Which Side of the Focal Plane Are You on?

Anita Sellent* and Paolo Favaro
Institut für Informatik und Angewandte Mathematik
Universität Bern, Switzerland
sellent@iam.unibe.ch, favaro@iam.unibe.ch

Abstract

Defocus blur is an indicator for the depth structure of a scene. However, given a single input image from a conventional camera one cannot distinguish between blurred objects lying in front or behind the focal plane, as they may be subject to exactly the same amount of blur. In this paper we address this limitation by exploiting coded apertures. Previous work in this area focuses on setups where the scene is placed either entirely in front or entirely behind the focal plane. We demonstrate that asymmetric apertures result in unique blurs for all distances from the camera. To exploit asymmetric apertures we propose an algorithm that can unambiguously estimate scene depth and texture from a single input image. One of the main advantages of our method is that, within the same depth range, we can work with less blurred data than in other methods. The technique is tested on both synthetic and real images.

1. Introduction

When a 3D object is imaged through a lens, objects at different distances to the camera are recorded with different sharpness of detail. The further away an object is placed from the focal plane of the camera, the blurrier it occurs in the image. By detecting the size of the out-of-focus blur, depth can be estimated from a single image [16, 21, 31]. While the blur-size is determined by the distance to the focal plane, the shape of the out-of-focus blur is determined by the shape of the aperture. In digital imaging, this shape usually resembles the eye and pupils of humans, i.e., they are approximately circular. However, to imitate the pupil shape of the human eye may not be the most advantageous choice for single image depth estimation. Indeed, it has been shown that coded apertures [14, 18, 27] enable more reliable identification of blur sizes. High frequency aperture masks preserve a larger portion of high frequency image details than conventional apertures. These details can be exploited for depth estimation. Still, defocus blur destroys or at least attenuates some high frequency image content. The larger the blur, the more image frequencies are lost, rendering depth estimation as well as deblurring more ambiguous and susceptible to image-noise [25]. Thus depth estimation based on defocus cues works best when the blur size is reasonably small.

Figure 1: Sepia officinalis (Cuttlefish) possesses high visual acuity eyes with a peculiar W shaped pupil, whose function has been connected to depth estimation [24].

Most previous single image depth estimation approaches place the entire scene on one side of the focal plane [14, 18]. When symmetric apertures are used, this is necessary to obtain an unambiguous depth estimate, Fig. 2. Objects between the camera lens and the focal plane produce an image of the lens aperture (a blur) with the same orientation, Fig. 2, top row. Objects beyond the focal plane produce an image of the lens aperture that is flipped both vertically and horizontally, Fig. 2, bottom row. When the aperture shape is symmetric, the original blur and its flipped version cannot be distinguished. For scenes with large depth differences, unique blurs for symmetric apertures can become quite large, sacrificing many frequencies from the sharp image. Asymmetric apertures, however, allow to distinguish between any position in space regardless of where the focal plane lies. The camera can thus focus in the middle of the scene so that smaller blur radii are obtained for the same range of depths. Or, for the same maximal blur radius, we increase the number of distinguishable blur levels by a factor of two. Asymmetric pupils have actually also evolved in

*Funded by DFG-Fellowship SE 2134/1

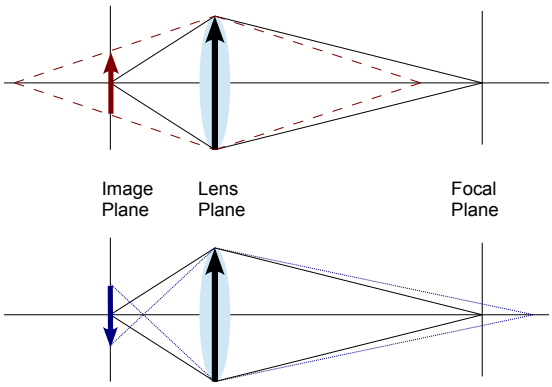


Figure 2: The defocus blur of an object located in front of (top) or behind (bottom) the focal plane are related by a horizontal and vertical reflection.

the animal kingdom, e.g., the eyes of cuttlefish, *sepia officinalis*, have a distinctive asymmetric W shape, Fig. 1.

We propose to exploit asymmetric apertures to estimate the complete range of depths in space given a single coded image. To maximize depth discrimination performance, we seek for the optimal asymmetric aperture mask by adapting a recent method [25] to dealing with depth levels on both sides of the focal plane, Sect. 3. For these asymmetric masks we derive an efficient depth estimation algorithm, Sect. 4. We can thus compare depth estimation on only one side of the focal plane to depth estimation on both sides of the focal plane, Sect. 5.

