DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

Depth-Guided Deblurring

Thomas Hach; Arnold & Richter Cine Technik (ARRI); Munich, Germany
Arvind Amruth; Technical University Munich (TUM); Munich; Germany

Abstract

In this paper, we propose a novel image deblurring framework, which noticeably improves the effectiveness and efficiency of state-of-the-art approaches. In professional imaging with its typical shallow depth-of-field, it is challenging to estimate the exact focus distance during recording, which often implies costly re-shooting. For the correction of blurred material in post-production, there exist a few deblurring methods, which are, however, challenged by working on real camera data due to noise and the general ill-posedness of the deblurring problem itself. Since the effective out-of-focus operating range of deblurring methods is small and the blur characteristics are strongly depth-dependent, we introduce a framework where a depth map and measured lens characteristics ingest into a selection of state-of-the-art deblurring methods. Therefor, we introduce a depth-dependent parameter selection concerning blur kernels and smoothing weights first. Second, using these parameters the out-of-focus areas are selectively deblurred in order to overcome the emergence of strong artifacts. The foundation for the provided evaluation is formed by a dataset with eight real images captured with a cinematic RGB plus depth camera containing multi-planar and in-scene depth-varying image content. Therein, we show visually and numerically that introducing our depth framework improves the deblurring performance and suppresses typical strong artifacts.

Introduction

In digital imaging, the physical representation of the world is projected onto the silicon sensor by the lens. In general, an image is naturally divided in focused and blurred regions. This holds for all imaging system within a certain distance interval around the focal distance, the projection is depicted at the imaging system's maximum modulation transfer function, which is related to the subjective term "sharpness". Elsewhere, the captured image becomes increasingly blurred, which is known as blur from defocus.

Modern consumer-grade cameras and mobile phone cameras are nowadays equipped with state-of-the-art auto-focus systems like phase auto-focus. Contrary in professional cinematography, focus is still adjusted manually. Although focusing is very reliable in the first case and utilizing a lot of human experience as well as high-level semantics in the second case, both approaches are prone to erronous focus estimation. Automatic systems may fail on assumption mismatch while human estimation is losing accuracy with, for example, increasing distance. Thus artifact-free deblurring approach are still highly required.

The approach presented in this work is biased to cinematic image creation while not being limited to it. In this particular case, we face the typical situation where the focus adjustment is slightly off its target distance. Combined with common shallow depth-of-field settings, the intended in-focus object is then blurred, depending on its depth profile. For example, targeting an actor's eyes often results in blurring the latter and focusing his ears. As research on deblurring shows, this problem fits to the restoration capabilities provided by naturally ill-posed deblurring approaches.

Furthermore, as blur caused by defocusing is depth dependent, we propose a system that acquires all necessary information to significantly reduce the ill-posedness. Thus, we employ a camera capturing color and dense depth information without occlusions at, in the context of the blur process, highly important object boundaries. Finally, the degree of blur introduced by the camera's lens is equivalent to the lens' point-spread function (PSF). Hence, we encounter by measuring those PSFs physically.

Our contribution in this paper is the finalization of 4 state-of-the-art deblurring methods in order to use prior information in form of measured depth and PSF information. Our dedicated evaluation, which uses an acquired deblurring data set with depth and color image ground truth information, underlines the achieved improvements in using 5 different methods with increasing complexity to utilize the aforementioned auxiliary information.

Prior Art

This section gives insight in the context of non-blind deblurring methods first. Afterwards, related methods in blind deblurring as well as methods using additional priors are discussed.

Non-Blind Deblurring

Already used for sharpening of analog film material decades ago, unsharp masking (USM), today, is still present as a standard method in relevant digital image processing tools. In USM, a sparse high-pass component of the image is added to the observed representation. We include USM in our testbed because it is still considered as significant method in image processing. All included methods are referred to as *selected methods* throughout the paper.

Van-Cittert iterations reveal the latent image by convolution with an iteratively optimized operator to mach the observed image [27]. The proposal of Van-Cittert converges to a typical objective function which is often used in MAP estimators for deblurring [9]. In contrast to the more recent methods, there is an absence of prior assumptions and hence only one free parameter.

The Wiener filter is designed to minimize the expected error $\mathbb{E}\{|L(f)-\hat{L}(f)|^2\}$ of the unknown sharp image L and the deblurred estimate $\hat{L}(f)=G(f)B(f)$ [30], where G denotes the transfer function of the Wiener filter and B refers to the blurry input. The Wiener filter allows for spectral attenuation of noisy frequency components while deblurring.

Richardson and Lucy (RL) [20, 16] describe a maximum likelihood estimation principle that is based on the Poisson distribution yielding an iterative computation rule [3]. We also choose RL for our testbed due to its successful application throughout research and perseverance although being proposed 40 years ago.

Similar to the Wiener filter and understood as the deblurring extension of the truncated SVD, spectral attenuation of singular components is also achieved by the Tikhonov regularization [26].

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

Levin et al. propose a fundamental method using natural image priors [13, 14]. The natural image prior is represented by $\rho(z)=|z|^{\alpha}$, which is either assumed to be Gaussian ($\alpha=2$) or sparse ($\alpha=0.8$). As the Gaussian prior does not necessarily favor the image edges over the noise, the sparse prior yields to concatenations of edges rather then emphasizing noise influences.

In general, recent methods tend to use Gaussian likelihoods and Laplacian priors on the gradients [1, 28].

Krishnan et al. generalize Levin's energy employing a hyper laplacian prior, where $0.5 \leq \alpha \leq 1$ [11]. The hyper laplacian prior fits to emperically estimated gradient distributions of natural images. We include and adopt their method, named Hyper-Laplacian-Prior (HLP), in our framework as one of the state-of-the-art bases due to its popularity and reported robustness as well as its applicability to local deblurring approaches [2, 8].

Fortunato et al. introduce a distinct representation of the image gradients by retrieving and deploying edges from a bilaterally filtered representation of the blurred observation to the optimization energy [5]. The final result for the vectorized latent sharp image l is then obtained solving a set of linear equations $Ul=v$, where U and v are given in Eq. 3. We also include this method, called Shape-Adaptive-Prior (SAP), in our framework as a representative of the most recent approaches.

Considering local deblurring requires the blur operator K to be fed with local blur kernels, which differ in each row. Hence, the system loses linear shift-invariant properties and thus computational speed by deploying Fourier methods. Local deconvolution is widely studied for motion blur rather than blur from defocus. The majority of methods in this field utilize the Richardson and Lucy energy. Methods by Whyte et al. and others differ mainly in estimating and applying the local functions into the framework [29, 32, 25, 24]. Further computational optimization can be found in Nagy et al. [18].

The method by Krishnan et al. was adopted by Couzine-Devy by employing a local convolution matrix [2]. In contrast, Sun et al. locally adopt priors utilizing the likelihood between a database of sharp images and deblurred image patches [22].

Blind Deblurring

By definition, blind deblurring comprises the alternating estimation of the sharp image and the point-spread function [9].

Krishnan et al. also present a method for blind deblurring based on their non-blind approach above. The blind deblurring method is based on a multiscale approach, where the PSF is estimated in the first scale level [12]. Having found the best matching operator K, it can be applied in the non-blind deblurring method to find the latent image L. If the PSF would be known a priori, computational costs were saved and erroneous PSF estimations prevented.

Additional Priors

Considering motion deblurring, Tai et al. provide a system of cameras which run at a low and high frame rate to find a solution for reducing motion blur [23].

