# Time-Constrained Photography
Supplementary material

Samuel W. Hasinoff Kiriakos N. Kutulakos
Frédo Durand William T. Freeman

# A Implementation details

## A.1 Camera models

**Standard camera**. We adopt a classic defocus model for the standard camera [10] that also accounts for diffraction. For a given DOF sub-interval $[d_1, d_2]$ we focus at $\frac{2d_1 d_2}{d_1+d_2}$, to equalize defocus blur at the endpoints.

**Coded aperture camera** [13, 22]. We simulated a $7\times7$ mask blocking half the light, optimized using the objective in [22]. We focus the coded aperture camera in the same way as the standard camera, to equalize defocus blur at DOF interval endpoints.

**Focus sweep** [9, 19]. We computed the OTF numerically, by integrating the PSF we use for the standard camera [10] over focus setting.

**Wavefront coding camera** [3]. As discussed in [14], wavefront coding corresponds to a 2D parabolic integration surface. We computed this exactly for a square aperture, accounting for boundary effects [16].

**Lattice-focal camera** [15]. We used an idealized analytic formula for the lattice-focal camera [15], ignoring discretization effects; in practice the DOF may not be evenly spanned by an integer square number of lens subsquares. The formula also assumes that the spectrum of a given lens instance matches well its expectation over a random selection of subsquare focal lengths.

## A.2 Target DOF Expansion

For all the cameras above, with the exception of the lattice-focal camera, we expanded the target DOF slightly to reduce variations at DOF endpoints. This expansion, while lowering average performance, is important for improving the worst-case SNR. In particular, the OTF magnitude at the endpoints of the DOF can be reduced by up to 50% [16]. Handling this effect is especially important for cameras that are approximately depth-invariant, such as focus sweep, because almost all of their OTF variation occurs at the DOF endpoints.

For the standard and coded aperture cameras we used a simple heuristic to expand the target DOF range: we choose expanded limits for the target DOF such that by focusing at each of those new endpoints, the resulting DOF predicted by geometric optics [8], assuming a circle of confusion of 1 pixel, falls just outside and adjacent to the original target DOF.

For the focus sweep camera, we optimized the amount of DOF expansion by exhaustive search, and found that 10-20% expansion was generally optimal. Similarly, for the wavefront coding camera, we optimized the curvature of the 2D parabolic integration surface by exhaustive search, per target DOF.

# B Image Restoration Method

We present a more detailed description our Bayesian image restoration and depth-from-defocus method and related error analysis, and also explain the steps we used to derive it.

## B.1 Image formation model

For completeness, we first review the image formation model in more detail.

**Blurred image formation with noise.** We collect a set of $N$ images $\{\mathbf{y}^1, \ldots, \mathbf{y}^N\}$. Our image formation model involves convolving the ideal image with an appropriate PSF $\boldsymbol{\phi}_d^{\mathcal{D}_k}$ representing the defocus, scaling it by the relative exposure level $\tau$, and corrupting it with random sensor noise $\mathbf{n}$:

$$\mathbf{y}^k \;=\; \tau\, \boldsymbol{\phi}_d^{\mathcal{D}_k} \otimes \mathbf{x} \;+\; \mathbf{n} \;\; . \tag{B.1}$$

where $d$ is the depth of the scene, and $\mathcal{D}_k$ is the DOF for which the $k$-th image is adjusted. We assume sensor noise to be Gaussian and constant over the image,

$$n(p) \;\sim\; \mathcal{N}\big(\,0\,,\,\eta^2\,\big) \;\; , \tag{B.2}$$

where $p$ is a pixel index, and the per-pixel noise variance, $\eta^2$, is a function of the relative exposure level $\tau$ (see Sec. 2).

**Scene model—depth.** Our model describes a scene that can be treated locally as a textured fronto-parallel patch, with random depth $d$. In the absence of prior information, we assume that scene depth is randomly sampled form within the target DOF, $\mathcal{D} = [d_1, d_2]$.

In particular, we assume that depth is drawn uniformly from the DOF on the *image-side*, according to the thin lens model [8]. In other words, for a lens with focal length $f$, we assume that the sensor-lens distance corresponding to the scene depth, $v = (1/f - 1/d)^{-1}$, is drawn uniformly from the range $[v_2, v_1]$, whose endpoints map analogously to the endpoints of the DOF.

