Time-Constrained Photography

Samuel W. Hasinoff [1,2]   Kiriakos N. Kutulakos [2]   Frédo Durand [1]   William T. Freeman [1]

[1] MIT CSAIL   [2] University of Toronto

Abstract

Capturing multiple photos at different focus settings is a powerful approach for reducing optical blur, but how many photos should we capture within a fixed time budget? We develop a framework to analyze optimal capture strategies balancing the tradeoff between defocus and sensor noise, incorporating uncertainty in resolving scene depth. We derive analytic formulas for restoration error and use Monte Carlo integration over depth to derive optimal capture strategies for different camera designs, under a wide range of photographic scenarios. We also derive a new upper bound on how well spatial frequencies can be preserved over the depth of field. Our results show that by capturing the optimal number of photos, a standard camera can achieve performance at the level of more complex computational cameras, in all but the most demanding of cases. We also show that computational cameras, although specifically designed to improve one-shot performance, generally benefit from capturing multiple photos as well.

1. Introduction

Recent years have seen many proposals for tightly integrating sensing, optics and computation in order to extend the capabilities of the traditional camera. Already, numerous "computational camera" designs exist for capturing photos with reduced motion blur [30, 20], post-capture refocusing capabilities [17, 33, 3], and an extended depth of field (DOF) [4, 11, 23, 17, 33, 19]. Although these designs differ in many respects, they all adhere to the principle of one-shot capture: the camera records a single image with a DOF constrained by the optics and an exposure time constrained by the available time budget (or by pixel saturation).

In this paper we show that one-shot capture is usually *not* optimal for extended-DOF photography, *i.e.*, it does not produce a well-focused image with the highest signal-to-noise ratio (SNR) for a desired DOF and time budget. Moreover, we show that this result applies to standard and computational cameras [4, 23, 17, 19] alike: image quality in both cases can often be improved by capturing many shots within a given time budget, rather than just one.

Our analysis is based on a key insight illustrated in Fig. 1: by spreading the time budget across several "underexposed" shots with different focus settings we can obtain reduced worst-case blur, at the expense of higher sensor noise. In particular, read noise leads to a penalty for each photo we capture, but Poisson-distributed photon noise *does not* penalize multiple shots. Since photon noise dominates
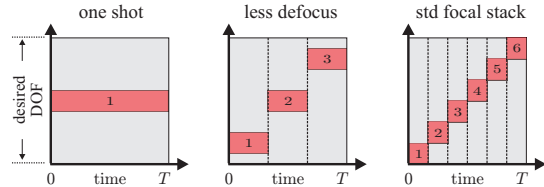

Figure 1. The time-slice advantage for a desired DOF and time budget T. One-shot capture gives the brightest image, but depths far from the lens' DOF (red rectangle) are blurred significantly. A standard focal stack "spans" the desired DOF: photos are exposed less but every depth is effectively blur-free in one of them. In this way, the number of shots acts as a balancing factor between under-exposure and worst-case blur. When photon noise dominates, the optimal-SNR capture strategy tilts even further to the right.

read noise under normal photographic conditions, the overall SNR usually tips in favor of splitting the time budget. We call this the *time-slice advantage*. By contrast, single-shot photography is only optimal for very limited time budgets, *i.e.*, when read noise becomes significant, or for cameras with high per-shot overhead.

In this paper we provide a detailed study of the time-slice advantage and use it for *optimal time-constrained photography*—creating an all-in-focus image with the highest SNR for a given camera design, time budget, target DOF, and average scene brightness. Working from first principles, we formulate all-in-focus photography as a frequency-based restoration problem that takes noisy and optically-blurred photos as input, and outputs a single, all-in-focus image for the target DOF. This leads to three basic questions:

- **camera-specific optimal time allocation:** given a camera's noise model and optical transfer function [29], how should we allocate the time budget to maximize the expected SNR of the all-in-focus image?
- **optics-independent performance bound:** what is the maximum attainable expected SNR for a given sensor across all possible optical transfer functions?
- **camera performance characterization:** how do the existing extended-DOF camera designs compare in terms of their attainable expected SNR, and how do they fare against the traditional camera?