Similarly, Li et al. make use of a tri-focal camera system, including the computation of depth maps for motion deblurring [15]. A method that makes explicit use of a depth map is proposed by Xu et al. targeting motion deblurring [31]. Therein, the depth map is obtained from a stereoscopic camera setup and the PSFs are estimated for all different depth layers. Yue et al. employ depth information to estimate the motion of the camera itself to support the PSF estimation step [33].

Obviously, current research strongly focuses on motion deblurring. Especially focus deblurring using a depth map has not yet been touched directly. Closely related is an approach of Navarro et al. where the depth-of-field in integral imaging systems is enlarged by using the inherent depth map [19]. Therein, the image is segmented in accordance to the depth map and a specialized RL approach is used for deblurring all single layers. Finally, the resulting images are blended to obtain a larger depth-of-field area.

Luft et al. apply depth-dependent stretching of local contrast at depth discontinuities [17]. Next to computing a high pass component of the depth map, they modulate color, luminance and local contrast of the input color image, yielding perceptually increased depth separations of the image objects.

Depth-Guided Deblurring Framework

Our methodical contribution enables the four selected deblurring methods to accept the required additional priors and to allow for different complexity levels of spatio-axial operation.

Methodology

All methods are provided with the captured and fully or partly blurred color image that represents the observation, available in the photo-linear domain B_{lin} and in the gamma-corrected monitor B_{mon} domain, and the corresponding depth map, provided as the captured TOF depth map D_{real} and as a ground truth depth map D_{gt}.

Furthermore, the blur matrix K^i for discrete distances, containing the according PSFs, is supplied. Besides the measured PSFs, there is also a Gaussian kernel provided in order to evaluate the benefit of using a measured ones. Next to conventional shift-invariant global matrices K^i, local matrices are employed in this work. Local is defined such that the row entries of K^i are selected depth-dependently.

Each method is also extended to handle a local regularization weight matrix $W=W(x,y)$. As the standard scalar weight controls the smoothing or sharpening of the entire image, the matrix W now allows for local control, which is the individual weighting of single pixels.

USM can accordingly be formulated as

$$S=B+W\circ\Gamma(B-(B*K^{i\in\{g,l\}})), \tag{1}$$

where $\circ$ denotes the Hadamard product, $*$ is the convolution operator and Γ is a thresholding operator yielding a sparse high frequency image. The latter is generated by subtracting a blurred representation from the blurry input image B. This subsequent blurring process is achieved using either the "global" shift-invariant K^g or "local" shift-variant blur Matrix K^l.

RL is handled slightly different. The smoothing or sharpening parameter is controlled by different amounts of iterations. Applying local weights does then mean that the image has to be stitched after different amounts of iterations for the according sub images.

HLP is modified such that we can integrate our weighting matrix W and convolution matrix K^i. Therefore, Eq. 4 in the paper of Krishnan et al. [11] can be re-written as

$$x= \mathscr{F}^{-1}\left(\frac{\mathscr{F}(F^1)^*\mathscr{F}(w^1)+\mathscr{F}(F^2)^*\mathscr{F}(w^2)+\frac{W}{\beta}\mathscr{F}(K^i)^*\mathscr{F}(B)}{\mathscr{F}(F^1)^*\mathscr{F}(F^1)+\mathscr{F}(F^2)^*\mathscr{F}(F^2)+\frac{W}{\beta}\mathscr{F}(K^i)^*\mathscr{F}(K^i)}\right), \tag{2}$$

where the gradient filters F^n, auxiliary variables w^n and the weight β stay unchanged. The last summand of the nominator and denominator

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

turn out to be dependent of the convolution matrix K^i and the blurred observation. In order to adapt it towards local changes, we transfer both summands to the time domain. Therefore, the part $\mathcal{F}(K^l)^*\mathcal{F}(B)$ turns into $\mathcal{F}(K^l * b)$. Computing the correct denominator term $\mathcal{F}(K^l)^*\mathcal{F}(K^l)$ introduces an enormous complexity when the entries of K^l are locally adaptive. Due to this computational demand, we approximate the result retaining the given computation in the frequency domain using K^g.

The adaptation of SAP requires a similar adoption like for HLP, yielding a modified set of linear equations

$$U = K^{gT}K^g + \sum_{s=1}^{5} \lambda_s d_s^T d_s, \qquad v = K^l b + \sum_{s=1}^{5} \lambda_s d_s^T w_s. \tag{3}$$

The blur matrix is replaced by its local counterpart K^l. Equally to the adaptation of HLP adaptation, the term $K^{gT}K^g$ in the first equation is computed with one, shift-invariant blur kernel, due to the computational complexity.

As this is a non-complete adaptation to local behavior in general, SAP and HLP are applied in a depth-guided sliced way with one single constant PSF for each slice. Furthermore, SAP incorporates two σ parameters adjusting the bilateral filter. The parameters were visually optimized for the real camera data and are set to $\sigma_{spat} = 10$ and $\sigma_{color} = 0.0004$.

PSF Prior

The proposed PSF measurement process is described in this section. Figure 1 shows a coarse illustration of the measurement setup. The general idea is to capture a light point, assumed to be a point source, which moves from pre-focus to post-focus while the lens settings are fixed. The employed lens in this paper is a *Zeiss Master Prime* 50mm, which was set to an aperture of $T = 1.3$ and a focus of 2m, which is an often used setup to achieve shallow depth of field.

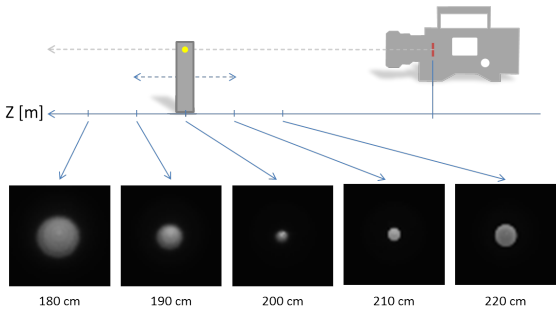


Figure 1: Experimental overview of our PSF measurement. The point light carrier is moved around the focal distance to stimulate the corresponding PSFs.

All deblurring methods in this work are applied to real images captured with an *ARRI Alexa* digital film camera mounted with the aforementioned lens. Therefore, the capturing device of the PSFs is chosen as the same camera model and lens to keep sensor and optical responses constant within the whole chain [7].

To acquire the PSFs for different depth levels, the measurement device, which is a *Trioptics ImagesMaster*, is equipped with a point light of circular shape with a diameter of 30µm. The carrier is then moved from pre-focus to post-focus in distance steps of 5cm. Figure 1 also depicts some of the resulting PSF measurements exemplarily. It shows the change of diameters and intensity structures with altered distance to the focus plane. Additionally the asymmetrical depth behavior of the lens is prominent.

All recording was done with the native sensor resolution in the camera's raw format to allow for utilization of the high spatial resolution and usage the linear data additionally. To reduce image noise, 25 frames were captured each. The subsequent temporal mean operation restores an estimate of the signal. Furthermore, an automatic, depth dependent control of the exposure time was implemented to keep the exposure constant and to retain the full dynamic range of the PSF.