**Scene model—Natural image prior.** We assume that the underlying in-focus image has texture following a gradient-penalizing Gaussian prior [13]:

$$\mathrm{vec}(\mathbf{x}) \;\sim\; \mathcal{N}\big(\mathbf{0}\,,\,\mathbf{s}\big) \;\; , \tag{B.3}$$

where vec$(\cdot)$ flattens the image to a 1D column vector. We can express the prior's inverse covariance explicitly as

$$\mathbf{s}^{-1} = \alpha(\mathbf{C}_{\mathbf{g}_x}^T \mathbf{C}_{\mathbf{g}_x} + \mathbf{C}_{\mathbf{g}_y}^T \mathbf{C}_{\mathbf{g}_y}) \ \ , \tag{B.4}$$

where $\mathbf{C}_\mathbf{h}$ denotes the banded matrix corresponding to convolution with filter $\mathbf{h}$; the filters $\mathbf{g}_x, \mathbf{g}_y$ take gradients in the $x, y$ spatial dimensions respectively; and $\alpha$ is a parameter fit to natural images [13].

**Frequency domain representation.** In the frequency domain, the optical transfer functions (OTFs), $\boldsymbol{\Phi}_d^{\mathcal{D}_k}$, correspond to the Fourier transforms of the PSFs. This representation lets us express compactly Eqs. (B.1)–(B.3) over all $N$ input photos, for each spatial frequency $\omega$ independently:

$$\Pr(\,\mathbf{Y}_\omega \,|\, X_\omega, d\,) \ = \ \mathcal{N}\big(\,\tau \boldsymbol{\Phi}_{d\omega} X_\omega \,,\, \eta^2 \mathbf{I}_N\,\big) \tag{B.5}$$

$$\Pr(\,X_\omega\,) \ = \ \mathcal{N}\big(\,0\,,\, S_\omega\,\big) \ \ , \tag{B.6}$$

where upper-case denotes the frequency domain; $\mathbf{I}$ is the identity matrix; the vector $\mathbf{Y}_\omega = [\,Y^1(\omega)\,\ldots\,Y^N(\omega)\,]^\mathrm{T}$ collects observations at frequency $\omega$ across the $N$ photos; the vector $\boldsymbol{\Phi}_{d\omega} = [\,\Phi_d^{\mathcal{D}_1}(\omega)\,\ldots\,\Phi_d^{\mathcal{D}_N}(\omega)\,]^\mathrm{T}$ collects coefficients of the OTFs corresponding to the input photos with depths of field $\mathcal{D}_1, \ldots, \mathcal{D}_N$; and $S_\omega = \big[\alpha(|G_x(\omega)|^2 + |G_y(\omega)|^2)\big]^{-1}$ is the per-frequency variance from the texture prior, where $\mathbf{G}_x, \mathbf{G}_y$ are the Fourier transforms of the gradient filters $\mathbf{g}_x, \mathbf{g}_y$.

## B.2 Image Restoration and Depth-from-Defocus

**Restoration with known depth.** When depth of the scene is known, the OTF vector $\boldsymbol{\Phi}_{d\omega}$ is fully specified by the camera design; restoration is equivalent to $N$-photo Wiener deconvolution.

In our Bayesian treatment, we compute the restored all-in-focus image as the maximum a posteriori (MAP) estimate, $\hat{X}_{d\omega}$. Given our image formation model, the posterior is Gaussian as well, and can be computed in closed form [2]. The MAP estimate then corresponds to the peak of this posterior:

$$\Pr(\,X_\omega \,|\, \mathbf{Y}_\omega, d\,) = \mathcal{N}\big(\,\underbrace{\tfrac{1}{\eta^2}(\tau \boldsymbol{\Phi}_{d\omega}{}^{*}\, \mathbf{Y}_\omega) V_{d\omega}}_{\text{MAP estimate, } \hat{X}_{d\omega}}\,,\, V_{d\omega}\,\big), \tag{B.7}$$

where $^*$ denotes the complex conjugate transpose and the variance of the MAP estimate is

$$V_{d\omega} \ = \ \left(\tfrac{1}{\eta^2}\|\tau \boldsymbol{\Phi}_{d\omega}\|^2 + \tfrac{1}{S_\omega}\right)^{-1} \ \ . \tag{B.8}$$

**Depth estimation.** Often we must estimate the unknown depth $d$ from the observations to specify the effective OTFs $\boldsymbol{\Phi}_{d\omega}$. To this end, we can evaluate the "evidence" for the observations $\mathbf{Y}_\omega$ over different depth hypotheses. Again, since our model involves Gaussians only, the probability distribution for this evidence can be computed in closed form [2]:

$$\Pr(\,\mathbf{Y}_\omega\,|d\,) = \mathcal{N}\big(\,0\,,\,\eta^2\mathbf{I}_N + S_\omega \mathbf{\Phi}_{d\omega}\mathbf{\Phi}_{d\omega}{}^*\,\big) \quad . \tag{B.9}$$

This expression marginalizes over all scenes $X_\omega$, and avoids degenerate models by accounting for the "volume" of scenes that explain the observations.

