Image translation for single-shot focal tomography: supplementary material

PATRICK LLULL, XIN YUAN, LAWRENCE CARIN, AND DAVID J. BRADY*

Fitzpatrick Institute for Photonics, Department of Electrical and Computer Engineering, Duke University, 129 Hudson Hall, Durham, North Carolina 27708, USA

*Corresponding author: david.brad@duke.edu

Published 17 September 2015

This document provides supplementary information to “Image translation for single-shot focal tomography,” <http://dx.doi.org/10.1364/optica.2.000822>. Herein, we describe the defocus model used in the primary manuscript, the deformable mirror calibration procedure, and details of our reconstruction algorithm. © 2015 Optical Society of America

<http://dx.doi.org/10.1364/optica.2.000822.s001>

S1. OPTICAL PROPAGATION MODEL

We assume a swept-focus or image translation system employs a rotationally-symmetric optical system of diameter A with a pupil function given by

$$P(x, y) = \text{circ}\left(\frac{\sqrt{x^2 + y^2}}{A}\right), \quad (\text{S1})$$

where $\text{circ}(\cdot)$ denotes the circle function. Incorporating the optical defocus for an object located at z_0 directly into the pupil function yields

$$P_{z_0}(x, y, z) = P(x, y) e^{-\frac{i\pi(x^2 + y^2)}{\lambda} \left| \frac{1}{F} - \frac{1}{z} - \frac{1}{z_0} \right|}, \quad (\text{S2})$$

where F is the focal length and z is the in-focus distance in object space. After a coordinate change to spatial frequency variables ($u = \frac{x}{\lambda z}$, $v = \frac{y}{\lambda z}$), we may express a defocused point spread function (PSF) as

$$h_{z_0}(x, y, z) = |\mathcal{F}\{P_{z_0}(u, v, z)\}|^2 \quad (\text{S3})$$

where $\mathcal{F}$ is the two-dimensional Fourier transform operator. An example defocused PSF with $A = 15\text{mm}$, $F = 50\text{mm}$, and $z_0 = \infty$ (with $z = 33\text{m}$ to optical infinity) is shown in Fig. 2 of the primary manuscript. These PSFs are integrated along the focal dimension to produce swept-focus and image translation PSFs.

S2. ALGORITHM DETAILS

Herein, we explain our image reconstruction algorithm in detail. We employ a patch-level alternating optimization method to find the 2D image and depth index corresponding to each patch. To illustrate the algorithm's flow, consider a small (presumably planar) patch $f_{z_0}^{(l)}$ (the superscript l denotes a patch-level entity) located a distance z_0 away from the camera, and assume we partition the focal volume into three focal channels as an example to demonstrate the algorithm. Please

note that our imaging system and algorithm can be readily extended to scenes with more layers. In the case of three focal channels, there are three possible outcomes for the detected image g that correspond to the three image translation kernels h_{s_2, z_i} , $i \in \{1, 2, 3\}$:

$$\begin{aligned} g_{z_1} &= h_{s_2, z_1} * f_{z_1} + \epsilon \\ g_{z_2} &= h_{s_2, z_2} * f_{z_2} + \epsilon \\ g_{z_3} &= h_{s_2, z_3} * f_{z_3} + \epsilon \end{aligned}$$

where $*$ denotes the convolution operator and ϵ is image noise. We observe a measurement $g^{(l)}$; thus, we need to select which image $g_{z_i}^{(l)}$ is correct (most similar to $g^{(l)}$). This equates to determining the object $f_{z_i}^{(l)}$ that is closest to the ground truth f_{z_0} , which motivates the first step of our algorithm. Our algorithm executes the following steps on each patch:

- 1) Given $g^{(l)}$ and $\{h_{s_2, z_i}^{(l)}\}, i = 1, 2, 3$, we estimate $\{\tilde{f}_{z_i}^{(l)}\}, i = 1, 2, 3$, using the hyperlaplacian priors deconvolution algorithm [1].
- 2) Select the best estimate $\hat{f}_{z_0}^{(l)}$ of the underlying scene $f_{z_0}^{(l)}$ and depth z_0 from $\{\tilde{f}_{z_i}^{(l)}\}, i = 1, 2, 3$.
- 3) Merge the images $\hat{f}_{z_0}^{(l)}$ and depths z_0 from each (overlapping) patch into a final image and depth map.