In order to be able to evaluate the benefit of the measured PSFs, a synthetic Gaussian kernel with depth-varying sizes is adduced in this work. The Gaussian PSF is motivated by its frequent usage in the community. However, the parameters of the Gaussian kernel are oriented on the real PSFs. Heuristically, a good starting value was found to be $\sigma = 2.5 \cdot d_{real}(z)$ with $d_{real}(z)$ being the diameter of the real PSFs at depth z.

Depth Models

The methodical approach in this paper is structured as follows. First, we explain how depth and PSF priors are effectively employed in a pre-learned deblurring framework for deblurring of single depth layers. Second, the approach is extended for an arbitrary number of depth layers.

Single-Layer Deblurring

The two main unknowns for deblurring are the regularization weight and the PSF. Naively applying the measured PSF referring to its measured location z might not lead to convincing results due to aberrations and the introduced measurement quantization. Hence, z is varied to query the best matching PSF within a certain neighborhood. Let $\hat{z}$ describe the optimized distance, this can be formulated as $PSF(\hat{z}) = PSF(z + \hat{o}(z))$. In this paper, the hat denotes all parameters optimized by this principle.

That means, for each input depth value z, a depth-dependent offset is selected. Therefor, a training procedure is required. Let D be a set of neighboring distances for a given focus set. The training phase maximizes the function ϕ, representing the relevant metric, which is adduced in the respective evaluation. All metrics accept the deblurring result L_{SLD} and the reference image R. This step is repeated for each depth plane $z \in D$ individually, denoted as

$$(\hat{o}(z), \hat{w}(z)) = \underset{o(z),w(z)}{\operatorname{argmax}}\ \phi(L_{SLD}, R) \qquad \forall z \in D. \tag{4}$$

After yielding $\hat{o}(z)$ and $\hat{w}(z)$ from this training on multiple planar (multi-planar) targets of the test data set, they can be applied to the selected deblurring methods as the distance of the object of interest in known by the captured depth map.

Still, deblurring methods are not capable of sharpening strongly blurred regions. Naively applying deblurring methods on the whole image will introduce severe artifacts. In order to overcome this problem, the Single Layer Deblurring (SLD) approach is introduced. The basic idea is to split the image into two regions based on the depth map, which is a secondary deployment of the latter. The first selection describes all areas where the deblurring result is used for the resulting image. The second selection describes all areas where the original image is used in the final image. Thereby the original background or foreground blur is effectively kept in the final image free of artifacts. To do so a validity matrix $V = V(x,y)$ is introduced. The matrix has

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

the same size containing only the values 0 and 1. Therein, a metrical range z_{range} can be defined to compute the validity map by

$$V(x,y) = \begin{cases} V(x,y) = 1, & D(x,y) - z_{range} < D(x,y) < D(x,y) + z_{range} \\ V(x,y) = 0, & \text{otherwise.} \end{cases} \tag{5}$$

Subsequently, the deblurred image L is computed by

$$L = V \circ L_{SLD} + (\mathbf{1}\mathbf{1}^T - V) \circ B. \tag{6}$$

L_{SLD} corresponds to a deblurrred image which results when deblurring the image applying using either naive or optimized PSFs and optimized weights. This approach allows for the explicit control of depth ranges in which the deblurring methods are applied. Occuring stitching artifacts can finally be overcome by a small smoothing kernel on V in order to achieve soft blended edges.

Multi-Layer Deblurring

Although SLD is very promising for scenarios where a single depth layer needs to be reconstructed, like in the eye and ear example from the introduction, we also propose a generalization of the SLD to Multi-Layer Deblurring (MLD). In particular, we want to stress to cases, which are designed for the comprehensive evaluation. First, we use only optimized weights and PSFs as measured and second, we also include optimized PSFs. This procedure allows for the specific measurement of the effectiveness of both optimization options.

In order to allow for per-pixel weights the matrix $W = W(x,y)$ is introduced. Again, depth-dependent weights are selected from the training phase. The procedure is applied for each depth slice to show the benefit when a local weight selection is performed. Additionally a depth-dependent extension $V_z(x,y)$ of the validity operator V is introduced. It will be denoted by where the z indicates the distinct depth layer. Thereby, the image can be reconstructed by computing the sum

$$L_{MLD} = \sum_z V_z \circ L_{SLD}(\hat{W}, PSF(z)), \tag{7}$$

where L_{SLD} again corresponds to one SLD result, however, neglecting optimized PSFs.

The final methodological step adds the selection of the optimized point-spread function. Besides creating local weights via $W(x,y)$, $hato(z)$ is evaluated for each image coordinate (x,y,z), where z is obtained from the depth map. This enables the optimized selection of blur kernels $PSF(\hat{z})$. The problem is similarly expressed as

$$L_{MLD} = \sum_z V_z \circ L_{SLD}(\hat{W}, PSF(\hat{z})). \tag{8}$$

Evaluation

There is no publicly available data set, which provides high resolution images, captured depth maps and a corresponding PSF archive. Thus, we generated a versatile data-set with different image content, amount of blur, depth variation and ground truth. After the data-set is explained in the following section, suitable metrics are presented. In the last part, the underlying questions are evaluated by subdividing the problem. First, the selected methods are applied on scenes which do not contain depth variation. This comprises investigations on the PSF type, the best photo-metric domain and color representation for deblurring. Next, the methods' parameters are trained on different depth levels. Finally, this information is utilized in the proposed approach for images which contain depth variations.

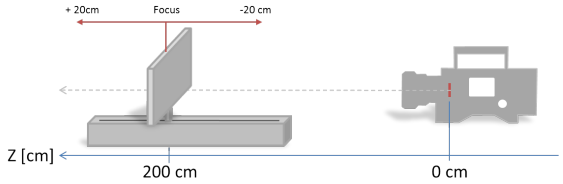


Figure 2: Measurement setup for the data set type plane. The focus distance of $z_{focus} = 2$m is fixed for all shots. The target with different textures slides in the range of the focus position ± 20 cm.

Data Set Generation

The data set provides the ability to obtain numerical and visual evaluation as well as adequate content to demonstrate the benefit of deblurring. The following list contains the core specification:

- **Realistic and depth varying blur**
  Whereas the majority of scientific contributions works on synthetically blurred data, we work on real camera data. This implies realistic blur, which is generated using a large physical aperture.
- **Sharp reference for visual and numerical comparison**
  Our data set allows for numerical evaluation, parameter learning and visual assessment. In particular, the human observer is given a reference to estimate the image sharpness after deblurring.
- **Corresponding depth map**
  To evaluate the given assumption, it is necessary to capture a dense depth map. In order to asses the quality of the captured depth map for deblurring, a ground truth depth map needs to be provided for this part of the data.
- **Variety of image content and depth variation**
  We evaluate the utilization of learned parameters. Under these requirements, the data set should exhibit a great amount of variety. Additionally, it should contain different types of depth structures. Thus, multi-planar and depth-varying scenes are required.
- **Different color domains for processing**
  Data in the *monitor* domain (ITU Rec. 709) has been passed through numerous processing steps. Since blurring has a physical origin, we provide the same content in the *monitor* and in the *linear* domain (sensor linear).

The deployed camera is a modified version of a RGB+Z camera by Hach et al. [6]. It allows capturing of spatio-temporal synchronized RGB and dense depth data without occlusions. The data is recorded in the camera's raw format allowing to transform the image into the domains *linear* and *monitor*.