To determine the marginal probability of our observations given a particular depth hypothesis, we evaluate the Gaussian defined by Eq. (B.9) using $\mathbf{Y}_\omega$. Fortunately, by the matrix inversion lemma and the matrix determinant lemma, we can express this in closed form:

$$\begin{aligned}\log\Pr(\,\mathbf{Y}_\omega\,|d\,) \;=\; \text{const} - \tfrac{1}{2}\Big[\,&\log\big(\eta^2 + S_\omega\|\tau\mathbf{\Phi}_{d\omega}\|^2\big) + \\ &\tfrac{1}{\eta^2}\|\mathbf{Y}_\omega\|^2 \;-\; \tfrac{1}{\eta^4}|\mathbf{\Phi}_{d\omega}{}^*\mathbf{Y}_\omega|^2\,V_{d\omega}\,\Big] \quad .\end{aligned} \tag{B.10}$$

This formula accounts for all joint information encoded in the $N$ photos [21, 17].

It is also particularly useful to reformulate Eq. (B.10) in the spatial domain, because this helps us describe spatially-varying scene depth. By straightforward algebraic manipulation of Eq. (B.10), involving the definitions in Eqs. (B.1)–(B.8), we obtain a 2D "image" representing the per-pixel evidence for a given depth hypothesis:

$$\begin{aligned}\log\Pr(\{\mathbf{y}^k\}\,|d\,) \;=\; \text{const} - \tfrac{1}{2}\Big[&\overbrace{\textstyle\sum_{k=1}^N(\mathbf{y}^k - \tau\,\boldsymbol{\phi}_d^{\mathcal{D}_k}\otimes\hat{\mathbf{x}}_d)^2}^{\text{restoration error}} + \\ &\underbrace{\text{unvec}\,\big(\text{vec}(\hat{\mathbf{x}}_d)^{\mathrm{T}}\mathbf{s}^{-1}\text{vec}(\hat{\mathbf{x}}_d)\big)}_{\text{prior term}} \;+\; \underbrace{\tfrac{1}{P}\textstyle\sum_\omega\log\big(\eta^2 S_\omega V_{d\omega}^{-1}\big)}_{\text{model complexity}}\Big],\end{aligned} \tag{B.11}$$

where the square in the term is pixel-wise; $P$ is the number of pixels in the image; unvec$(\cdot)$ returns the 1D flattened image to 2D form; and $\hat{\mathbf{x}}_d$ is the MAP estimate in the spatial domain, assuming depth $d$ for the whole image. Using Parseval's rule we can verify that $\sum_{x,\,y}\log\Pr(\{\mathbf{y}^k\}\,|d\,) = \sum_\omega\log\Pr(\,\mathbf{Y}_\omega\,|d\,)$.

The terms labeled in Eq. (B.11) show how our Bayesian approach to depth estimation goes beyond evaluating which depth hypothesis minimizes reconstruction error (*i.e.*, computing the maximum-likelihood estimate). What distinguishes our approach from classic methods are the terms for the prior model of the scene and an Occam factor penalizing OTFs with more complex structure. This last term enables exact model selection, which is often handled with such approximations as the Bayesian Information Criterion (BIC) [2].

To estimate the per-pixel probability that the scene is a particular depth we can use Eq. (B.11) and apply Bayes' rule to compute:

$$\Pr(d\,|\,\{\mathbf{y}^k\}\,) \;=\; \frac{\prod_\omega\Pr(\,\{\mathbf{y}^k\}\,|d\,)}{\sum_{d'}\Big[\,\prod_\omega\Pr(\,\{\mathbf{y}^k\}\,|d'\,)\,\Big]} \quad , \tag{B.12}$$

where all products and sums are pixel-wise, and $\Pr(d)$ has been factored out because its distribution is uniform.

This last formula provides us with all the tools we need to specify our Bayesian reconstruction algorithm. On a greedy per-pixel basis, we can apply Eq. (B.12) directly to obtain $d^* = \arg\max_d \Pr(d \,|\, \{\mathbf{y}^k\})$, and create a composite restored image by copying pixels from the corresponding MAP restorations over depth.

In practice, we use a Markov random field (MRF) [1] instead, to regularize this estimate and favor piecewise-smoothness (Fig. 5). For our experiments we used graph cuts with the expansion move [12]; we specified the "data term" as $-\log \Pr(d \,|\, \{\mathbf{y}^k\})$; we specified the "smoothness term" as $8.0 \cdot \min\{|\ell_1 - \ell_2|, 5\}$, for labels pairs $(\ell_1, \ell_2)$ defined by 4-neighborhoods on the pixel grid; and we defined the labels according to a discretization of the DOF into 200 depths (distributed evenly on the image-side).