2. Related Work

Introducing masks into the aperture of a conventional camera, we profit from research that was previously published. Levin *et al*. [14] and Veeraraghavan *et al*. [27] first use coded masks in photography to estimate both depth and focused images. The approaches rely on deblurring to identify the blur that causes the recorded image. An earlier contribution of Dowski and Cathey [7] and later work by Martinello and Favaro [17, 18] circumvent the problem of deblurring the image at all depth levels for depth estimation. They use either filters complementary to the aperture mask [7], a bayesian approach [17] or an efficient filtering approach [18] to estimate depth from a single coded aperture image. In particular, the latter shows good performance also in face recognition [22] and in video applications [19]. Unfortunately, it is invariant to the flipping of the mask, Sect. 4.

Usually, a scene consists of several objects at different depth levels. If for depth estimation deblurring is performed on the entire image [14, 27], ringing artifacts can propagate and encourage wrong depth estimates. In setups with multiple images, boundaries for piecewise space variant blur can be inferred, e.g., via a variational approach [9]. For single image deblurring, segmentation for large blur differences can also be inferred via an edge map [2]. Once regions of constant depth are known, space variant deblurring can be applied directly [2]. Suitable regularization in this approach allows to preserve the grey-value level of each segment. In our approach we obtain a segmentation of the image using a scale estimation approach [18] and can thus apply piecewise constant deblurring.

Previous deblurring based approaches to single image depth estimation use the reconstruction error [14, 27] and the kurtosis [27] of the image to detect the correct blur scale. However, in conjunction with our deblurring method we found that these criteria are not equally well suited to identify the correct orientation of a blur kernel. Instead, other reference-free quality metrics can be applied, as they have recently been introduced to assess the quality of denoised [30] or deblurred [6, 3] images. In particular the latter two approaches focus on the detection of ringing artifacts with a criterion that can be computed sufficiently fast to be evaluated repeatedly.

Many of the works using coded defocus blur for depth estimation also considers which possible mask is optimal for that purpose. Levin *et al*. [14] consider the Kullbach-Leibler divergence between statistical formulations of the blurring process. The masks with optimal depth discrimination - obtained by a full-search of all binary masks - all turn out to be symmetric. We obtained similar symmetric masks in Ref. [25]. Additionally, our formulation and optimization scheme allow to find the masks quickly. Other approaches [27] to find aperture masks require them to be broadband in order to destroy as little frequency content as possible. Zhou and Nayar [29] and Masia *et al*. [20] focus on the deblurring capabilities of coded aperture masks and ignore depth estimation. Similarly to Levin *et al*., these latter approaches require a complete search to find the optimal mask. However, the masks they propose turn out to be asymmetric and can be used in our depth estimation algorithm, see Sect. 5.

Currently, asymmetric point-spread functions are exploited for depth estimation using multiple input images [11] or active illumination [10]. Asymmetric masks are also frequently used in wavefront coding [8], where the point-spread function is designed to be invariant to depth. In contrast, our focus is passive depth estimation from a single input image using only a conventional camera equipped with a coded aperture.

3. Optimized Masks for Depth Estimation on both Sides of the Focal Plane

For accurate depth estimation, images at different depths should be sufficiently different from each other. That is,

(a) 3×3 (b) 5×5 (c) 7×7 (d) 11×11 (e) 13×13

Figure 3: Top row: optimized asymmetric masks on different grid-sizes with transmission coefficients in $[0, 1]$. Bottom row: masks obtained by thresholding the transmission coefficients to $\{0, 1\}$.

for a given texture at a given set of distances from the focal plane, the blurred versions of that texture should differ as much as possible from each other. We implemented this idea into a mask evaluation criterion [25]. The aperture of the camera is divided into an $n \times n$ grid of elementary apertures with transmission coefficients $0 \leq \alpha_i \leq 1$. Each elementary aperture i contributes a depth-dependently blurred image g_i^d. These elementary images sum up to form the final image on the sensor plane $g_\alpha^d = \sum_i \alpha_i g_i^d$. To keep the brightness of the final image independent from the open apertures, the normalization $\sum_i \alpha_i = 1$ is used. For different blur levels d_1 and d_2 the difference of the blurred versions of a texture f is considered, i.e., $Df_{1,2} = \|g_\alpha^{d_1} - g_\alpha^{d_2}\|_2^2$. To obtain a good depth discrimination, this distance should be maximal.

While depth from defocus approaches use several images to resolve the ambiguity between texture and blur, single image approaches rely on the presence of depth-specific structures that are unique to each scale. Therefore, in Ref. [25] a projector P is introduced to eliminate all frequency content that is common in natural images. So the distance $PDf_{1,2} = \|Pg_\alpha^{d_1} - Pg_\alpha^{d_2}\|_2^2$ is considered to find optimized masks for single image depth estimation. This distance is averaged over a database F of sharp textures and a given set of depth levels $B = \{d_1, \ldots, d_k\}$. Finally, a term is added that encourages open elementary apertures. Then optimized aperture masks are obtained by solving the problem

$$\max_\alpha \quad \sum_{d_1 \in B} \sum_{d_2 \neq d_1} \sum_{f \in F} PDf_{1,2} - \lambda \|\alpha\|_2^2 \qquad (1)$$

$$s.t. \quad \sum_{i=1}^{n^2} \alpha_i = 1,\ 0 \leq \alpha_i \leq 1\ \forall i \in \{1, \ldots, n^2\}$$

