus via markov random fields," in *GraphiCon'2009*, 2009, pp. 74–80.

[9] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[10] S. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, 1994.

[11] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 266–274, 1995.

[12] H. Attouch, J. Bolte, and B. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Math. Prog.*, vol. 137, no. 1-2, pp. 91–129, 2013.

[13] T. P. Dinh, H. Le, H. L. Thi, and F. Lauer, "A Difference of Convex Functions Algorithm for Switched Linear Regression," to appear in IEEE Transactions on Automatic Control.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[15] J. Cai, S. Osher, and Z. Shen, "Split bregman methods and frame based image restoration," *Multiscale Modeling & Simulation*, vol. 8, no. 2, pp. 337–369, 2010.

[16] T. Valkonen, "A primal-dual hybrid gradient method for non-linear operators with applications to mri," *Inverse Problems*, vol. 30, no. 5, p. 055012, 2014.

[17] S. Andriani, H. Brendel, T. Seybold, and J. Goldstone, "Beyond the kodak image set: A new reference set of color image sequences," in *2013 20th IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 2289–2293.

[18] M. Bachmayr and M. Burger, "Iterative total variation schemes for nonlinear inverse problems," *Inverse Problems*, vol. 25, 2009.

Appendix: Proof of Theorem (3.1)

Proof. Let $\hat{d}$ be a critical point and denote $\hat{g} = K\hat{d}, \hat{b} \in \frac{\alpha\tau}{\lambda}\partial\|\hat{g}\|_{2,1}$. Furthermore, denote $d_e^k = d^k - \hat{d}$, $g_e^k = g^k - \hat{g}$, $b_e^k = b^k - \hat{b}$, $p_e^k = \nabla D(d^k) - \nabla D(\hat{d})$ and $q_e^k \in \partial\|g^k\|_{2,1} - \partial\|\hat{g}\|_{2,1}$. Then the optimality conditions arising from algorithm (3) yield

$$\begin{aligned} 0 =&\lambda K^T(Kd_e^{k+1} - g_e^k + b_e^k) + d_e^{k+1} - d_e^k + \tau p_e^k \\ 0 =&-\lambda(Kd_e^{k+1} - g_e^{k+1} + b_e^k) + \alpha\tau q_e^{k+1} \end{aligned}$$

Taking the inner product of the first equation with d_e^{k+1} yields

$$\begin{aligned} 0 =&\lambda\langle Kd_e^{k+1} - g_e^k + b_e^k, Kd_e^{k+1}\rangle + \|d_e^{k+1}\|^2 - \langle d_e^k, d_e^{k+1}\rangle + \tau\langle p_e^k, d_e^{k+1}\rangle \\ =&\lambda\langle Kd_e^{k+1} - g_e^k + b_e^k, Kd_e^{k+1}\rangle + \frac{1}{2}\left(\|d_e^{k+1}\|^2 + \|d_e^k - d_e^{k+1}\|^2 - \|d_e^k\|^2\right) + \tau\langle p_e^k, d_e^{k+1}\rangle \\ =&\frac{\lambda}{2}\left(\|Kd_e^{k+1} - g_e^k\|^2 + \|Kd_e^{k+1}\|^2 - \|g_e^k\|^2\right) + \lambda\langle b_e^k, Kd_e^{k+1}\rangle \\ &+ \frac{1}{2}\left(\|d_e^{k+1}\|^2 + \|d_e^k - d_e^{k+1}\|^2 - \|d_e^k\|^2\right) + \tau\langle p_e^k, d_e^{k+1}\rangle \end{aligned}$$

The inner poduct of the second equation with g_e^{k+1} results in

$$\begin{aligned} 0 =&-\lambda\langle Kd_e^{k+1} - g_e^{k+1} + b_e^k, g_e^{k+1}\rangle + \alpha\tau\langle q_e^{k+1}, g_e^{k+1}\rangle \\ =&\frac{\lambda}{2}\left(\|Kd_e^{k+1} - g_e^{k+1}\|^2 + \|g_e^{k+1}\|^2 - \|Kd_e^{k+1}\|^2\right) - \lambda\langle b_e^k, g_e^{k+1}\rangle + \alpha\tau\langle q_e^{k+1}, g_e^{k+1}\rangle \end{aligned}$$

Adding the two estimates above leads to

$$\begin{aligned} 0 =&\frac{\lambda}{2}\left(\|Kd_e^{k+1} - g_e^{k+1}\|^2 + \|g_e^{k+1}\|^2 - \|g_e^k\|^2 + \|Kd_e^{k+1} - g_e^k\|^2\right) + \lambda\langle b_e^k, Kd_e^{k+1} - g_e^{k+1}\rangle \\ &+ \frac{1}{2}\left(\|d_e^{k+1}\|^2 + \|d_e^k - d_e^{k+1}\|^2 - \|d_e^k\|^2\right) + \tau\langle p_e^k, d_e^{k+1}\rangle + \alpha\tau S_R(g^{k+1}, \hat{g}), \end{aligned} \tag{4}$$

where we used the fact that $\langle q_e^{k+1}, g_e^{k+1}\rangle = S_R(g^{k+1}, \hat{g})$. We use the update formula for b^{k+1} to obtain that

$$\begin{aligned} \|b_e^{k+1}\|^2 &= \|b_e^k\|^2 + \|Kd_e^{k+1} - g_e^{k+1}\|^2 - 2\langle b_e^k, Kd_e^{k+1} - g_e^{k+1}\rangle, \\ \Rightarrow\langle b_e^k, Kd_e^{k+1} - g_e^{k+1}\rangle &= \frac{1}{2}(\|b_e^{k+1}\|^2 - \|b_e^k\|^2 - \|Kd_e^{k+1} - g_e^{k+1}\|^2). \end{aligned} \tag{5}$$