## B.3 Error Analysis

**Expected error with known depth.** The expected squared error of the MAP reconstruction takes on a well-known form when depth is known [2]:

$$\mathrm{E}\Big[|X_\omega - \hat{X}_{d\omega}|^2\Big] \;=\; V_{d\omega} \quad , \tag{B.13}$$

in expectation over noise and the image prior. To relate this result to squared error in the spatial domain, we sum over frequency and apply Parseval's rule to obtain $\mathrm{E}\big[\|\mathrm{vec}(\mathbf{x} - \hat{\mathbf{x}}_d)\|^2\big] \;=\; \sum_\omega V_{d\omega}$.

**Unknown depth.** More interestingly, we can also derive the expected squared error for a MAP reconstruction given *incorrectly* estimated OTFs. If the true scene depth is $d$ but its estimated value is $\hat{d}$, we can show that the expected squared error is

$$\mathrm{E}\Big[|X_{d\omega} - \hat{X}_{\hat{d}\omega}|^2\Big] \;=\; \underbrace{\Big(\tau^2 \tfrac{S_\omega}{\eta^2} \|\varDelta\|^2 - (\varDelta + \varDelta^*)\Big) \tfrac{\tau^2}{\eta^2} {V_{\hat{d}\omega}}^2}_{\text{depth estimation error}} \;+\; \underbrace{V_{\hat{d}\omega}}_{\text{known-depth error}} \quad , \tag{B.14}$$

in expectation over noise and the prior, where the scalar $\varDelta = {\mathbf{\Phi}_{\hat{d}\omega}}^* (\mathbf{\Phi}_{d\omega} - \mathbf{\Phi}_{\hat{d}\omega})$ measures discrepancy between the true and estimated OTFs, and the notation $X_{d\omega}$ emphasizes that the true depth of the scene $X_\omega$ is $d$. This follows from straightforward, if tedious, algebra.

**Error analysis for an imaging system overall.** To estimate the *overall* performance of an imaging system we must consider performance not only over the distribution of underlying images (Eq. (B.6)), but also over the distribution of our depth estimates.

Suppose that the true depth of the scene is $d$. Under the assumption that our depth uncertainty, $\Pr(\,\hat{d}\,|d\,)$, is independent of scene texture, we obtain

$$\mathrm{E}\big[\|\mathbf{x}_d) - \mathrm{vec}(\hat{\mathbf{x}}_{\hat{d}}\|^2\big] \;=\; \sum_{\omega} \int_{\hat{d}} \Pr(\,\hat{d}\,|d\,)\,\mathrm{E}\big[|X_{d\omega} - \hat{X}_{\hat{d}\omega}|^2\big]\,\mathrm{d}\hat{d} \;\; . \tag{B.15}$$

In practice, we compute this expectation using Monte Carlo sampling, evaluating the probability of obtaining different depth estimates $\hat{d}$ using Eq. (B.12), for a collection of samples for the scene $\mathbf{x}$ and sensor noise at depth $d$.

The reason why we can assume that depth uncertainty is independent of scene texture follows from information theory. Since most scenes $\mathbf{x}$ constitute "typical sequences" in a formal sense [5], for a sufficiently large image patch, the law of large numbers ensures that drawing even a single sample leads to a tight $(\delta, \epsilon)$-bound on its estimate of expected error; therefore $\Pr(\,\hat{d}\,|d\,)$ and $\mathrm{E}\big[|X_{d\omega} - \hat{X}_{\hat{d}\omega}|^2\big]$ can be thought of as decoupled. Empirically, we found that very few samples for the scene texture (*i.e.,* less than 20) were indeed needed for the approximation in Eq. (B.15) to be accurate.

# C Derivation of Upper Bound in Eq. (12)

We closely follow the treatment in [15], extending the derivation to handle circular apertures and incorporating our novel upper bound, namely the final term of Eq. (12). We first review basic terminologies from 4D light fields in the Fourier domain along the lines of [18, 20]. We then prove an upper bound on the magnitude of the defocus kernel for a given spatial frequency (Lemma 1). Then we show that the worst-case upper bound involves evenly spreading the energy available over a lower-dimensional 3D manifold (Lemma 2). Finally, we show another bound on all coefficients of the defocus kernel (Lemma 3), and combine these results to obtain Eq. (12).

## C.1 Frequency analysis of depth of field in 4D

Our analysis is based on geometric optics and the light field. We first review how the light field can be used to analyze cameras [18, 20]. It is a 4D function $\ell(x, y, u, v)$ describing radiance for all rays in a scene, where a ray is parameterized by its intersections with two parallel planes, the $uv$-plane and the $xy$-plane [18]. We assume the camera aperture is positioned on the $uv$-plane, and $xy$ is a plane in the scene. $x, y$ are spatial coordinates and the $u, v$ coordinates denote the viewpoint direction.