Steps 1) and 2) are performed for all overlapping patches; step 3) is performed once after steps 1) and 2) finish.

Step 1) of our algorithm (represented as red boxes in Fig. S1) employs a patch-level implementation of the hyperlaplacian priors deconvolution algorithm proposed by [1] to estimate the patch-level objects. The patches overlap according to Fig. S2 to improve the

Fig. S1. Image translation processing pipeline. Each step of our reconstruction algorithm is represented by a different box fill color. Red and gray colors respectively denote processing occurring in the first and second step. Blue and green boxes denote processing responsibilities of the third step; green boxes represents the final results. The steps within the red dashed lines are repeated for each patch and are performed in parallel.

quality and robustness of the image and depth estimation. We perform each deconvolution in parallel to greatly reduce the reconstruction time. The algorithm solves the following problem for the l^{th} image patch [1]:

$$\tilde{f}_{z_i}^{(l)} = \min_{f_{z_i}^{(l)}} \sum_{i=1}^N \left(\frac{\lambda}{2} (f_{z_i}^{(l)} * h_{s_2, z_i}^{(l)} - g^{(l)})_i^2 + \sum_{j=1}^2 |(f_{z_i}^{(l)} * c_j)_i|^\alpha \right), \quad (\text{S4})$$

where N represents the number of pixels within the l^{th} patch and c_j are the derivative filters $[1, -1]$ and $[-1, 1]^T$. λ is a regularization weight that balances the costs associated with the fitting term the hyperlaplacian (of power α) term. The values of λ and α are tuned to 2000 and .66 in our experiments, respectively.

Fig. S2. Overlapping patch concept drawing. The red outline denotes a local window of size 100×100 pixels; the green shaded box illustrates the current patch; the blue dotted lines denote the positions the patch slides in sub-window increments. Eq. (S5) is evaluated for each patch at each sub-window shift prior to voting.

After solving for a patch's candidate deblurred images $\{\tilde{f}_{z_i}^{(l)}\}, i = 1, 2, 3$, we are ready to proceed to step 2) of our reconstruction algorithm (shown as the gray box in Fig. S1). Here, we solve the following equation (a copy of Eq. (4) in the primary manuscript) to estimate the true patch-level object and depth:

$$\hat{f}_{z_o}^{(l)} = \arg \min_{\tilde{f}_{z_i}^{(l)}} \|g^{(l)} - h_{s_2, z_i}^{(l)} * \tilde{f}_{z_i}^{(l)}\|_2^2 + \tau \text{TV}(h_{s_2, z_i}^{(l)} * \tilde{f}_{z_i}^{(l)}). \quad (\text{S5})$$

Fig. S3. The advantage of using overlapping patches.

The regularization weight τ is set to -0.3 in our experiments; hence, this term rewards sharp data estimates $h_{s_2, z_i}^{(l)} * \tilde{f}_{z_i}^{(l)}$. Though this differs from traditional penalty functions that penalize object estimates $\tilde{f}_{z_i}^{(l)}$ containing spurious gradient content or noise, we regularize this way because the reconstructions from Eq. (S4) are already penalized for noise and ringing artifacts during step 1) of our algorithm. To select negative total variation on $g^{(l)}$ means we reward measurement estimates that are more informative from a spatial frequency perspective. This relates directly to the rotating PSFs (and the corresponding transfer functions): spatial frequency content of the image $g^{(l)}$ will change based on which kernel $h_{s_2, z_i}^{(l)}$ is used; thus, if the wrong kernel is modeled, the spatial frequency content in $g^{(l)}$ will be strictly lower than if the correct kernel is used. This is explained more precisely in [2]. We additionally reward total variation of the data estimate because the ℓ_2 data fitting term of Eq. S5 by itself is prone to errors associated with measurement noise in $g^{(l)}$ which is not present in $h_{s_2, z_i}^{(l)} * \tilde{f}_{z_i}^{(l)}$, since $\tilde{f}_{z_i}^{(l)}$ is already regularized in step 1). Since we're observing a blurred image rather than a coded image, contrast is attenuated due to the low transfer function values; this renders measurement noise a significant