Due to the linearization we have the term $\langle p_e^k, d_e^{k+1}\rangle$ in our current estimate instead of the symmetric Bregman distance (which we would obtain by replacing p_e^k by p_e^{k+1}. Thus, we estimate

$$\begin{aligned} \langle p_e^k, d_e^{k+1}\rangle =&S_D(d^k, \hat{d}) + \langle p_e^k, d_e^{k+1} - d_e^k\rangle \\ =&S_D(d^k, \hat{d}) - S_D(d^k, d^{k+1}) + \langle p^{k+1}, d^{k+1} - d^k\rangle - \langle \hat{p}, d^{k+1} - d^k\rangle \\ \geq&S_D(d^k, \hat{d}) - S_D(d^k, d^{k+1}) + D(d^{k+1}) - D(d^k) - \langle \hat{p}, d^{k+1} - d^k\rangle, \end{aligned} \tag{6}$$

where we used $p^{k+1} \in \partial D(d_{k+1})$ along with the convexity of D for the last inequality. Inserting (5) and (6) into (4) yields

$$\begin{aligned}
0 \geq & \frac{\lambda}{2}\left(\|Kd_e^{k+1} - g_e^{k+1}\|^2 + \|g_e^{k+1}\|^2 - \|g_e^k\|^2 + \|Kd_e^{k+1} - g_e^k\|^2\right) \\
& + \frac{\lambda}{2}(\|b_e^{k+1}\|^2 - \|b_e^k\|^2 - \|Kd_e^{k+1} - g_e^{k+1}\|^2) \\
& + \frac{1}{2}\left(\|d_e^{k+1}\|^2 - \|d_e^k\|^2\right) + \frac{1}{2}\|d_e^k - d_e^{k+1}\|^2 - \tau S_D(d^k, d^{k+1}) \\
& + \tau S_D(d^k, \hat{d}) + \alpha\tau S_R(g^{k+1}, \hat{g}) + \tau(D(d^{k+1}) - D(d^k) - \langle \hat{p}, d^{k+1} - d^k\rangle) \\
\geq & \frac{\lambda}{2}\left(\|g_e^{k+1}\|^2 - \|g_e^k\|^2 + \|b_e^{k+1}\|^2 - \|b_e^k\|^2 + \|Kd_e^{k+1} - g_e^k\|^2\right) + \frac{1}{2}\left(\|d_e^{k+1}\|^2 - \|d_e^k\|^2\right) \\
& + \tau S_D(d^k, \hat{d}) + \alpha\tau S_R(g^{k+1}, \hat{g}) + \tau\left(D(d^{k+1}) - D(d^k) - \langle \hat{p}, d^{k+1} - d^k\rangle\right),
\end{aligned}$$

where we used the assumption $\frac{1}{2}\|d_e^{k+1} - d_e^k\|^2 - \tau S_D(p^k, p^{k+1}) \geq 0$ for the second inequality. Now we can sum over this inequality from $k = 0$ to n to obtain

$$\begin{aligned}
0 \geq & \frac{\lambda}{2}(\|g_e^{n+1}\|^2 - \|g_e^0\|^2 + \|b_e^{n+1}\|^2 - \|b_e^0\|^2) + \frac{1}{2}(\|d_e^{n+1}\|^2 - \|d_e^0\|^2) \\
& + \frac{\lambda}{2}\sum_{k=0}^{n}\|Kd_e^{k+1} - g^k\|^2 + \tau\sum_{k=0}^{n}\left(S_D(d^k, \hat{d}) + \alpha S_R(g^{k+1}, \hat{g})\right) \\
& + \tau\underbrace{\left(D(d^{n+1}) - D(\hat{d}) - \langle \hat{p}, d^{n+1} - \hat{d}\rangle\right)}_{\geq 0} - \tau\left(D(d^0) - D(\hat{d}) - \langle \hat{p}, d^0 - \hat{d}\rangle\right),
\end{aligned}$$

such that we finally obtain

$$\begin{aligned}
& \frac{\lambda}{2}(\|g_e^0\|^2 + \|b_e^0\|^2) + \frac{1}{2}\|d_e^0\|^2 + \tau\left(D(d^0) - D(\hat{d}) - \langle \hat{p}, d^0 - \hat{d}\rangle\right) \\
\geq & \frac{\lambda}{2}\sum_{k=0}^{n}\|Kd_e^{k+1} - g^k\|^2 + \sum_{k=0}^{n}\left(S_D(d^k, \hat{d}) + \alpha S_R(g^{k+1}, \hat{g})\right).
\end{aligned}$$

Since the sums are bounded for all n and the summands are nonnegative, we can conclude their convergence to zero with a rate as least as fast as $1/k$, which yields the assumption. $\square$