An important property is that the light rays emerging from a given physical point correspond to a 2D plane in 4D of the form

$$x = su + (1-s)p_x, \quad y = sv + (1-s)p_y \;, \tag{C.16}$$

whose slope $s$ encodes the object's depth:

$$s = (d - d_o)/d \;, \tag{C.17}$$

where $d$ is the object depth; $d_o$ the distance between the $uv$, $xy$ planes [4, 11, 7]. The offsets $p_x$ and $p_y$ characterize the location of the scene point within the plane at depth $d$.

Each sensor element gathers light over its 2D area and the 2D aperture. This is a 4D integral over a set of rays, and under first order optics (paraxial optics), it can be modeled as a convolution [20, 18]. A shift-invariant kernel $k(x, y, u, v)$ determines which rays are summed for each element, as governed by the lens. Before applying imaging noise, the value recorded at a sensor element is then:

$$\tilde{\mathbf{y}}(x_0, y_0) = \iiiint k(x_0 - x, y_0 - y, -u, -v)\, \ell(x, y, u, v)\, dxdydudv\ . \qquad \text{(C.18)}$$

If the local scene depth, or slope, is known, the noise-free defocused image $\tilde{\mathbf{y}}$ can be expressed as a convolution of an ideal sharp image $\mathbf{x}$ with a PSF $\boldsymbol{\phi}_s$: $\tilde{\mathbf{y}} = \boldsymbol{\phi}_s \otimes \mathbf{x}$. For a given slope $s$, this PSF is fully determined by projecting the 4D lens kernel $k$ along the direction $s$:

$$\boldsymbol{\phi}_s(x, y) = \iint k(x, y, u + sx, v + sy) dudv\ . \qquad \text{(C.19)}$$

That is, we simply integrate over all rays $(x, y, u + sx, v + sy)$ corresponding to a given point in the $xy$-plane (see Eq. C.16).

Now that we have expressed defocus as a convolution, we can analyze it in the frequency domain. Let $K(\omega_x, \omega_y, \omega_u, \omega_v)$ denote the 4D lens spectrum, the Fourier transform of the 4D lens kernel $k(x, y, u, v)$. As the PSF $\boldsymbol{\phi}_s$ is obtained from $k$ by projection (Eq. (C.19)), by the Fourier slice theorem, the OTF $\boldsymbol{\Phi}_s$ is a slice of the 4D lens spectrum $K$ in the orthogonal direction [20, 16]:

$$\varPhi_s(\omega_x, \omega_y) = K(\omega_x, \omega_y, -s\omega_x, -s\omega_y)\ . \qquad \text{(C.20)}$$

Below we refer to slices of this form as *OTF-slices*, because they directly provide the OTF-the frequency response due to defocus at a given depth. These are slanted slices whose slope is orthogonal to the object slope in the primal light field domain. Low spectrum values in $K$ leads to low magnitudes in the OTF for the corresponding depth.

**The dimensionality gap.** As described above, scene depth corresponds to slope $s$ in the light field. It has, however, been observed that the 4D light field has a *dimensionality gap*, in that most slopes do not correspond to a physical depth [7, 20]. Indeed, the set of all 2D planes $x = s_u u + p_x$, $y = s_v v + p_y$ described by their slope $s_u, s_v$ and offset $p_x, p_y$ is 4D. In contrast, the set corresponding to real depth, i.e. where $s = s_u = s_v$, is only 3D, as described by Eq. (C.16).

The dimensionality gap is particularly clear in the Fourier domain [20]. Consider the 4D lens spectrum $K$, and examine the 2D slices $K_{\omega_{x_0,y_0}}(\omega_u, \omega_v)$, in which the spatial frequencies $\omega_{x_0}, \omega_{y_0}$ are held constant. We call these $\omega_{x_0,y_0}$-*slices*. In flatland, $\omega_{x_0,y_0}$-slices are vertical slices. Following Eq. (C.20), we note that the set of entries in each $K_{\omega_{x_0,y_0}}$ participating in the OTF for any depth is restricted to a 1D line:

$$K_{\omega_{x_0,y_0}}(-s\omega_{x_0}, -s\omega_{y_0}) \ , \tag{C.21}$$

for which $\omega_u = -s\omega_{x_0}, \ \omega_v = -s\omega_{y_0}$. For a fixed slope range (DOF) the set of entries participating in any OTF $\mathbf{\Phi}_s$ is a 1D segment.

**Relation between slope and defocus.** We seek to capture a fixed depth range $[d_1, d_2]$. To simplify the light field parameterization, we select the location of the $xy$ plane according to the harmonic mean $d_o = \frac{2d_1d_2}{d_1+d_2}$, corresponding to the point at which one would focus a standard lens to equalize defocus diameter at both ends of the depth range [8]. This maps the depth range to the symmetric slope range $[-S/2, S/2]$, where $S = \frac{2(d_2-d_1)}{d_2+d_1}$ (Eq. (C.17)). Under this parameterization the defocus diameter corresponding to a given slope $s$, as measured on the $xy$ plane in the scene, can be expressed simply as $A|s|$.