To create a blurred and a sharp image, we employ two different lens apertures and a moving slider shifting the object of interest in and out of the depth-of-field area. For the used lens, the largest aperture setting is $T = 1.3$. Given a focus setting of 2 m, the pre-focus point situates at $f_{pre}^{1.3} = 1.98$ m and the post-focus distance measures $f_{post}^{1.3} = 2.02$m. In contrast, for the smallest aperture setting of $T = 16$, both distances are $f_{pre}^{16} = 1.75$ m and $f_{post}^{16} = 2.31$ m, respectively, assuming a circle-of-confusion of 13 µm. Between the range of f_{pre} and f_{post} the image content is depicted at the system's maximum sharpness and increasingly blurred elsewhere.

Blur variation is created by shifting the object of interest in depth. Starting from 20 cm in front up to 20 cm behind the focus, the object

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

was moved by a precise and automated slider with a stepping of 2 cm while having the aperture set to $T = 1.3$ and $T = 16$, respectively. Additionally, the exposure is controlled to account for the attenuation by the aperture. The overall setup can be seen in Figure 2.

Figure 4 shows an overview of the entire data set. For the planar test charts, there exists also a reference value for the depth map. Images that are referred to as non-planar, contain depth variations. It means that besides the object of interest, there are foreground or background objects. This setup is intended to evaluate the degradation caused by deblurring methods in intact image regions or, to the contrary, the preservation effectiveness of selective deblurring methods.

The last group of images contains a synthetic depth reference. Whereas the *planar* images are assigned a reference depth by measuring the unique position of the target, these images contain in-scene depth variations. Hence, the ground truth was drawn manually using the precisely measured geometry of the scene. This group of images can be seen in Figure 4i and Figure 4j. For the remainder of the evaluation all planar multi-planar images are referred to as **D1** and all depth-varying images are referred to as **D2**.

Metrics

For numerical evaluation, we chose 2 often used metrics from literature and define a third metric, which provided visually convincing results in the realm of deblurring.

The PSNR Human Visual System (PSNRHVS), proposed by Egiazarian et al. is based on the usual PSNR. To optimize the metric for human observers, they suppose that the HVS is more sensitive to low-frequency changes instead of changes in the high frequencies. They feed this weighting into the computation of the PSNR [4].

The Visual Image Fidelity (VIS) by Sheikh et al. investigates on statistical properties of natural images [21].

Our third metric, which we call FCORR is computed by measuring the cross-correlation of the spectra between the reference image and the deblurred image. Before converting to the frequency domain, the mean is subtracted to focus on the structural elements and to ignore small intensity offsets in the data set. The correlation is measured per color channel and subsequently, the mean cross-correlation value of all three color channels is computed.

Visual and Numerical Results

An overview of the evaluation can be obtained from Figure 3. Therein, the evaluation process is depicted in accordance to the raised questions. The letters on the right of the image identify the part problems and correspond to:

(A) Evaluation of the optimal domain type, color type (RGB or Luminance) and blur kernel type. This evaluation is done on the data set **D1**.

(B) After the basic parameters have been evaluated a detailed, depth-dependent optimization of the blur kernel offset and weights is performed. Similar to part A, the multi-planar test data **D1** is used.

(C) Having trained the parameters, the SLD and MLD approaches are evaluated using the data set **D2**.

Experiment A: Blur Kernel, Domain and Color

The blurred image can either be provided in *linear* (d_l) or *monitor* (d_m) domain. Furthermore, it is possible to work either on

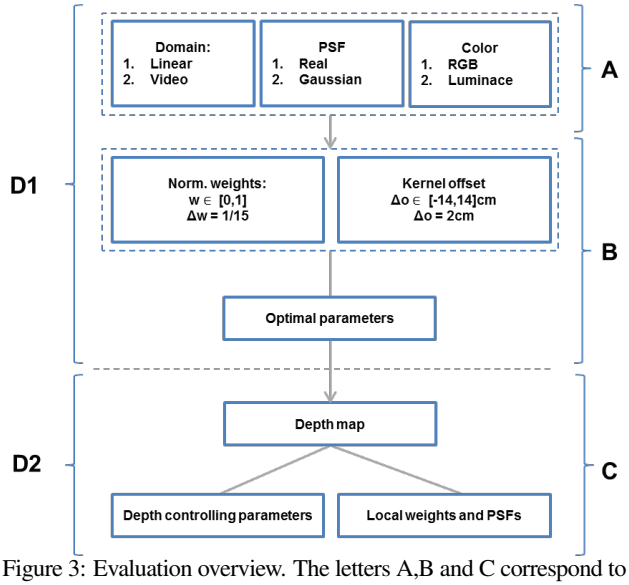


Figure 3: Evaluation overview. The letters A,B and C correspond to different evaluation parts. The identification D1 and D2 correspond to subsets of our data set. D1 represent multi-planar images and D2 contains images with depth variation.

all RGB channels separately (c_r) or on the luminance channel only (c_l). The last and most important parameter is the usage of the PSF type. It is either possible to work with the real measured PSFs (p_r) or on synthetic Gaussian kernels (p_s).

This investigation describes part **A** from Figure 3. In there the three parameters d, c and p are varied. For the evaluation, the blur kernel offset $\hat{o}(z)$ and the weight $\hat{w}(z)$ were optimized for each input combination. This experiment was done on a subset of different depth layers. For the four test shots of **D1**, the distances 190 cm and 210 cm were selected. To get an overview of the test results, Figure 9 gives the numerical results for all four selected methods while altering the inputs.

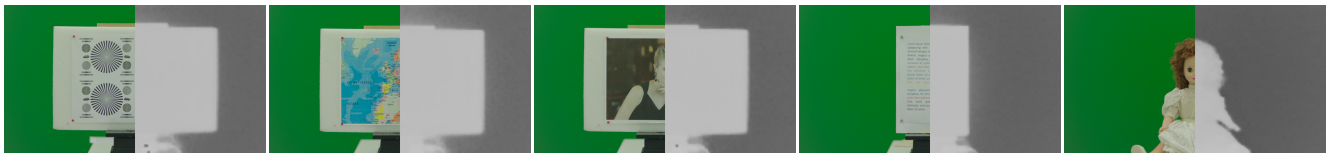
The question arises which of the varied inputs should be used and which general trend is observable. Table 1 shows a summary of the previously obtained numerical results. The values were computed, by fixing the parameter of interest and computing the mean of the remaining parameters. For example, when computing numerical differences for d_l and d_m, each of the remaining variations c_r, c_l, p_r and p_s were computed and subsequently the average value of the metric is picked.

The x marker in the table denotes that the corresponding input performs better. Better means, that the majority of metrics voted for the input to be the best for the selected method. Majority means that at least 2 out of 3 results evaluated the input to be better. A star on the corresponding x indicates that 3 of 3 metrics voted equally.

The usage of either a real or synthetic PSF shows the biggest disagreement and varies for each method. However, an interesting observation can be done. It shows that the two older methods USM and RL benefit from a real kernel whereas the more recent methods HLP and SAP tend to benefit from the usage of synthetic kernels. That means, approaches which are based on a prior minimization scheme work better with synthetic kernels. A strong agreement can be derived for the question on working on each color channel separately or on luminance. All methods work better for the luminance channel. The selection of linear or monitor domain varies strongly in dependence of the method.