Our answer to these questions can be viewed as complementing and generalizing several lines of recent work.

Closest to our work, Hasinoff and Kutulakos [10] studied a related problem in extended-DOF photography: minimizing the time it takes to capture a given DOF while maintaining ideal exposure. Their work considers multiple photos, but it ignores the effect of noise and uses a basic view of

DOF that applies only to conventional cameras. By contrast, our analysis applies to computational cameras as well, and is specifically designed to consider the tradeoff between defocus blur and noise under constrained time budgets.

Although not applicable to extended DOF, synthetic shutter speed imaging [30] is similar in spirit. This approach used bursts of photos to mitigate motion blur due to camera shake, but did not investigate the tradeoff between blur and under-exposure. In another context, multiple photos were shown to be beneficial compared to using a polarizer for hazy scenes [31]. In this respect, our work can be viewed as identifying a similar multi-photo advantage for extended-DOF photography, with a full theoretical treatment.

Since we take advantage of a new photographic tradeoff, our work is also related to tradeoffs for other degrees of freedom (*e.g.*, changing the aperture [10], the pixel-to-ray mapping [24, 18, 19], or lighting [28]). The time-slice advantage is orthogonal to these studies, does not involve modifying optics, and can lead to further quality improvements.

At the heart of our approach lies a derivation for the expected worst-case SNR of a restored, all-in-focus image given a camera, time budget, DOF, and number of photos. We derive analytic formulas for expected restoration error and characterize our uncertainty in estimating scene depth using Monte Carlo integration. While computing such an expectation is novel, our derivation is related to shape-from-focus [14, 16, 32], and defocus [25, 6, 7], and all-in-focus restoration [26, 2, 9]. Indeed, our derivation includes a Bayesian depth-from-defocus and restoration algorithm that handles any camera design (*e.g.*, [4, 23, 17]) and jointly considers any number of photos.

Our work offers five contributions over the state of the art. First, it introduces the time-slice advantage as a powerful tool for high-SNR imaging over large DOFs. The advantage applies to both conventional and computational cameras, and exploits their noise properties [12, 22] and their increasing capture rates. Second, our framework enables derivation of expected SNR for any given sensor/lens combination under geometric optics. Third, this leads to a novel way of comparing all-in-focus imaging performance: we fix the time budget and estimate SNR with the number of photos optimized for each design. This "levels the playing field" across different optical designs and, we believe, is a more representative performance metric. Fourth, we show that conventional cameras perform very well when placed under this new light—for a broad range of time budgets, they match the performance of the best-performing extended-DOF cameras. Fifth, we confirm the time-slice advantage experimentally, through controlled tests with real scenes and high-end digital cameras.

2. Optimal Time-Constrained Photography

Given a camera, a time budget, and a target DOF, we seek the *optimal time slice*—the number of photos that maximizes all-in-focus performance within that DOF. To achieve
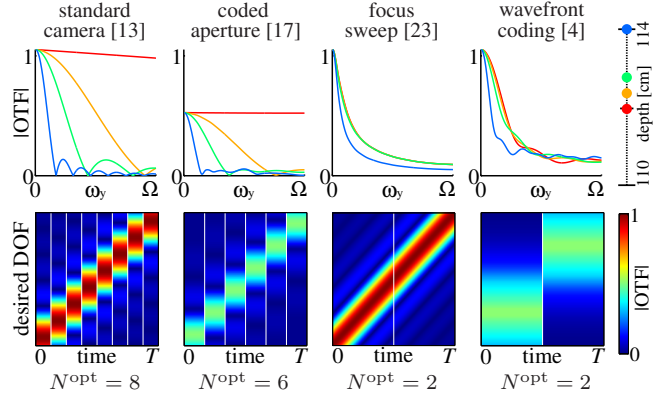