## C.2 Energy bound on 4D spectrum

The first step in our derivation is to show that the energy in a 4D lens spectrum is bounded. We derive the available energy budget using a direct extension of the 1D case [6, 16].

**Lemma 1** *For a circular aperture with diameter $A$, the total energy in each $\omega_{x_0,y_0}$-slice is bounded by $\frac{\pi}{4}A^2$:*

$$\iint |K_{\omega_{x_0,y_0}}(\omega_u, \omega_v)|^2 d\omega_u d\omega_v \leq \frac{\pi}{4}A^2 \ . \tag{C.22}$$

*Proof:* This is similar to the budget proof in [16]. Basically $K_{\omega_{x_0,y_0}}(\omega_u, \omega_v)$ is the 2D Fourier transform of

$$\iint k(x,y,u,v)e^{-2i\pi(\omega_{x_0}x+\omega_{y_0}y)}dxdy \ . \tag{C.23}$$

Since the amount of energy that can pass via an aperture area in a given integration time is bounded, the norm of each element is bounded by

$$\left|\iint k(x,y,u,v)e^{-2i\pi(\omega_{x_0}x+\omega_{y_0}y)}dxdy\right|^2 \leq 1 \ . \tag{C.24}$$

Since the aperture size is $\frac{\pi}{4}A^2$ (that is, the above integral is non-zero for an area of $\frac{\pi}{4}A^2$ only) we get that the total norm is bounded by $\frac{\pi}{4}A^2$:

$$\iint \left|\iint k(x,y,u,v)e^{-2i\pi(\omega_{x_0}x+\omega_{y_0}y)}dxdy\right|^2 dudv \leq \frac{\pi}{4}A^2 \ . \tag{C.25}$$

By Parseval's theorem, the square integral is the same in the dual and the primal domains, thus:

$$\iint |K_{\omega_{x_0,y_0}}(\omega_u, \omega_v)|^2 d\omega_u d\omega_v \leq \frac{\pi}{4}A^2 \ . \tag{C.26}$$

□

## C.3 Worst-case MTF over the 3D focal manifold

In this section we derive a bound on the magnitude of the OTF for defocus, also known as the modulation transfer function (MTF). In particular we seek to maximize the MTF $|\Phi_s(\omega_{x,y})|$ in the worst case, over all slopes $s \in [-S/2, S/2]$ and over all spatial frequencies $\omega_{x,y}$. Since the OTFs are slices from the 4D lens spectrum $K$ (Eq. (C.20)), this is equivalent to maximizing the spectrum on the focal segments of $K$. As in the 1D space-time case [16], optimal worst-case performance can be realized by spreading the energy budget uniformly over the range of slopes.

Given a power budget for each $\omega_{x_0,y_0}$-slice, the upper bound for the defocus MTF concentrates this budget on the 1D focal segment only. Distributing energy over the focal manifold requires caution, however, because the segment effectively has non-zero thickness due to its finite support in the primal domain. If a 1D focal segment had zero thickness, its spectrum values could be made infinite while still obeying the norm constraints of Lemma 1. As we show below, since the primal support of $k$ is finite ($k$ admits no light outside the aperture), the spectrum must be finite as well, so the 1D focal segment must have non-zero thickness.

**Lemma 2** *For a circular aperture with diameter $A$, the worst-case defocus MTF for the range $[-S/2, S/2]$ is bounded. For every spatial frequency $\omega_{x,y}$:*

$$\min_{s\in[-S/2,S/2]} |\Phi_s(\omega_x, \omega_y)|^2 \le \frac{\frac{2}{3}\beta(\omega_{x,y})A^3}{S\|\omega_{x,y}\|} \ , \tag{C.27}$$

*where the factor*

$$\beta(\omega_{x,y}) = \frac{\|\omega_{x,y}\|}{\max(|\omega_x|, |\omega_y|)}\left(1 - \frac{\min(|\omega_x|, |\omega_y|)}{3\cdot\max(|\omega_x|, |\omega_y|)}\right) \tag{C.28}$$

*is in the range $[\frac{5\sqrt{5}}{12}, 1] \approx [0.93, 1]$.*

*Proof:* For each $\omega_{x_0,y_0}$-slice $K_{\omega_{x_0,y_0}}$ the 1D focal segment is of length $S\|\omega_{x_0,y_0}\|$. We first show that the focal segment norm is bounded by $\frac{2}{3}A^3$, and then the worst-case optimal strategy is to spread the budget evenly over the segment.