In accordance with previous approaches [14, 27, 18], Ref. [25] considers only depth levels on one side of the focal plane. All the resulting aperture masks turn out to be symmetric. In our approach we consider depth levels on both sides of the focal plane. By using a set $\tilde{B}$ of depth levels on both sides of the focal plane, we obtain the asymmetric masks shown in Fig. 3, top row. In particular we use the same database F of natural textures, the same parameter settings and the same strategy to find suitable regularization parameter as in Ref. [25]. Experiments in this work also show that the optimized masks perform best in noise free images. If noise is present in the image, binarized masks perform better than their continuous valued counterparts [25]. For binarization only values larger than $\frac{1}{n^2}$ are set to 1; all other values are set to 0. The thresholded versions of our asymmetric masks are shown in Fig. 3, bottom row. For all experiments in Sect. 5 we use the thresholded versions of the masks.

4. Depth Estimation

In our approach we use asymmetric masks to obtain different point-spread functions (PSFs) for objects located in front and behind the focal plane. Theoretically, we could deblur the image with a set of known blur kernels, i.e., perform the approach suggested by Levin *et al.* or by Veeraraghavan *et al.* [14, 27] with all scales and both possible orientations of the blur kernel. However, the scale estimation method of Martinello and Favaro [18] does not require to solve the inverse problem of deblurring for depth estimation. As it is more accurate and much faster than the deblurring approaches, we profit from this algorithm to obtain the scale of the defocus blur. We show that the approach is invariant to the orientation of the blur kernel even for asymmetric aperture masks. To determine the orientation of the blur kernel, we still resort to deblurring. However, with the known scale and known neighborhood, the two required deblurring operations - one for each orientation of the kernel - can be performed quickly and accurately. By applying reference-free quality metrics to the deblurred images, we can learn to distinguish the correct orientation of the blur kernel.

4.1. Scale Estimation

The depth estimation algorithm in Ref. [18] describes the coded defocus blur of an $m \times m$ image patch g at a certain depth d as a linear operation on the sharp image f,

$$g_d = H_d f \quad . \qquad (2)$$

The basic assumption of the algorithm is that all blurred natural images at depth level d are contained within a subspace. The projection to the perpendicular space of this subspace, $H_d^\perp$ is such that $H_d^\perp g_d = 0$ and can be obtained in two ways. Formally,

$$H_d^\perp = I - H_d (H_d^\top H_d)^{-1} H_d^\top \qquad (3)$$

where I is the identity matrix and $H^\top$ indicates the transpose of a matrix H. Eq. (3) can be derived from solving

Eq. (2) via pseudo-inverse. This notation is helpful to show the invariance of the approach to flipped blur kernels. For a region of constant depth, the blurring with a flipped kernel can be written as $g_{\tilde{d}} = H_d^\top f$. Using Eq. (3) we obtain $H_d H_{\tilde{d}}^\perp g_d = 0$ and thus $H_{\tilde{d}}^\perp g_d \in (H_{\tilde{d}}^\perp \cap H_d^\perp)$. As projections are idempotent, we can derive from $H_d^\perp H_{\tilde{d}}^\perp g = H_{\tilde{d}}^\perp g$ and Eq. (3) that actually $H_d^\perp = H_{\tilde{d}}^\perp$, i.e., the linear subspaces for the kernel and its flipped versions are identical.

In practice, $H_d^\perp$ can be learned from sample images: For a set of sharp image patches from a database $f \in \tilde{F}$, images at all the blur levels are recorded or synthesized. All experiments in this article use synthetically blurred images with a calibrated PSF. For each depth level d and texture f the synthesized images are stacked in a vector h_f^d and arranged in a matrix F_d. This matrix is decomposed via singular value decomposition to $F_d = U_d S_d (V_d)^\top$. By thresholding the singular values, $U_d = [P_d, Q_d]$ can be divided into two subspaces where P_d corresponds to the large singular values and spans the subspace of blurred textures at level d. Q_d is perpendicular to the subspace P_d. The operator $H_d^\perp = Q_d (Q_d)^\top$ is the projection to the subspace Q_d.

For a given set of blur scales $D = \{|d_i| \text{ , } d_i \in B\}$ scale identification reduces to finding the most perpendicular subspace. As a single image may contain objects at different depths, we consider small $m \times m$ patches to determine the depth level of each center independently. To obtain smooth depth maps, we regularize the depth estimation and consider

$$d = \arg\min_{d \in D} \|H_d^\perp g\|^2 + \beta |\nabla d| \qquad (4)$$

with $\beta > 0$, which can be solved via graph-cuts [5, 12, 4].