Figure 2. Representing blur and DOFs in the frequency domain. **Top:** OTF magnitude for four color-coded scene depths, for various camera designs. The plots show a 1D frequency slice, $\omega = (\frac{1}{2}\omega_y, \omega_y)$, up to the Nyquist limit, Ω. All plots assume an 85mm focal length, an f/1.2 aperture, and the Canon 1D Mark II sensor. For comparison, the DOF of the standard camera covers just 1/13th of the range $[110, 114]\,\mathrm{cm}$. **Bottom:** OTF magnitude for the same depth range, a time budget of $0.1T^{\mathrm{opt}}$, and each design's optimal time slice. Each plot shows $|\Phi_d^{\mathcal{D}}(\omega)|$ for each of the N^{opt} photos, for $\omega = (0.2, 0.4)\Omega$. The non-zero regions of $|\Phi_d^{\mathcal{D}}(\omega)|$ are analogous to the red rectangles in Fig. 1, but for a specific frequency. When ω changes, these widths change too.

this, we estimate the expected worst-case performance from capturing N photos and then optimize it:

$$N^{\mathrm{opt}} = \arg\max_N \ \mathrm{SNR}(N) \ . \qquad (1)$$

The SNR-based measure in Eq. (1), and the optimization it leads to, forms the core of our approach. We define this measure below in terms of three models: for the lens, for idealized image formation, and for sensor noise.

Lens model. Under first-order geometric optics [29], the optical properties of a lens can be expressed as a convolution that maps a planar, "blur-free" scene at depth d to a blurred image on the sensor plane. The kernel of this convolution, the lens point-spread function $\phi_d^{\mathcal{D}}$, depends on the plane's depth as well as the depth of field, $\mathcal{D} = [d_1, d_2]$, for which the lens is adjusted.

For lenses with complex optical properties, the Fourier transform of the point-spread function $\Phi_d^{\mathcal{D}}$ is convenient because it allows analysis of each spatial frequency independently. This representation, also called the optical transfer function (OTF), describes how the lens attenuates and shifts a given frequency ω on the scene plane (Fig. 2, top).

Ideal image formation. The relation between exposure time and pixel intensity is particularly important in our analysis. To factor out absolute scene brightness, we represent exposure time as a fraction τ of the time required for an ideal exposure. We also call τ a photo's *exposure level*.

Guided by auto-exposure criteria [15], we consider a photo **y** to be ideally exposed if its mean intensity is 13% of the maximum possible value $y_{\max}$. We use T^{opt} to denote the exposure time required to achieve this level.
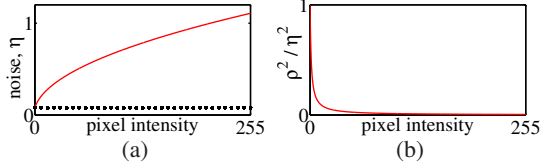
Figure 3. (a) Sensor noise (red) as a function of intensity for the Canon 1D Mark II at ISO 100 [5], for $y_{\max} = 255$. Noise is well-modeled by Poisson-distributed photon arrivals, plus read noise (dashed). (b) Read noise as a fraction of the overall variance.

Note that spreading a time budget T among N photos leads to an exposure level of $\tau = \frac{1}{N}\frac{T}{T^{\mathrm{opt}}}$ for each photo. If $\tau < 1$, these photos will be under-exposed.

In the absence of noise, the above considerations define our image formation model as a scaled 2D convolution:

$$\mathbf{y} \;=\; \tau\,\boldsymbol{\phi}_d^{\mathcal{D}} \otimes \mathbf{x}_d \;, \qquad (2)$$

where $\mathbf{x}_d$ is a planar scene at depth d that is modeled as an ideally-exposed, blur-free image.