Considering the measurement of the real PSFs, the overall process requires a noticeable amount of efforts. This begins with that

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168



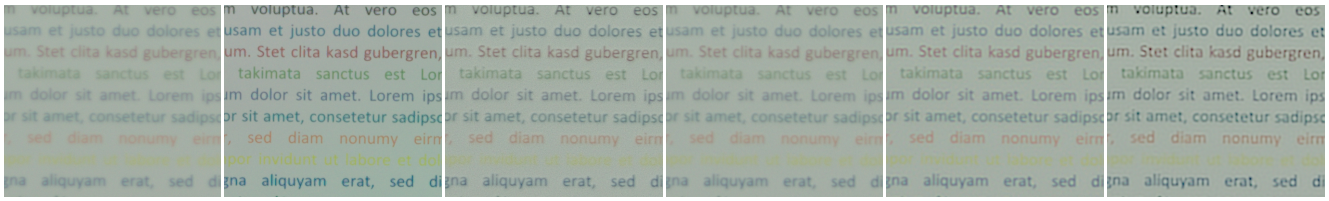
(a) *SiemensPlane* (b) *WorldPlane* (c) *WomenPlane* (d) *TextPlane* (e) *Doll*

(f) *Doll* (g) *DollForeground* (h) *TextScene* (i) *DofFan* (j) *DofLyingPlane*

Figure 4: Overview of the data set with corresponding depth maps

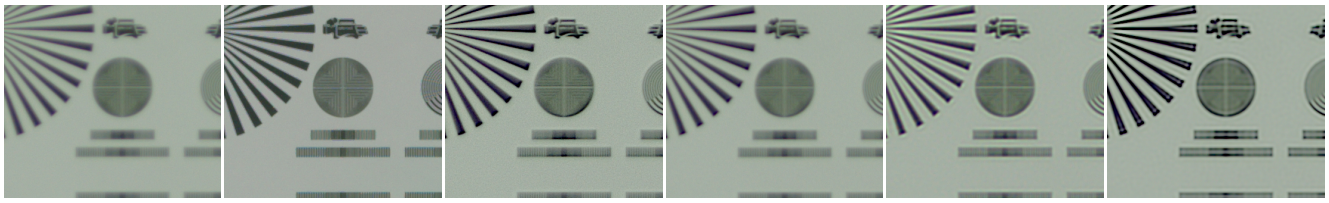
(a) Blurred input (b) Reference (c) USM (d) RL (e) HLP (f) SAP

Figure 5: Application of the trained parameters for the test shot *Doll* at a distance of 190 cm (best viewed in electronic version. Zoom for more details)



(a) Blurred input (b) Reference (c) USM (d) RL (e) HLP (f) SAP

Figure 6: SLD using trained parameters for the test shot *TextPlane* at a distance of 190 cm (best viewed in electronic version. Zoom for more details)



(a) Blurred input (b) Reference (c) USM (d) RL (e) HLP (f) SAP

Figure 7: SLD using the trained parameters for the test shot *SiemensPlane* at a distance of 190 cm (best viewed in electronic version. Zoom for more details)

(a) Blurred (b) Reference input (c) BHLP (d) USM (e) HLP (f) SAP

Figure 8: Visual comparison of non-blind and blind deblurring approaches for the shot 'WomanPlane' at 190cm (best viewed in electronic version. Zoom for more details)

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

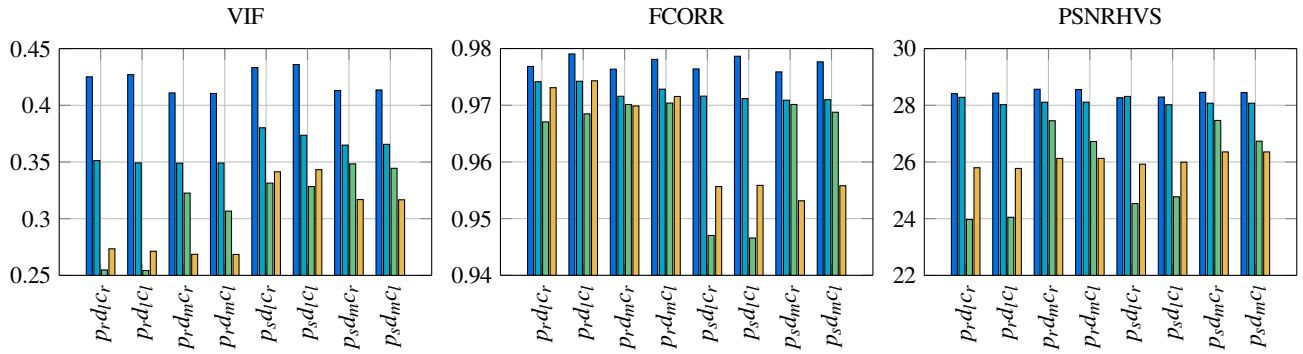


Figure 9: The three metrics VIF, FCORR and PSNRHVS for varying input combinations of the domain type (linear d_l or monitor d_m), PSF type (real p_r and synthetic Gaussian p_s) and color model (RGB c_r or on luminance only c_l). The values are averaged for the distances 190 cm and 210 cm for all four planar scenes. The colorization in the plot corresponds to the following methods:
— USM — RL — HLP — SAP

Table 1: Majority votes for the different input types p, c and d. A x indicates that 2 out of three metrics voted for the distinct input type to be the majority vote. A star indicates that all three metrics agree on the same result. The values are computed by fixing one parameter and averaging the result for the remaining parameters. The distances 190 cm and 210 cm were used for all four planar test shots of the data set D1.

Type	USM	RL	HLP	SAP
PSF - real - p_r	x	x	-	
PSF - syn - p_s	-	-	x	x
Color - rgb - c_r	-	-	-	
Color - lum - c_l	x	x	x	x
Domain - lin - d_l	x	x*	-	x
Domain - mon - d_m	-	-	x*	

fact that PSFs behave differently for all physical lens setting and the measurement has to be done with special hardware and appropriate illumination circumstances. When weighing this fact towards the benefit in the presented results, the measurement efforts hardly pay off for very high quality applications. However, in our experiment, the synthetic PSFs were created by deriving their sizes from the sizes of the measured PSFs. Thus, it is of great benefit, if at least the PSFs' sizes are known. As it is shown in the following parts of the evaluation, the knowledge of the PSF structure does not deliver convincing improvement. All the more, the PSFs' depth-dependent sizes are of promising benefit.

Experiment B: Optimizing Depth-Dependent Parameters

In this experiment, the selected methods are applied to the four planar test shots on each of the 21 depth layers. The evaluation shows the benefit of the depth-guided parameter selection in terms of the best PSF and the best weight per method. The first task is optimizing the weights and PSF selection for each depth plane for each test image.

In order to point out the parameters' depth dependency and sensitivity, the four implementations noted in Figure 10 were created.

The chosen metric is the FCORR metric. The results can be obtained in Figure 10. The horizontal axis shows the distance sweep from -20 cm to 20 cm around the focus point at $z_{foc} = 2$m in steps of $\Delta z = 2$cm. The measurements itself was computed by sweeping both parameters $o(z)$ and $w(z)$ creating a 2D feature space containing the numerical results for each depth level. The range for variation of $o(z)$ was limited to -14 cm and to 14 cm with respect to the initial depth position. The step width was $\Delta z = 2$cm. The distances which are used in this part of the evaluation originate from precisely measuring the ground truth distance. Since the test was fulfilled on the four planar scenes individually, the resulting best parameter coordinates are averaged and then considered as learned.

The green lines show, that adapting the PSF accordingly already results in a great benefit. But, if the PSF is not adapted and remains as measured but the weight is optimized, the benefit is also remarkable. Hence both parameters strongly depend on depth and should be adapted accordingly.