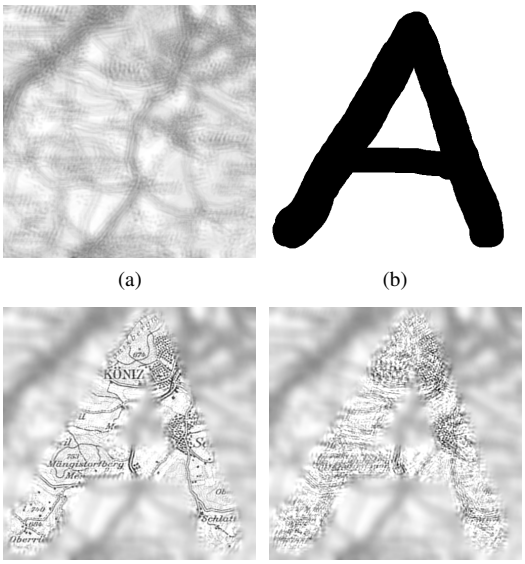
If not indicated otherwise, we use a patch size $m = 27$, 20 equally spaced blur levels between 1 and 20 pixels, and $\beta = 0.0005$ for normalized images.

4.2. Blur Kernel Orientation Estimation

The algorithm of the previous section allows to estimate the scale of the blur at each pixel. To determine the sign of the distance to the focal plane we deblur the image. Deblurring the entire image with the same blur kernel can cause ringing artifacts from pixels with different scales. We therefore consider a region A_d of pixels with the depth scale d. Within the region we want the deblurred image $\tilde{f}_d$ and the input image g to correspond under the blurring operation, i.e., $H_d \tilde{f}|_{A_d} = g|_{A_d}$ where $|_{A_d}$ is the restriction to A_d. Outside the region A_d the image $\tilde{f}_d$ should continue smoothly. Thus we construct a weight function

$$w(x) = \begin{cases} \frac{1}{dist(\bar{A}_d)} & \text{if } x \in A_d \\ 1 & \text{if } x \notin A_d \end{cases} \qquad (5)$$

where $\bar{A}_d$ is the complement of A_d in the image domain. This weight function is 1 outside of A and decreases linearly



(a) (b) (c) (d)

Figure 4: To account for depth variations, we deblur a given input image (a) with a mask (b) that only considers pixels with a certain scale. If using the correct orientation of the kernel for deblurring, we obtain a basically artifact-free image (c). Using the flipped kernel, we obtain an image with different characteristics (d).

with increasing distance to the boundary of A. So we can find the deblurred image f_d by minimizing the energy

$$E(f_d) = \|H_d f_d|_{A_d} - g|_{A_d}\|_2^2 + \gamma w \|\nabla f_d\|_2^2 \quad . \qquad (6)$$

We find an optimum of this functional via conjugate gradient using forward differences for the image gradient $\nabla f \approx Df$ and thus $\nabla E = H_d^\top H_d f|_{A_d} - H_d^\top g|_{A_d} + \gamma w D^\top D f$. As this simple deblurring algorithm does not consider occlusions, it generally requires the sharp image also beyond the boundaries. We therefore enlarge the region A_d in our implementation by half the blur kernel size. For all experiments, we use $\gamma = 0.005$ and 50 iterations in the conjugate gradient approach.

Given locally deblurred images for each possible orientation of the blur, f_{+d} and f_{-d}, we want to decide which orientation is the correct one. For this we consider Gaussian windows around a pixel x, i.e., p_{+d} and p_{-d} with the same patch size as for scale estimation. As we use only weak regularization of the image, the reconstruction error [14, 27] for both setups is approximately the same and unsuitable for the decision. In our experiments, we found the following measures to work best:

- As proposed by Veeraraghavan *et al*. [27] we consider the kurtosis of a patch p to detect ringing artifacts. $F_1(x) = \sum_{y \in p} (\frac{\nabla p(y) - \bar{\nabla} p}{\sigma_p})^4$ where $\bar{\nabla} p$ and σ_p are the

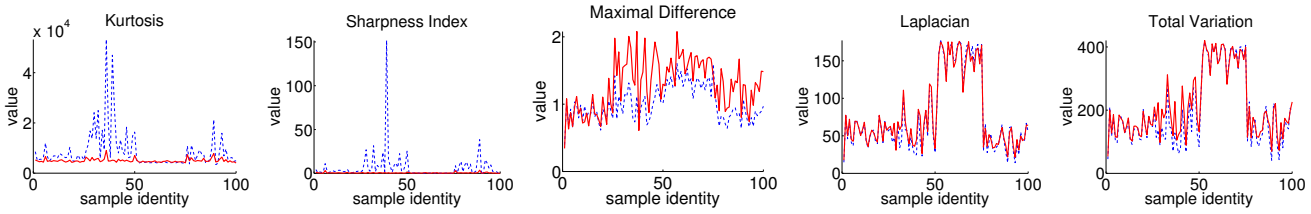


Figure 5: Deblurring an image with the correctly orientated (dotted) or flipped (solid) blur kernel results in differences in the considered, reference-free quality metrics. Here the values of each considered metric for 100 sample patches are shown. We learn the relative weight of each metric via a support vector machine.

mean and the standard deviation of the gradient of the patch.