Sensor noise model. The benefits of the time slice tradeoff are greatly enhanced by the fact that the dominant source of noise in digital cameras is generally *not* constant read noise (Fig. 3). To capture the essential properties of noise, we use the well-known affine noise model [12, 22, 31], which depends on a photo's exposure level. In particular, we assume that a captured photo follows a normal distribution with a mean given by Eq. (2) and per-pixel variance

$$\eta^2 \;=\; \underbrace{\tau \cdot (0.13 y_{\max})\,/\,g}_{\text{photon noise}} \;+\; \underbrace{\rho^2}_{\text{read noise}} \;. \qquad (3)$$

The first term above represents the Poisson distribution of photon arrivals; it depends linearly on the exposure level, the ideal mean intensity $0.13 y_{\max}$, and an ISO-dependent gain factor $1/g$. The second term represents constant read noise and quantization.

For simplicity, Eq. (3) assumes constant noise variance over the image, computed using the mean image intensity, $\tau(0.13 y_{\max})$. This model is sufficient to determine SNR, since the linearity of affine noise implies that mean pixel variance is a function of mean image intensity.[1,2]

Note that spreading a time budget over N photos incurs *no* penalty from photon noise, whose variance falls off according to $1/N$ as well. This significantly increases the time-slice advantage: for most of the intensity range of real sensors, photon noise dominates over read noise (Fig. 3b); within that range, spreading the budget reduces worst-case blur without a large noise penalty.

Scene-independent restoration error. Our goal is to spread the time budget in a way that enables optimal restora-

tion of any scene inside a given DOF. This requires answering two questions: (1) how do we measure the restoration error from a given set of input photos, and (2) how do we quantify, in a scene-independent way, the performance of a strategy that spreads the budget across N of them?

Equations (2) and (3) model the degradation that occurs in photos of a scene $\mathbf{x}_d$ at depth d. Given an all-in-focus restoration, $\hat{\mathbf{x}}$, of this scene, we use $\|\mathbf{x}_d - \hat{\mathbf{x}}\|^2$ to measure restoration error. To obtain a scene-independent measure of performance, we use the peak signal-to-noise ratio (PSNR) at the worst-case depth:

$$\mathrm{SNR}(N) = \frac{P \cdot y_{\max}^2}{\max_d \mathrm{E}[\,\|\mathbf{x}_d - \hat{\mathbf{x}}\|^2\,]} \;, \qquad (4)$$

where P is the number of pixels per image. This measure is inversely proportional to the expected error over all possible scenes at the worst-case depth, which we quantify below.

3. The Expected Restoration Error

A key step in our approach is to quantify the error for a given camera design if we spread the time budget over N photos. We achieve this by applying a Bayesian analysis in the frequency domain, which allows us to characterize image restoration in detail. This analysis can be also viewed as a general depth-from-defocus [25, 7] and restoration [9] algorithm that handles any camera design and jointly handles any number of photos (including $N=1$).

Simplified scene model. To make our analysis tractable, we model the scene as a textured, fronto-parallel plane whose depth d is uniformly distributed in the target DOF. We assume its texture follows a gradient-penalizing Gaussian prior [17]: $\Pr(\mathbf{x}_d) \propto \prod_p \exp(-\frac{1}{2}\alpha[(\mathbf{x}_d \otimes \mathbf{g}_x)^2 + (\mathbf{x}_d \otimes \mathbf{g}_y)^2])$, where $\mathbf{g}_x, \mathbf{g}_y$ are gradient filters for the x, y spatial dimensions, α is a parameter fit to natural images, p indexes over pixels, and the squaring is element-wise.