To simplify notations, we consider the case $\omega_{y_0} = 0$ since the general proof is similar after a basis change. For this case, the 1D focal segment is a horizontal line of the form $K_{\omega_{x_0,y_0}}(\omega_u, 0)$. For a fixed value of $\omega_{x_0}$, this line is the Fourier transform of:

$$\iiint k(x, y, u, v)e^{-2i\pi(\omega_{x_0}x+0y+0v)}dxdydv \ . \tag{C.29}$$

By showing that the total power of Eq. (C.29) is bounded by $A^3$, Parseval's theorem gives us the same bound for the focal segment.

Since the exposure time is assumed to be 1, we collect unit energy through every $u, v$ point lying within the clear aperture[1]:

$$\iint k(x, y, u, v) dx dy = \begin{cases} 1 & u^2 + v^2 \leq \frac{1}{4}A^2 \\ 0 & \text{otherwise} \end{cases} \quad . \tag{C.30}$$

A phase change to the integral in Eq. (C.30) does not increase its magnitude, therefore, for every spatial frequency $\omega_{x_0, y_0}$,

$$\left| \iint k(x, y, u, v) e^{-2i\pi(\omega_{x_0} x + \omega_{y_0} y)} dx dy \right| \leq 1 \ . \tag{C.31}$$

Using Eq. (C.31) and the fact that the circular aperture is width $\sqrt{A^2 - 4u^2}$ along on the $v$-axis, we obtain:

$$\left| \iiint k(x, y, u, v) e^{-2i\pi\omega_{x_0} x + 0y + 0v} dx dy dv \right|^2 \leq A^2 - 4u^2 \ . \tag{C.32}$$

On the $u$-axis, the aperture has width $A$, corresponding to its diameter. By integrating Eq. (C.32) over $u$ we see the power is bounded by $\frac{2}{3}A^3$:

$$\int \left| \iiint k(x, y, u, v) e^{-2i\pi(\omega_{x_0} x + \omega_{y_0} y)} dx dy dv \right|^2 du \leq \frac{2}{3}A^3 \ . \tag{C.33}$$

Since the left-hand side of Eq. (C.32) is the power spectrum of $K_{\omega_{x_0, y_0}}(\omega_u, 0)$, by applying Parseval's theorem we see that the total power over the focal segment is bounded by $\frac{2}{3}A^3$ as well:

$$\int |K_{\omega_{x_0, y_0}}(\omega_u, 0)|^2 d\omega_u \leq \frac{2}{3}A^3 \tag{C.34}$$

Since the focal segment norm is bounded by $\frac{2}{3}A^3$, and since we aim to maximize the worst-case magnitude, the best we can do is to split the budget uniformly over the length $S\|\omega_{x_0, y_0}\|$ focal segment, which bounds the worst MTF power by $\frac{2}{3}A^3 / S\|\omega_{x_0, y_0}\|$. In the general case, Eq. (C.33) is bounded by $\frac{2}{3}\beta(\omega_{x,y})A^3$ rather than $\frac{2}{3}A^3$, and Eq. (C.27) follows. $\square$

## C.4 Bound on coefficients of the MTF

We identify one more bound on the MTF, which comes from applying another conservation argument over the *full* domain to all the coefficients in the MTF.

**Lemma 3** *For a circular aperture with diameter $A$, every coefficient of the MTF is bounded by $\frac{\pi}{4}A^2$:*

$$|\Phi_s(\omega_x, \omega_y)| \ \leq \ \frac{\pi}{4}A^2 \quad . \tag{C.35}$$

[1] If an amplitude mask is placed at the aperture (e.g. coded aperture) the energy will be reduced and the upper bound still holds.

*Proof:* It is clear that the DC component, in particular, is bounded by the support in the primal domain, $\frac{\pi}{4}A^2$. However, other frequencies share this bound as well. Following from the definition of the 4D Fourier transform we can show:

$$
\begin{aligned}
|\Phi_s(\omega_x,\omega_y)| &= |K(\omega_x,\omega_y,-s\omega_x,-s\omega_y)| && \text{(C.36)}\\
&= \left|\iiiint k(x,y,u,v)e^{-2i\pi(\omega_x x+\omega_y y-s\omega_x u-s\omega_y v)}dxdydudv\right| && \text{(C.37)}\\
&\le \iiiint \left|k(x,y,u,v)e^{-2i\pi(\omega_x x+\omega_y y-s\omega_x u-s\omega_y v)}\right| dxdydudv && \text{(C.38)}\\
&= \iiiint |k(x,y,u,v)|dxdydudv && \text{(C.39)}\\
&\le \frac{\pi}{4}A^2 \quad . && \text{(C.40)}
\end{aligned}
$$

The derivation is elementary, relying only on the triangle inequality applied to the modulus function, combined with the fact that the kernel is real in the spatial domain. □