- The sharpness index is an image metric formalised by Blanchet and Moisan [3] to be sensible also to ringing artifacts. It can be computed fast on a local patch via $F_2(x) = SI(p) = -log_{10}\Phi\left(\frac{\mu - TV(p)}{\sigma}\right)$ where $TV(p)$ is the total variation of p, $\mu = \mathbb{E}TV(p)$ and $\sigma^2 = Var(TV(p))$ its expectation and variance, respectively, and Φ the tail of the Gaussian distribution, [3].
- When an image is deblurred with a different kernel than the true blur kernel, more extreme values are required to compensate for the wrong information [15]. We therefore measure the spread of deblurred patches, i.e. $F_3(x) = max(p) - min(p)$
- The Laplacian of an image can be used as a regularization term for image deblurring $F_4(x) = \sum_{y \in p} |\Delta p(y)|$ [13]. We use it as a ringing detector.
- Total variation is used as regularization on the boundary in our deblurring algorithm. Choosing a small value for γ, however, still allows for the evaluation of total variation as a quality measure $F_5(x) = \sum_{y \in p} |\nabla p(y)|$.

All these qualities measures are summarized in a feature vector $F^+(x)$ and $F^-(x)$ for deblurring with the two orientations of the kernel, Fig. 5. The relative importance of each feature is then learned via a support vector machine (SVM) [23]. We train the SVM on images blurred synthetically with one of the blur kernels and deblur it with both kernels to learn a vector of weights ν so that $\nu^\top(F^+ - F^-) > 0$ for patches blurred with the original kernel.

To propagate the flipping also in regions with low texture, we regularize the flipping decision

$$sign(x) = \max_{s \in \{+,-\}} s\nu^\top(F^+ - F^-) + \zeta \sum_{y \in N_x} \delta(s(x), s(y))$$

with $\zeta > 0$ and N_x a spatial neighborhood of x. We use graph cuts to obtain a global solution to this problem, [5]. Note that on the binary decision problem the algorithm is indeed guaranteed to find the globally optimal solution. In our experiments, we use 8-connected neighborhoods and, if not indicated otherwise, $\zeta = 3$.

After both scale and sign have been estimated, an all-in-focus image that is sharp everywhere can be generated. As the images already have been deblurred for the flipping detection, a sharp image can be simply reconstructed by assigning to each pixel the value of the image patch deblurred with the correctly orientated kernel.

5. Experiments

We validate our contributions in several experiments. First, we perform an evaluation on synthetic images with known ground truth. Then we show results on real imagery.

5.1. Synthetic Experiments

To evaluate our asymmetric masks and to validate our approach we first run the algorithm on a stair data set with 20 equally spaced levels, using three different textures, see Ref. [25]. For each texture we determine the mean squared blur scale error between the ground truth blur scale l_{gt} and the estimated blur scale l_{est} on the occlusion-free part of the image Ω, i.e. $e = \frac{1}{|\Omega|}\sum_{x \in \Omega}(l_{est} - l_{gt})^2$. We then average the mean squared error over the three data sets.

As aperture masks block the incoming light to a different degree, we evaluate the robustness of the masks against noise. Gaussian noise of zero mean and given variance σ^2 is added to an image, where the largest transmission value is set to 1. Then the images are normalized to $[0, 1]$. Thus images with more open apertures are less noisy than images with few open apertures. For more accurate evaluation, we set the regularization parameter of the scale estimation $\beta = 0$ in the synthetic experiments.

First, we generate a reference depth estimation error by considering a depth volume on only one side of the focal plane with a blur size between 1 and 22 pixel. The parameters of the algorithm are set to the same range of depth level. The dashed lines in Fig. 6 show that in this setup the depth estimation error increases considerably with the noise level.