Figure 11 shows a comparison of the best results of each selected method and additionally the metrics values, if no deblurring is used. Again the distances reach from -20 cm to 20 cm cm with respect to the focus at 2 m. Starting in the focus plane and slowly increasing the distance, the first distinction can be done between the methods. The two methods USM and RL perform best near the focus distance since the can be parameterized in such a way that they don't have any significant influence. This does not hold for the two methods HLP and SAP. Both methods are based on a minimization scheme. The methods don't behave well near the focus plane since the weight cannot be set in a way that the image is kept untouched. An improvement in comparison to the 'no deblurring' reference can be observed at -3cm and $+10$cm around the focus point. Concerning this observation, all methods nearly behave the same. When traveling even further away from the focus plane, especially towards the negative distance, the method RL is slightly ahead of the competition. Considering the other methods, the is no clear ranking as they behave similarly with respect to FCORR. Note, in the depth-of-field zone (DOF), no method can be better than no deblurring.

The next step of the evaluation contains learning of depth-dependent parameters. The purpose of evaluating each depth layer builds the foundation to find the best parameters for depth offsets and correct weights. This is beneficial for three reasons. First, if the depth value of a planar scene is known, it is possible to directly apply the parameters instead of the sweeping for the best. Second, if the scene contains in scene depth variation the parameters can be applied locally. Third, the typical lack of a sharp reference, which the methods are

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

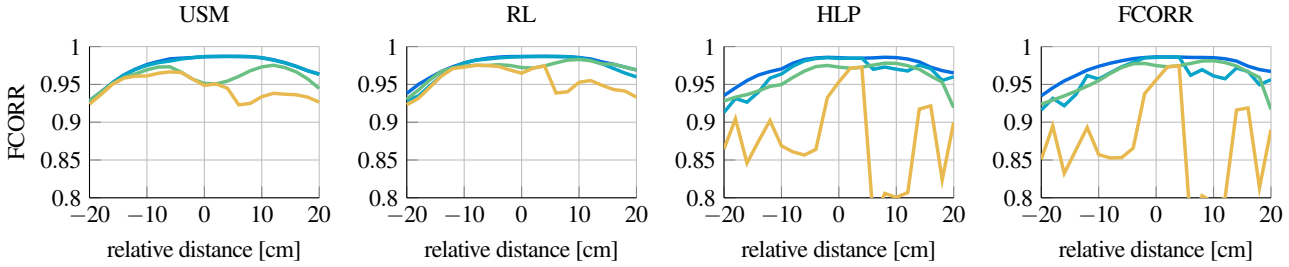


Figure 10: Mean FCORR results for the four different adaptations of parameters to asses the influence of weight and PSF optimization. Shown are all four methods including a depth sweep over all planar scenes. The four adaptions are: — $\mathbf{PSF}(\mathbf{z})$ and $\mathbf{w}(\mathbf{z})$, PSF as measured and fixed weight — $\mathbf{PSF}(\hat{\mathbf{z}})$ and $\mathbf{w}(\mathbf{z})$, PSF offset corrected and fixed weight — $\mathbf{PSF}(\mathbf{z})$ and $\mathbf{w}(\hat{\mathbf{z}})$, PSF as measured and optimal weight — $\mathbf{PSF}(\hat{\mathbf{z}})$ and $\mathbf{w}(\hat{\mathbf{z}})$, PSF offset corrected and optimal weight

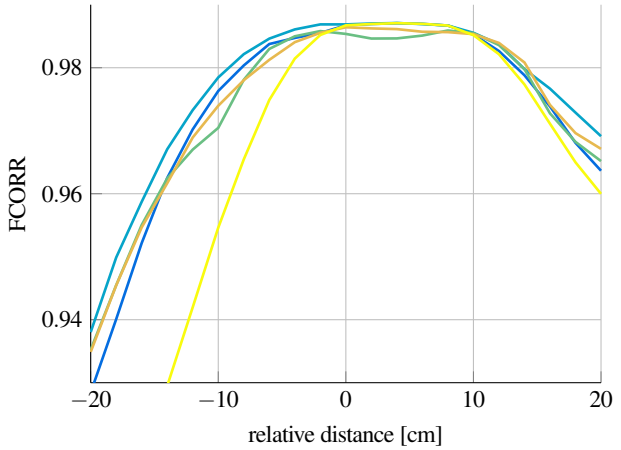


Figure 11: Shows all four methods with optimal parameters according to FCORR per depth plane. It is also shown how the metrics react, if no deblurring is performs. Close to the focus distance no deblurring seems to be performing best. But at about −3cm and +10cm around focus, the deblurring methods show an improvement. At the outer positions the recent approaches HLP and SAP perform best.
— USM — RL — HLP — SAP — No deblurring

optimized to, is overcome. The application of the trained parameters for each method can be seen in the Figures 5, 6 and 7.

In there, the resulting images of the depth value 190 cm are depicted. For the trained parameters the method USM delivers the visually most convincing results although it introduces an amplification of noise. The approach RL also delivers convincing results whereas the details are not re-covered entirely. In general, the RL approach turns out to be more hesitant in terms of sharpening the image. The HLP results are comparable to the USM results. HLP, however, still introduces noticeable artifacts on intensity and depth discontinuities. This effect gets even worse for the method SAP.

Comparison to Blind Deblurring

The presented methods originate from the field of non-blind deblurring. This means that they are provided with a PSF matrix in order to perform the deblurring procedure. Our numerical comparison can be obtained from Figure 12. Therein, all four planar shots at a distance of 190 cm are computed with the blind method Blind Hyperlaplacian Prior (BHLP), in comparison to the non-blind methods with the trained parameters. The weight parameter λ of the method BHLP was sweeped in order to get the best results. The initial blur kernel was selected to be a 2×2 pixel area filled with ones in a black patch of size 25×25. While sweeping the regularization weights, the standard values, which can be found in the paper of Krishnan et al. were used for all other parameters [12].

Taking a look at the numerical results, BHLP does not outperform the non-blind methods in general. However, in a few cases, BHLP outperforms the non-blind methods. In order to provide a better impression, a visual comparison is also given in Figure 8. The results show that BHLP introduces a great amount of sharpness but also provokes artifacts, especially at edges. The non-blind methods act more preserving. Furthermore, it has to be kept in mind that blind deblurring methods require an enormous amount of time due the alternating minimization scheme optimizing the image and the PSFs. The method BHLP took 64 times longer than the non-blind version of HLP.

Experiment C: Depth Models

This section focuses on depth variations. Thus, the proposed SLD and MLD approaches are evaluated.

Single-Layer Deblurring

Starting with an comparison between the ground truth and the measured depth map, Figure 13 shows the FCORR results, when the trained parameters are used in SLD. The selection of parameters via the measured depth map yields results, which are comparable to using the ground truth depth. Although this observation is slightly dependent on the used deblurring method, there is no evidence that shows mediocre results when using the real depth map. On the numerical basis of this evaluation, we state that the depth precision, delivered by the upscaled PMD 19kS3 depth map, is sufficient for selecting PSFs and weights from a pre-learned library.