## C.5 Upper bound: Proof of Eq. (12)

*Proof:* By combining Lemmas 2-3, we obtain:

$$
\begin{aligned}
\min_{s\in[-S/2,S/2]} |\Phi_s(\omega_x,\omega_y)|^2 &\le A^4\cdot\min\left\{\frac{\frac{2}{3}\,\beta(\omega_{x,y})}{AS\|\omega_{x,y}\|}\ ,\ \frac{\pi^2}{16}\right\} && \text{(C.41)}\\
&= A^4\cdot\min\left\{\frac{\beta(\omega_{x,y})}{3\|\omega_{x,y}\|b_{\max}}\ ,\ \frac{\pi^2}{16}\right\} && \text{(C.42)}
\end{aligned}
$$

where $b_{\max} = A\frac{S}{2}$ is the blur diameter at the endpoints of the DOF, $[d_1, d_2]$. This reproduces the upper bound in Eq. (12), for the case of $N = 1$. □

Note by further analyzing Eq. (C.42) we can show there is no benefit to dividing the time budget for the worst-case upper bound, *i.e.*, that $N^* = 1$. As in Sec. B.1, define the vector $\mathbf{\Phi}_{s\,\omega_{x,y}} = \frac{1}{N}[\,\Phi_s^{\mathcal{D}_1}(\omega_{x,y})\ \ldots\ \Phi_s^{\mathcal{D}_N}(\omega_{x,y})\,]^{\mathrm{T}}$ corresponding to the OTF coefficients for the $N$ photos, for a scene whose depth corresponds to a slope of $s$. Dividing the time budget in this way reduces the power for each observation by $1/N^2$. However, since each of the $N$ photos is responsible for $1/N$ of the DOF, the maximum blur diameter is reduced to $b/N$. Putting this together, we obtain

$$
\begin{aligned}
\min_{s\in[-S/2,S/2]} \|\mathbf{\Phi}_{s\,\omega_{x,y}}\|^2 &\le N\cdot\left[\frac{A^4}{N^2}\cdot\min\left\{\frac{1}{3\|\omega_{x,y}\|(b_{\max}/N)}\ ,\ \frac{\pi^2}{16}\right\}\right] && \text{(C.43)}\\
&\le A^4\cdot\min\left\{\frac{1}{3\|\omega_{x,y}\|b_{\max}}\ ,\ \frac{\pi^2}{16N}\right\} && \text{(C.44)}
\end{aligned}
$$

which shows that the upper bound is indeed highest for $N = 1$.

The same conclusion that $N^* = 1$ also holds if the input photos are not used to divide the DOF, but instead to capture $N$ identical photos, each corresponding to the regular one-shot upper bound with $1/N$ of the time budget.

# References


[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In *SIGGRAPH*, pp. 294–302, 2004.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.

[3] W. T. Cathey and E. R. Dowski. New paradigm for imaging systems. *App. Optics*, 41(29):6080–6092, 2002.

[4] J.-X. Chai, S.-C. Chan, H.-Y. Shum, and X. Tong. Plenoptic sampling. In *SIGGRAPH*, pp. 307–318, 2000.

[5] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, 1991.

[6] A. R. FitzGerrell, j. Edward R. Dowski, and W. T. Cathey. Defocus transfer function for circularly symmetric pupils. *App. Optics*, 36(23):5796–5804, 1997.

[7] X. Gu, S. J. Gortler, and M. F. Cohen. Polyhedral geometry and the two-plane parameterization. In *Rendering Techniques*, pp. 1–12, 1997.

[8] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. In *ECCV*, vol. 4, pp. 45–59, 2008.

[9] G. Häusler. A method to increase the depth of focus by two step image processing. *Optics Communications*, 6(1):38–42, 1972.

[10] H. H. Hopkins. The frequency response of a defocused optical system. *Proc. of the Royal Society of London, Series A*, 231(1184):91–103, 1955.

[11] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *SIGGRAPH*, pp. 297–306, 2000.

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

[13] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*, 2007.

[14] A. Levin, W. T. Freeman, and F. Durand. Understanding camera trade-offs through a Bayesian analysis of light field projections. In *ECCV*, vol. 4, pp. 88–101, 2008.

[15] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4D frequency analysis of computational cameras for depth of field extension. In *SIGGRAPH*, 2009.

[16] A. Levin, P. Sand, T. S. Cho, F. Durand, and W. T. Freeman. Motion-invariant photography. In *SIGGRAPH*, 2008.

[17] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009.

[18] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH*, pp. 31–42, 1996.

[19] H. Nagahara, S. Kuthirummal, C. Zhou, and S. Nayar. Flexible depth of field photography. In *ECCV*, vol. 4, pp. 60–73, 2008.

[20] R. Ng. Fourier slice photography. In *SIGGRAPH*, pp. 735–744, 2005.

[21] A. Rav Acha and S. Peleg. Two motion-blurred images are better than one. *PRL*, 26(3):311–317, 2005.

[22] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *SIGGRAPH*, 2007.