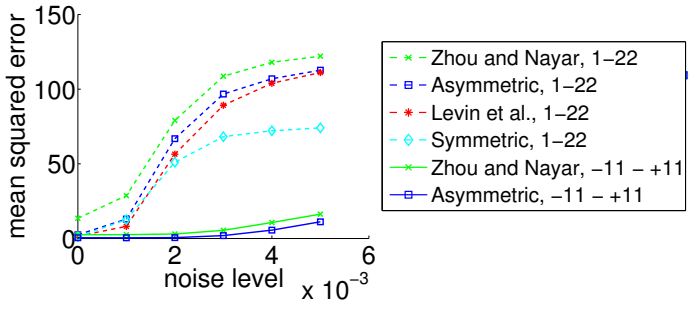
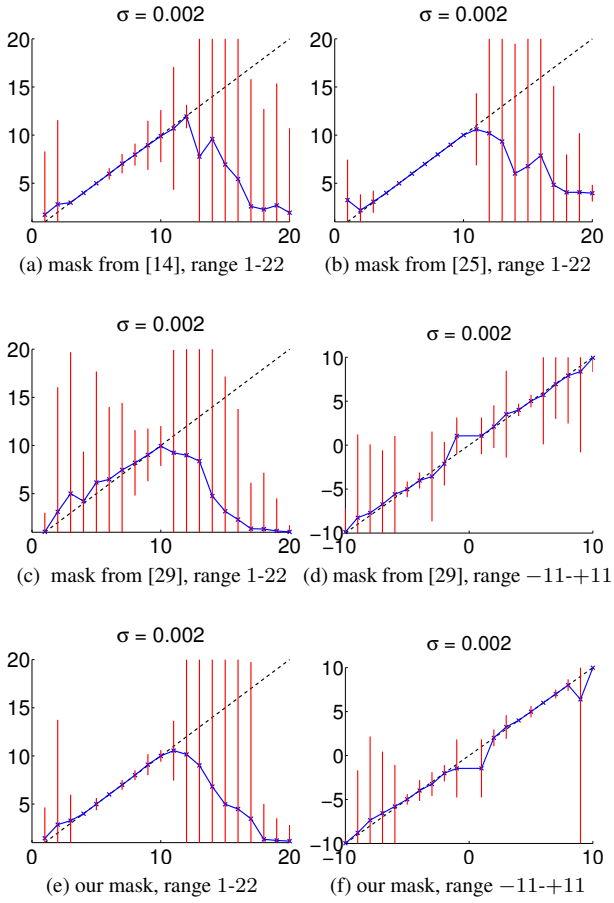


Figure 6: For the approach with blur sizes in the range 1-22 pixels (dashed), symmetric masks (red star [14], cyan diamonds [25]) have a smaller depth estimation error than asymmetric masks for the same noise level. Our approach with blur sizes in the range from -11-$+11$ pixel (solid) allows for more accurate depth estimation. Our optimized asymmetric mask (blue squares) is even more accurate than comparable asymmetric masks [28] (green crosses).

Thereby, the error of asymmetric masks is higher than the error of the symmetric masks from Levin *et al*. [14] and Ref. [25].

Second, we consider a depth volume centered on the focal plane so that blur sizes have a range between -11 and -1 pixel and 1 and 11 pixel. Note that a blur size of ± 1 pixel corresponds to the original sharp image and the kernel orientation is indeed undetermined. We find our algorithm to be considerably more robust to noise than the reference estimation on one side of the focus plane, Fig. 6. Also, the optimized mask from Sect. 3 performs slightly better in depth estimation than the reference mask [29] of the same resolution that was originally designed for deblurring. Both masks have comparative performance on the deblurring-based sign detection, i.e., for the representative noise level $\sigma = 0.002$ the reference masks estimates the sign correctly for 94% of the pixels (including the blur sizes of ± 1) while our optimized masks estimated 93% of the signs correctly. However, for the scale estimation on the same noise level, our mask obtains an average mean squared blur level error of 0.56 while the reference mask has an error of 3.00.

The increased robustness of our algorithm in comparison to single-sided scale estimation on the same depth volume becomes obvious in the error-plots for the moderate noise level $\sigma = 0.002$, Fig. 7. While the detection of small blur sizes is still quite accurate, noise induces the algorithm to confuse large blur sizes with very small blur sizes. The sign



(a) mask from [14], range 1-22 (b) mask from [25], range 1-22

(c) mask from [29], range 1-22 (d) mask from [29], range -11-$+11$

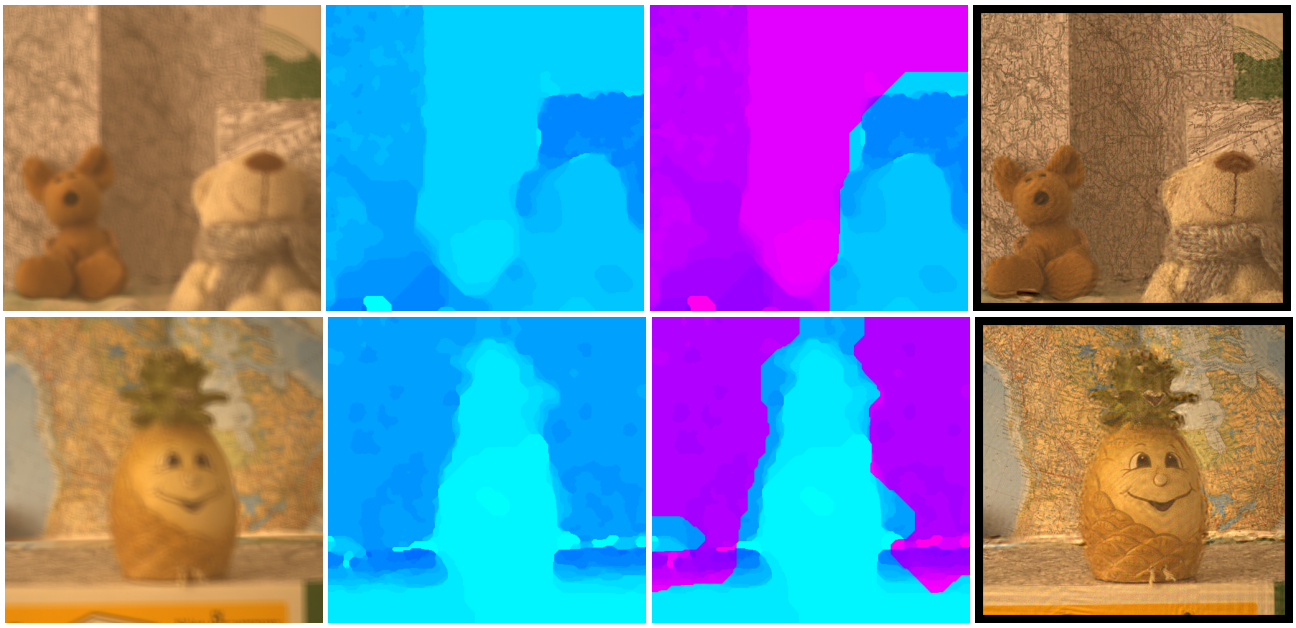
(e) our mask, range 1-22 (f) our mask, range -11-$+11$