Figure 14 shows an application of the SLD procedure. Figure 14a shows the originally blurred image. The object of interest, the doll, was located 14 cm in front of the focus point. Thus, the doll is slightly blurred. The background objects also show a strong amount of blur as they are far out of focus. Globally applying the deblurring method SAP introduces a sharpened doll, but also the background gets strongly degraded since the single weight and PSF are not adequate for deblurring the strongly blurred background regions as shown in Figure 14b. Using SLD instead is depicted in Figure 14c. The doll clearly gains sharpness whereas the background is left untouched. The mask was created using the real processed depth map, which was upscaled using the popular joint-bilateral filter [10]. Then, the face region was selected in order to derive the depth position of the face

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

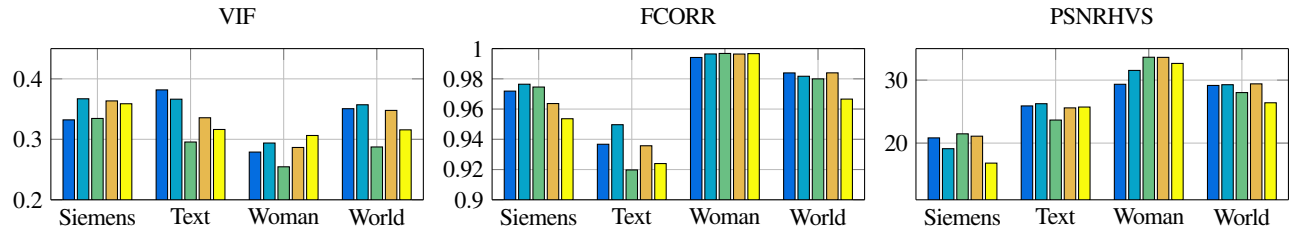


Figure 12: Comparison of non-blind and blind deblurring methods. Results are shown for the depth plane at 190 cm for all planar test images. Blind deblurring using only the sizes of the measured PSFs is comparable to the results of the non-blind methods using the full PSF information.

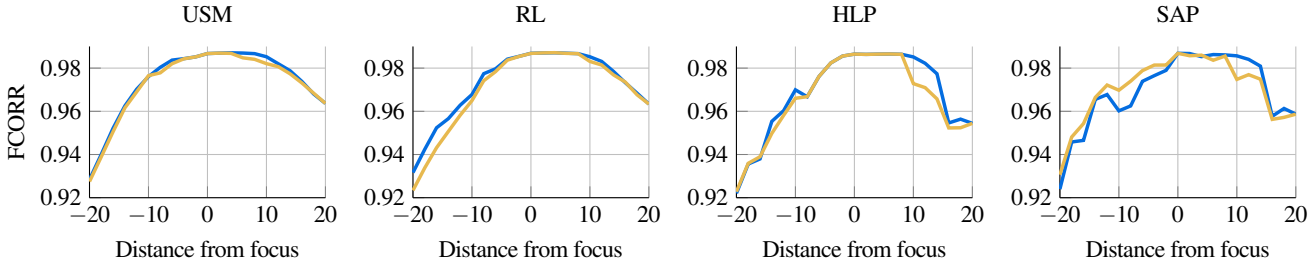


Figure 13: FCORR results for the selection of the pre-trained parameter by either the ground truth distance information or the depth map.

by the depth map. Using $z_{range} = 10$cm, a depth interval was selected which guided the process. The used $PSF(\hat{z})$ and $w(\hat{z})$ were gathered from the parameter optimization for the metric FCORR.

Considering the quality of the depth map, it can be seen that the selection in the hair region of the doll is tricky, when inspecting Figure 14c. However in this region, the depth map measured more of the background since the infra-red light is partly transmitted by the hair resulting in miss-measurement. Regarding the black tape on text the plane, it can be observed that the blurred parts are used in the stitched image instead of the deblurred counterparts. This can be traced back to depth measurement artifacts.

Concluding, the depth-selective approach SLD delivers yet convincing results. In the introductory example, it is not unusual to utilize a small DOF. This means that the object of interest is intended to be sharp while foreground and background are blurred in order to direct the viewer's attention of the portrait. If the focus is only slightly wrong or the actor stepped slightly off his position, the blurred parts do not change significantly while the face becomes inconveniently blurred. Deblurring the object of interest using SLD, delivers a great benefit if the depth map is accurate. Problems occur at corrupted depth measurements, which require a correction prior to deblurring.

Multi-Layer Deblurring

To add more depth layers for deblurring, the method MLD was introduced. Therein, first approach ($PSF(z)+W(\hat{z})$) simply adapts the optimized weight in dependence on depth and the second ($PSF(\hat{z})+W(\hat{z})$) additionally optimizes the PSF selection. Thereby, the effect of using the optimized PSF selection can be evaluated. The results, utilizing the selected method SAP and the real depth map, are shown in Figure 15. Whereas deblurring shows convincing results in this setup, it is still problematic, that the usage of deblurring methods might sharpen the image but manipulates it in a way that it can be visually distinguished from the unprocessed or only slightly processed regions.

Conclusion

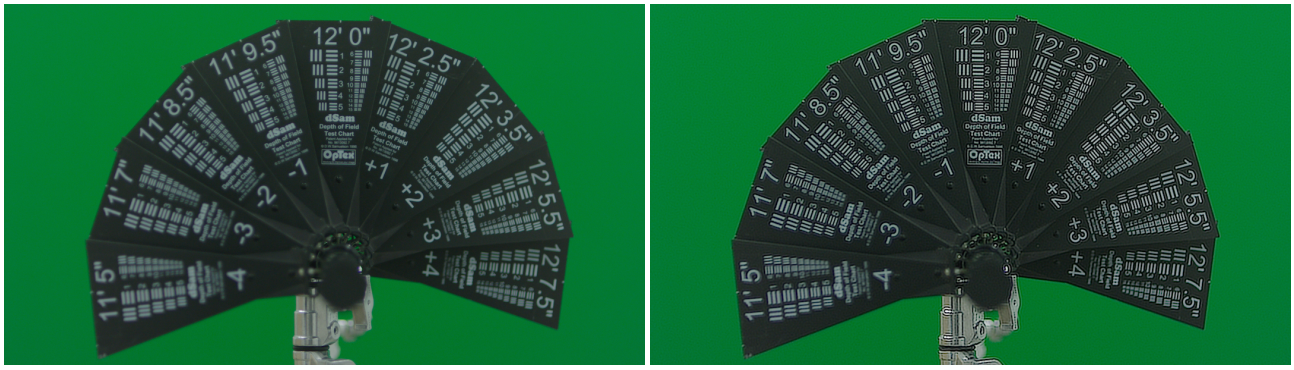
In this work, we discussed a comprehensive defocus deblurring framework, which was ignited by the availability of a color image accompanied by a dense depth map. The methodological part addresses the enhancement of four state-of-the-art deblurring methods. Therefore, we provided single-layer and multi-layer deblurring strategies to include depth information, measured lens point-spread functions and depth-selective weighting. This includes a learning strategy to apply optimized parameters to general reference-free data as well as an a priori captured point-spread-function archive to overcome scene unknowns.

We finally conclude that using the learned parameters and depth-selective processing yields superior deblurring results while strongly suppressing artifacts. This is confirmed by numerical and visual evaluation on our captured test data set. Thus, we argue that dense depth maps are already highly beneficial towards depth-guided deblurring approaches, although they still noticeably lack in quality.