Image formation in the frequency domain. Since the observed photos, $\mathbf{y}^1, \ldots, \mathbf{y}^N$, are modeled by convolutions and Gaussians, the analysis simplifies greatly in the frequency domain. In particular, this allows us to express compactly Eqs. (2) and (3) over all N input photos, for each spatial frequency ω independently:

$$\Pr(\mathbf{Y}_\omega \,|\, X_\omega, d) \;=\; \mathcal{N}\big(\tau\boldsymbol{\Phi}_{d\omega} X_\omega\,,\,\eta^2 \mathbf{I}_N\big) \qquad (5)$$
$$\Pr(X_\omega) \;=\; \mathcal{N}\big(0\,,\,S_\omega\big)\,, \qquad (6)$$

where X_ω is the Fourier coefficient of $\mathbf{x}_d$ for frequency ω; the vector $\mathbf{Y}_\omega = [\,Y^1(\omega)\,\ldots\,Y^N(\omega)\,]^{\mathrm{T}}$ collects the observations at frequency ω across the N input 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 for the input photos, adjusted to capture the depths of field $\mathcal{D}_1, \ldots, \mathcal{D}_N$ respectively; $\mathbf{I}_N$ is the $N \times N$ identity matrix; and $S_\omega = \big[\alpha(|G_x(\omega)|^2 + |G_y(\omega)|^2)\big]^{-1}$ is the per-frequency prior texture variance, where $\mathbf{G}_x, \mathbf{G}_y$ are vectors containing the Fourier-transformed gradient filters.

---

[1] For an in-focus image, mean pixel variance defines SNR completely. Even when deblurring (Sec. 3), spatial noise variations have negligible impact on SNR. We initially modeled per-pixel noise [8], but mean intensity gave the same SNR to ± 0.01 dB, with huge computational gain.

[2] Since we effectively average over ≥ 1 Mpixels, our model for sensor noise is Gaussian by the central limit theorem, even for low photon counts.

Bayesian all-in-focus restoration. When depth is known, the OTF vector $\mathbf{\Phi}_{d\omega}$ is fully specified by the camera design; restoration is then equivalent to N-photo Wiener deconvolution.[3] In our Bayesian treatment, the restored all-in-focus image corresponds to the maximum a posteriori (MAP) estimate [21], $\hat{X}_{d\omega}$, defined as the peak of

$$\Pr(X_\omega \,|\, \mathbf{Y}_\omega, d) \;=\; \mathcal{N}\big(\tfrac{1}{\eta^2}(\tau\mathbf{\Phi}_{d\omega}{}^*\mathbf{Y}_\omega)V_{d\omega}\,,\, V_{d\omega}\big), \quad (7)$$

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

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

For the case of unknown depth, we can express the per-pixel evidence for a specific depth using the observations, by evaluating $\Pr(\mathbf{Y}_\omega \,|\, d)$ in the spatial domain:

$$\log \Pr(\{\mathbf{y}^k\}\,|\,d) = \text{const} - \tfrac{1}{2}\Big[\overbrace{\textstyle\sum_{k=1}^{N}(\mathbf{y}^k - \tau\,\boldsymbol{\phi}_d^{\mathcal{D}_k}\otimes\hat{\mathbf{x}}_d)^2}^{\text{restoration error}} +$$
$$\underbrace{\alpha\big[(\hat{\mathbf{x}}_d\otimes\mathbf{g}_x)^2+(\hat{\mathbf{x}}_d\otimes\mathbf{g}_y)^2\big]}_{\text{prior term}} + \underbrace{\tfrac{1}{P}\textstyle\sum_\omega \log\big(\eta^2 S_\omega V_{d\omega}^{-1}\big)}_{\text{model complexity}}\Big], \quad (9)$$

where squaring in the first two terms is element-wise. This formula marginalizes over all scenes $\mathbf{x}$ exactly, accounting for all joint information encoded in the N photos [27, 21] and the volume of scenes explaining those observations.[4]

The all-in-focus image can then be computed by evaluating Eq. (9) for a set of depth hypotheses. In particular, we apply Bayes' rule to obtain the probability, $\Pr(d\,|\,\{\mathbf{y}^k\})$, of a given depth d; then we apply Eq. (7) to obtain the MAP restoration using the most probable depth.

Deriving the expected restoration error. When the scene depth is known, Eq. (7) tells us that the expected error of the MAP estimate is simply $\mathrm{E}[|X_\omega - \hat{X}_{d\omega}|^2] = V_{d\omega}$. This formula connects to $\mathrm{SNR}(N)$ in Eq.(4) via Parseval's rule, by which $\mathrm{E}\big[\|\mathbf{x}_d - \hat{\mathbf{x}}\|^2\big] = \sum_\omega V_{d\omega}$.