Figure 7: For one sided depth estimation with blur sizes $1 - 22$ pixels all masks show a breakdown at large blur levels for the moderate noise level $\sigma = 0.002$. The black dotted line shows the correct blur scale level, the blue line the average of the estimated blur scale level and the red bar thrice the standard deviation of the estimate. For depth estimation with blur sizes -11-$+11$ our algorithm can cover the depth volume more robustly.

detection, however, is sufficiently robust to deal with the noise. Only for the smallest blur size of ± 1 pixel, the algorithm cannot detect the flipping of the blur kernel but can only give the information that the object is within the depth of field.

5.2. Real Imagery

We acquire real images by introducing the binarized mask in Fig. 3b into a Canon EF 50mm f/1.8 II lens [1] that we attach to a Canon EOS 5D, Mark II camera. We calibrate the camera by acquiring a single PSF from a calibration dot. All other PSF scales on both sides of the focal



(a) input image (b) blur scale map (c) signed blur map (d) deblurred image

Figure 8: Starting with a single input image, (a) our algorithm first determines the blur scale (b), color coded with [26]. Then deblurring is used to determine if the PSF is flipped, resulting in a signed depth map, (c). The signed depth map has generally smaller blur sizes which give a more accurate depth estimation result, Fig. 7 and also give a good deblurring result (d).

plane are generated synthetically from the measured image. In each of the acquired scenes the focal plane is placed approximately in the middle of the visible depth volume. The smaller blur scales in the middle of the scenes are visible in the depth scale image in column (b). To obtain optimal results for the sign estimation, we set $\zeta = 17$ for the scene *animals* (top row) and $\zeta = 3$ for the scene *pinapple* (bottom row). In both scenes the sign detection is able to distinguish objects placed in front of the focal plane from objects behind it. We observe that the sign detection tends to assign the orientation of the foreground to pixels belonging to the background around occlusion boundaries. As the algorithm does not include an occlusion model, and only the foreground is visible to the entire aperture, this is the expected behavior. Also, in the *pinapple* scene we observe an occasional sign confusion at the low textured regions close to the focal plane.

By considering the deblurred images in column (d) and the closeups in Fig. 9 we observe a good reconstruction of fine image details and only few ringing artifacts. Especially in the closeups of the large blur scales, the high quality of the reconstruction of our algorithm becomes evident. Note that depth estimation without flipping would have to deal with twice the amount of blur and thus make the reconstruction of details much harder.

6. Conclusion

In this work we exploit the ability of asymmetric masks to distinguish every depth in space, regardless of whether one is in front of or behind the focal plane. We optimized asymmetric aperture masks for their depth discrimination capability. These masks allow to focus coded aperture cameras in the middle of a scene rather than very close to the camera or at infinity. The most immediate benefit is that the image captured with a central focus plane is less blurry and hence retains more texture detail. The overall improvement in depth estimation and image restoration has been validated experimentally on both synthetic and real images.

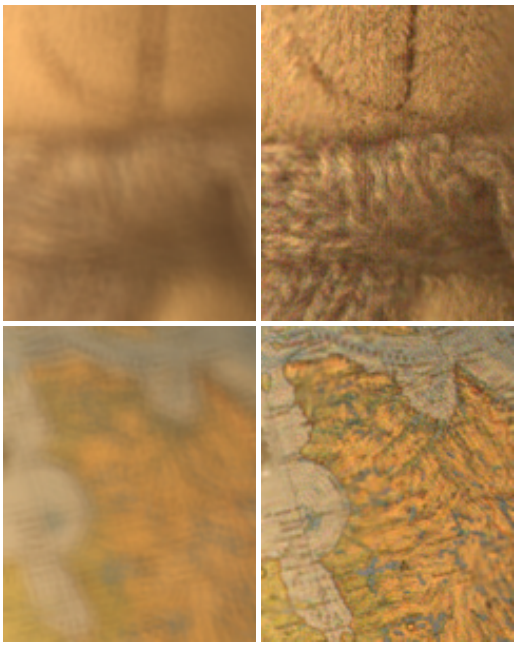
References

[1] Y. Bando. How to disassemble the canon EF 50mm f/1.8 II lens, 2013. http://web.media.mit.edu/ bandy/rgb/disassembly.pdf.