portion of $\|g^{(l)} - h_{s_{2,z_i}}^{(l)} * \tilde{f}_{z_i}^{(l)}\|_2^2$. We believe better regularizers exist in this framework; however, we found this selection to work well in our experiments.

After obtaining each overlapping patch's final object and depth estimates, we proceed to our algorithm's third and final step (shown as blue and green boxes in Fig. S1). Here, we use a weighted voting scheme to estimate the sub-window shifted depth indices and object. We then merge the depth indices into a final depth map and the image patches into an all-in-focus image as shown in Fig. 7 of the primary manuscript. The difference between using overlapping patches and non-overlapping patches may be visualized in Fig. S3; note in particular the blocky structure and less-accurate depth indices.

Though this approach is capable of relatively fast all-in-focus object and depth map reconstructions, we believe there may be alternative ways to address the following disadvantages: 1) the algorithm performs N_z deconvolutions per patch; 2) Eq. (S5) does not directly exploit the linear relationship of peak PSF angle versus depth.

S3. DEFORMABLE MIRROR CALIBRATION PROCEDURE

In the experiments, we use a 17-electrode micromachined-membrane deformable mirror (DM) with two tip/tilt actuators (OKOTech's MMDM-17-TT). Though MEMS beam steering mirrors allow for faster tip/tilt rates over larger angular extents, they generally have small working aperture diameters (usually 1-7mm) and would require a separate focal modulator for our application. We choose the DM to allow for a larger working diameter (and hence lower F/#) and to facilitate component synchronization and alignment. As mentioned

Fig. S4. DM responses before and after linearization. The solid blue and red dashed curves show the parabolic relationship between input voltage and the optical power of the DM; the solid black curve shows the correction factor ranging from 1 to 2.1. The solid purple curve shows the measured response of the DM after applying the correction factor to the input voltage. The green dashed line is the target linear relationship.

in the primary manuscript, our DM's 17 electrodes can change the provided optical power between 0 and .7 diopters, which equates to a focusing range of infinity to 1.42 meters from the camera. Additionally, the DM's two tip/tilt actuators can provide $\pm 2\text{mrad}$ of angular motion, which equates to 15 detector pixels for our setup. Nineteen channels of our 40-channel DAC board and amplifier interface with the mirror and one channel provides external trigger signals to the camera. A host computer is connected via ethernet to this DAC board and communicates data to these 20 channels via UDP. Though this system is quite

capable for image translation, we must characterize it on global (i.e. all electrodes) and local (each electrode) scales before performing image translation experiments.

A. Focal sweep linearization

We modulate the DM's optical power by applying a flat voltage to all 17 electrodes; doing so changes the surface profile from flat to paraboloidal. The resulting optical power varies approximately quadratically with input voltage; we linearize this behavior to ensure each focal channel is allotted the same portion of the overall integration time Δt according to traditional focal sweeping [3].

To linearize the DM's optical response, we apply a correction factor (a ratio of the desired linear response to the actual response) to the input voltage level according to the target optical power. We calibrate the DM's response by sequentially applying 14 equispaced voltages (between 0 and 176V, the maximum operating voltage) to all electrodes. For each input voltage value, we record the object-space position (in diopters) that yields the sharpest contrast. After finding the correction factor, we repeat this experiment, this time scaling the input voltage by the correction factor.

Fig. S4 shows the optical power as a function of input voltage, a polynomial fit to the optical power curve, the correction factor, and the linearized optical response of the DM. The solid purple curve shows the measured optical power from the corrected input voltages after linearizing the response; we can see that it closely approximates the desired linear relationship (green dotted line).