To treat the case of unknown depth, it is possible to show that the expected error of MAP restoration with an *incorrect* depth estimate, $\hat{d}$, has a closed form

$$\mathrm{E}\Big[|X_\omega - \hat{X}_{\hat{d}\omega}|^2\Big] =$$
$$\underbrace{\Big(\tau^2\tfrac{S_\omega}{\eta^2}|\Delta|^2 - (\Delta+\Delta^*)\Big)\tfrac{\tau^2}{\eta^2}V_{\hat{d}\omega}{}^2}_{\text{depth estimation error}} + \underbrace{V_{\hat{d}\omega}}_{\text{known-depth error}}, \quad (10)$$

where the scalar $\Delta = \mathbf{\Phi}_{\hat{d}\omega}{}^*(\mathbf{\Phi}_{d\omega}-\mathbf{\Phi}_{\hat{d}\omega})$ measures discrepancy between the true and estimated OTFs, and the true depth of X_ω is d. The formula consists of a bias term plus the basic known-depth restoration error.

To derive a scene- and depth-independent expected error,

---

[3]For camera designs that yield approximately depth-invariant blur [4, 11, 23], knowing the scene depth is not necessary for restoration.

[4]See the supplementary materials [1] for more details.

we take expectation over our depth estimate as well. More specifically, under the assumption that our depth uncertainty, $\Pr(\hat{d}\,|\,d)$, is independent of scene texture, we obtain

$$\mathrm{E}\big[\|\mathbf{x}_d - \hat{\mathbf{x}}\|^2\big] = \sum_\omega \int_{\hat{d}} \Pr(\hat{d}\,|\,d)\, \mathrm{E}\big[|X_\omega - \hat{X}_{\hat{d}\omega}|^2\big]\mathrm{d}\hat{d}. \quad (11)$$

In practice, we compute this expectation using Monte Carlo sampling, using Eq. (9) to compute the probability of obtaining different depth estimates $\hat{d}$ over a set of sampled noisy observations of a scene at depth d.

Scenes with depth variation. Fig. 4 shows restoration results from applying Eqs. (7)-(9) to photos of a simulated fronto-parallel scene. However, our analysis in Eq. (9) also lets us handle scenes with depth variations. Assuming a locally fronto-parallel scene, we can estimate per-pixel depths using $\hat{d} = \arg\max_d \Pr(d\,|\,\{\mathbf{y}^k\})$ and copy pixels from the corresponding constant-depth MAP restorations. In practice, we use a Markov random field (MRF) [2] to regularize this estimate and to favor piecewise smoothness (Fig. 5).

4. Optics-Independent Upper Bound

While restoration error depends greatly on camera design, we can establish an upper bound on the OTF magnitude (Fig. 2) for the worst-case depth and spatial frequency. Following [19], the total power useful for reconstructing each spatial frequency is bounded by $\frac{2}{3}A^3$, where A is the diameter of a circular aperture. The desired bound corresponds to spreading power evenly over the 3D manifold in the frequency domain corresponding to in-focus imaging [24, 19]:

$$|\Phi_d^{\mathcal{D}}(\omega)|^2 \;\leq\; A^4\cdot\min\left\{\tfrac{\beta(\omega)}{3\|\omega\|b_{\max}}\,,\,\tfrac{\pi^2}{16}\right\}, \quad (12)$$

where $b_{\max} = A\cdot\big(\tfrac{d_2-d_1}{d_2+d_1}\big)$ is the largest blur diameter over the target DOF $\mathcal{D} = [d_1, d_2]$, measured at depth $\tfrac{2d_1 d_2}{d_1+d_2}$ in the scene; and $\beta(\omega)$ is a factor in the range $[0.93, 1]$ [19].