[2] L. Bar, N. Sochen, and N. Kiryati. Restoration of images with piecewise space-variant blur. In *Scale Space and Variational Methods in Computer Vision*, pages 533–544. Springer, 2007.

[3] G. Blanchet and L. Moisan. An explicit sharpness index related to global phase coherence. In *ICASSP*, pages 1065–1068. IEEE, 2012.

[4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *T-PAMI*, 26(9):1124–1137, 2004.



(a) input (b) deblurred

Figure 9: A closeup from the input images and the deblurring result shows how efficiently the considerable blur is removed and fine detail is reconstructed.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *T-PAMI*, 23(11):1222–1239, 2001.

[6] F. Calderero and P. Moreno. Evaluation of sharpness measures and proposal of a stop criterion for reverse diffusion in the context of image deblurring. In *VISAPP 2013*, Barcelona, 02/2013 2013.

[7] E. Dowski and W. Cathey. Single-lens single-image incoherent passive-ranging systems. *Applied Optics*, 33(29):6762–6773, 1994.

[8] E. Dowski and W. Cathey. Extended depth of field through wave-front coding. *Applied Optics*, 34(11):1859–1866, 1995.

[9] P. Favaro, M. Burger, and S. Soatto. Scene and motion reconstruction from defocused and motion-blurred images via anisotropic diffusion. In *Computer Vision–ECCV*, pages 257–269, 2004.

[10] B. Girod and E. H. Adelson. System for ascertaining direction of blur in a range-from-defocus camera, Oct. 23 1990. US Patent 4,965,442.

[11] A. Greengard, Y. Schechner, and R. Piestun. Depth from diffracted rotation. *Optics Letters*, 31(2):181–183, 2006.

[12] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *T-PAMI*, 26(2):147–159, 2004.

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

[14] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. *TOG*, 26(3):70, 2007.

[15] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding blind deconvolution algorithms. *T-PAMI*, 33(12):2354–2367, 2011.

[16] J. Lin, X. Ji, W. Xu, and Q. Dai. Absolute depth estimation from a single defocused image. *IEEE T-IP*, 22:4545–4550, 2013.

[17] M. Martinello, T. Bishop, and P. Favaro. A bayesian approach to shape from coded aperture. In *Proc. ICIP*, pages 3521–3524. IEEE, 2010.

[18] M. Martinello and P. Favaro. Single image blind deconvolution with higher-order texture statistics. *Video Processing and Computational Video*, pages 124–151, 2011.

[19] M. Martinello and P. Favaro. Depth estimation from a video sequence with moving and deformable objects. In *Proc. IET*, 2012.

[20] B. Masia, L. Presa, A. Corrales, and D. Gutierrez. Perceptually optimized coded apertures for defocus deblurring. In *Computer Graphics Forum*. Wiley Online Library, 2012.

[21] V. P. Namboodiri and S. Chaudhuri. Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera. In *Proc. CVRP*, pages 1–6. IEEE, 2008.

[22] M. Nishiyama, A. Hadid, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi. Facial deblur inference using subspace analysis for recognition of blurred faces. *T-PAMI*, 33(4):838–845, 2011.

[23] S. Nowozin and C. Lampert. *Structured learning and prediction in computer vision*, volume 6. Now publishers Inc, 2011.

[24] F. Schaeffel, C. J. Murphy, and H. C. Howland. Accommodation in the cuttlefish (sepia officinalis). *Journal of Experimental Biology*, 202(22):3127–3134, 1999.

[25] A. Sellent and P. Favaro. Optimized aperture shapes for depth estimation. *Pattern Recognition Letters*, 40:96–103, 2014.

[26] A. Sellent, P. Lauer, D. Kondermann, and J. Wingbermühle. A toolbox to visualize dense image correspondences. Technical report, Heidelberg University, 2012.

[27] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *TOG*, 26(3):69, 2007.

[28] C. Zhou, S. Lin, and S. Nayar. Coded aperture pairs for depth from defocus. In *Proc. ICCV*, pages 325–332. IEEE, 2009.

[29] C. Zhou and S. K. Nayar. What are Good Apertures for Defocus Deblurring? In *Proc. ICCP*, Apr 2009.

[30] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *T-IP*, 19(12):3116–3132, 2010.

[31] S. Zhuo and T. Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852–1858, 2011.