B. System aberration correction

The previous section presented a correction factor required to linearize the global response of the mirror. Herein, we show how we correct each electrode locally to compensate for static optical aberrations. Since our objective lens is presumably well-corrected (an MTF value of .2 at 160 cyc/mm according to Edmund Optics' datasheet), the beamsplitter and DM are the main contributors to the system's aberrations.

Fig. S5. Image of a wheel captured by the experimental prototype. (left) without astigmatism correction, (right) after astigmatism correction. The tangential and sagittal foci are noticeably farther apart before correction.

The DM is fabricated with substantial oblique Zernike astigmatism, which must be corrected directly. Fig. S5 illustrates the severity of the astigmatism with an image of a resolution target wheel captured by the static image translation prototype.

We can approximate a continuous surface s given the 17 DM vectorized modal influence functions $\{d_i\}, i \in \{1, 2, \dots, 17\}$ (provided by OKOTech). Let $D = [d_i^T], \forall i$ denote a matrix of the influence functions and let (r, ϕ) denote the radial and azimuthal polar coordinates. The aberration compensating-surface $S(r, \phi)$ may be represented as a flat surface minus a linear combination of the Zernike polynomials $Z_n^m(r, \phi)$ (where n, m are the Zernike indices) that comprise the overall system aberrations:

$$S(r, \phi) = 1 - \sum_m \sum_n a_{m,n} Z_n^m(r, \phi) + b_{m,n} Z_n^{-m}(r, \phi). \quad (\text{S6})$$

According to mirror-specific data provided by OKOTech, $b_{2,2} = 1.007$ is the dominant coefficient (oblique astigmatism) at an angle of $\phi = 54^\circ$. We experimentally verified this by configuring the DM as the test arm of a Michelson interferometer (Fig. S6) and by analyzing the resulting interferogram using the OpenFringe program, which yielded a dominant coefficient of $b_{2,2} = 1.306$ in addition to much smaller coefficients for coma and spherical aberration. The astigmatism does not vary with optical power supplied by the DM. The DM surface profile needed to eliminate this astigmatism is portrayed in Fig. S7.

We denote a discrete approximation of the target surface as s . Given an array of input voltages v sent to the DM electrodes, a surface profile is generated as a linear combination of the actuator influence functions

$$s = Dv. \quad (\text{S7})$$

To approximate this surface, we use the pseudoinverse estimator to solve for the voltage array v :

Fig. S6. Michelson interferometer configuration (a) and close-up interferogram (b) to characterize the system's static aberrations. The laser's central wavelength is 543nm.

$$\hat{v} = D^\dagger s. \quad (\text{S8})$$

According to data provided by OKOTech, D 's smallest (17th) normalized singular value is $\sim .01$, rendering Eq. (S8) well posed. Our mirror's approximation to the desired surface, which is given by $D\hat{v}$, is shown in Fig. S7.

Fig. S7. Deformable mirror surface plots for the target astigmatic surface s (left) and reconstructed surface (right). The peak-to-valley wavefront error is approximately $.6\mu\text{m}$, or 1 wave.

In practice, applying this wavefront correction requires a 60V peak-to-peak difference relative to the relaxed positions, which reduces the

workable voltage values by 34.1%. This results in a corrected focal workspace of .46 diopters (a sweep range from infinity to 2.16 meters), which suffices for our 50mm, F/3.3 prototype's .313-diopter target focal sweep range.

REFERENCES

1. D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in "Advances in Neural Information Processing Systems," (2009), pp. 1033–1041.
2. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in "ACM Transactions on Graphics (TOG)," , vol. 26 (ACM, 2007), vol. 26, p. 70.
3. D. Miao, O. Cossairt, and S. K. Nayar, "Focal sweep videography with deformable optics," in "Computational Photography (ICCP), 2013 IEEE International Conference on," (IEEE, 2013), pp. 1–8.