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

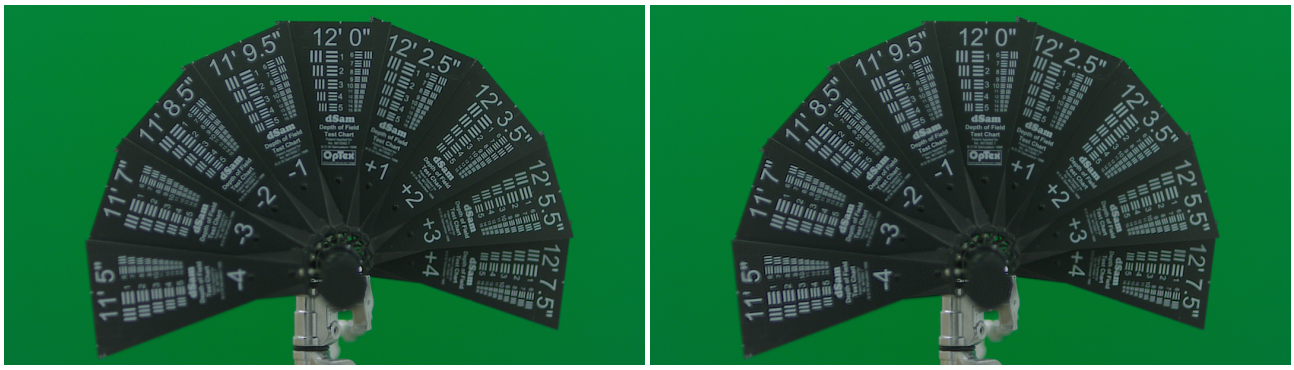
(a) The blurred input image (b) Applying global deblurring method (c) Deblurring only applied in the region of interest

Figure 14: Depth selective image deblurring applied on the test shots *Doll* and *TextScene* by only deblurring the object of interest while leaving the background unprocessed (SLD). Selecting the interval, deblurring is applied in, is guided by the captured depth map. The doll is located 14 cm in front of the focus. The text plane is located 20 cm in front of the focus. The used selected method is SAP.



(a) The blurred test image *DofFan*

(b) Globally deblurring with SAP with only one optimized PSF and weight for the entire image



(c) SAP, PSF(z) + W($\hat{z}$) and real depth map

(d) SAP, PSF($\hat{z}$) + W($\hat{z}$) and real depth map

Figure 15: Visual comparison of increasing complexity in our depth models using the test image *DofFan*. Employing the learned and optimized PSFs and weights in 15d shows the most convincing results. The employed selected method was SAP.

References

[1] T. F. Chan and C.-K. Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, 1998.

[2] F. Couzinie-Devy, J. Sun, K. Alahari, and J. Ponce. Learning to Estimate and Remove Non-uniform Image Blur. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1075–1082, 2013.

[3] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. 3D Microscopy Deconvolution using Richardson-Lucy Algorithm with Total Variation Regularization [Research Report] RR-5272 . Technical report, INRIA, 2004.

[4] K. Egiazarian, J. Astola, and N. Ponomarenko. New full-reference quality metrics based on HVS. volume 4, 2006.

[5] H. E. Fortunato and M. M. Oliveira. Fast high-quality non-blind deconvolution using sparse adaptive priors. *The Visual Computer*, 30(6-8):661–671, 2014.

[6] T. Hach and J. Steurer. A novel RGB-Z camera for high-quality motion

DOI: 10.2352/ISSN.2470-1173.2016.19.COIMG-168

picture applications. In *Proceedings of the 10th European Conference on Visual Media Production*, pages 1–10, 2013.

[7] T. Hach, J. Steurer, A. Amruth, and A. Pappenheim. Cinematic Bokeh rendering for real scenes. In *Proceedings of the 12th European Conference on Visual Media Production*, pages 1–10, 2015.

[8] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Scholkopf. Fast removal of non-uniform camera shake. In *2011 IEEE International Conference on Computer Vision*, pages 463–470, 2011.

[9] J. Jia. Mathematical models and practical solvers for uniform motion deblurring. In *Motion Deblurring*, pages 1–30. Cambridge University Press, Cambridge, 2009.

[10] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Transactions on Graphics*, 26(3):96, 2007.

[11] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-Laplacian priors. In *Advances in Neural Information Processing Systems*, pages 1033–1041, 2009.

[12] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 233–240, 2011.

[13] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(99):70–10, 2007.

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1964–1971, 2009.

[15] F. Li, J. Yu, and J. Chai. A hybrid camera for motion deblurring and depth map super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[16] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79(6):745–754, 1974.

[17] T. Luft, C. Colditz, and O. Deussen. Image enhancement by unsharp masking the depth buffer. *ACM Transactions on Graphics*, 25(3):1206, 2006.

[18] J. G. Nagy and D. P. O'Leary. Restoring Images Degraded by Spatially Variant Blur. *SIAM Journal on Scientific Computing*, 19(4):1063–1082, 1998.

[19] H. Navarro, G. Saavedra, M. Martinez-Corral, M. Sjostrom, and R. Olsson. Depth-of-Field Enhancement in Integral Imaging by Selective Depth-Deconvolution. *Journal of Display Technology*, 10(3):182–188, June 2015.

[20] W. H. Richardson. Bayesian-Based Iterative Method of Image Restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972.

[21] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[22] L. Sun, S. Cho, J. Wang, and J. Hays. Good Image Priors for Non-blind Deconvolution. In *Computer Vision – ECCV 2014*, pages 231–246. Springer International Publishing, 2014.

[23] Y.-W. Tai, H. Du, M. S. Brown, and S. Lin. Image/video deblurring using a hybrid camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[24] Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-Lucy Deblurring for Scenes under a Projective Motion Path. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1603–1618, 2011.

[25] M. Temerinac-Ott, O. Ronneberger, R. Nitschke, W. Driever, and H. Burkhardt. Spatially-variant Lucy-Richardson deconvolution for multiview fusion of microscopical 3D images. In *8th IEEE International Symposium on Biomedical Imaging*, pages 899–904, 2011.

[26] A. N. Tikhonov and V. Arsenin. *Solutions of ill-posed problems*. Vh Winston, 1977.

[27] P. H. Van Cittert. Zum Einfluß der Spaltbreite auf die Intensitätsverteilung in Spektrallinien. II. *Zeitschrift für Physik*, 69(5-6):298–308, 1931.

[28] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[29] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform Deblurring for Shaken Images. *International Journal of Computer Vision*, 98(2):168–186, 2011.

[30] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press, Cambridge, MA, 1949.

[31] L. Xu and J. Jia. Depth-aware motion deblurring. In *IEEE International Conference on Computational Photography*, pages 1–8, 2012.

[32] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics*, 26(3):1, 2007.

[33] T. Yue, J. Suo, and Q. Dai. High-Dimensional Camera Shake Removal With Given Depth Map. *IEEE Transactions on Image Processing*, 23(6):2688–2703, 2014.

Author Biography

Thomas Hach received his B.Sc. and his Dipl.-Ing. degree from the Technical University Munich (TUM) in 2011 and 2013, respectively. Since then he is working as part of the R&D department of Arnold & Richter Cine Technik. His current focus lies on depth sensing and sensor fusion signal processing, which is part of his doctorate supervised by the chair for data processing at TUM.

Arvind Amruth received his B.Sc. for Media-Technology from Ilmenau University of Technology (TUI, 2013). Subsequent M.Sc. graduation in electrical and information technology was gained from Technical University Munich (TUM, 2015). After looking back on a three year student trainee time in the R&D department for ARRI he currently works as junior engineer for image sensors at ARRI.