Our bound tightens the result in [19] with a new bound over the *full* domain, corresponding to the final term in Eq. (12). By way of example, the DC component is bounded trivially by the spatial support, $\tfrac{\pi}{4}A^2$. Other frequencies share this bound as well, because Fourier coefficients are bounded by the integral of the modulus of the function; since our function is real and non-negative, this is simply the DC. More details are in the supplementary material [1].

5. The Time-Slice Advantage

To characterize the optimal time slices and their effect on SNR, we simulated several camera designs and estimated the restoration error in Eq. (4) over a large space of conditions—varying target DOF size, time budget, exposure level, noise parameters, and camera overhead—representing a wide range of photographic scenarios. This allowed us to assemble a detailed picture of the time-slice advantage, and of how these designs compare.
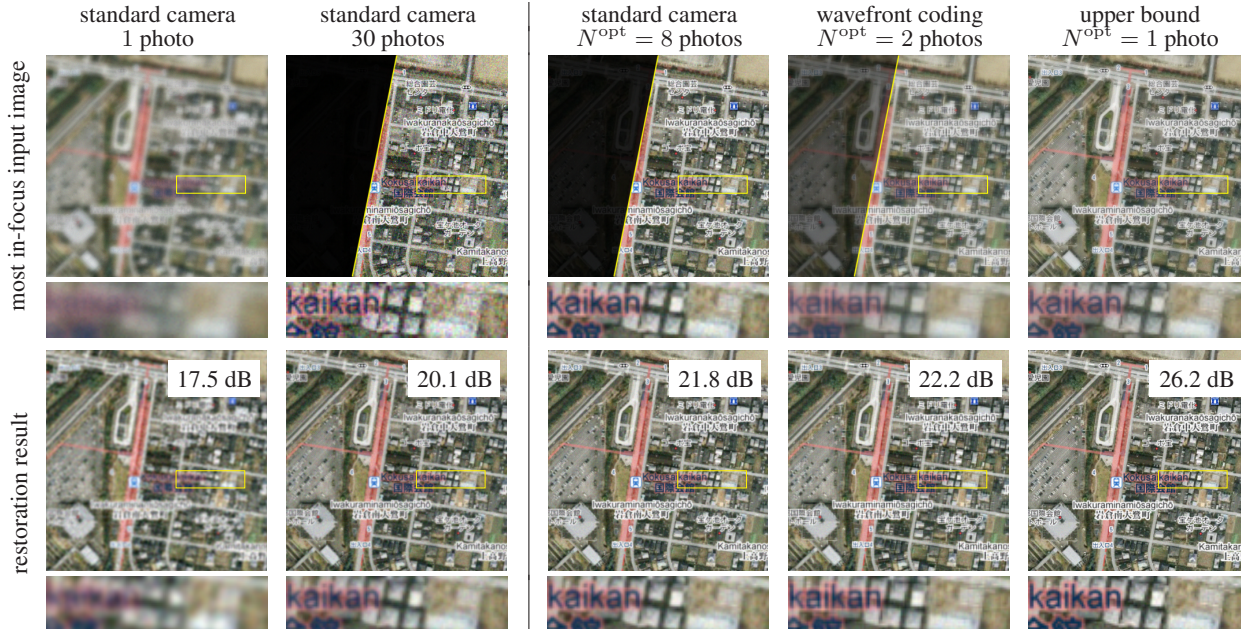
Figure 4. Image restoration under a time budget. We simulated photos of a street map (at the worst-case depth), and show results for several camera designs capturing various number of photos. We imposed a time budget of $0.1T^{\mathrm{opt}}$, and applied the same conditions as in Fig. 2. The left-most columns correspond to a suboptimal number of photos, leading to excessive defocus blur and noise respectively. **Top row:** most in-focus photo of all the observations, scaled for display. The split images show underexposure due to dividing the time budget. **Bottom row:** restoration assuming known depth. For this strongly-textured scene the Gaussian prior leads to conservative denoising.
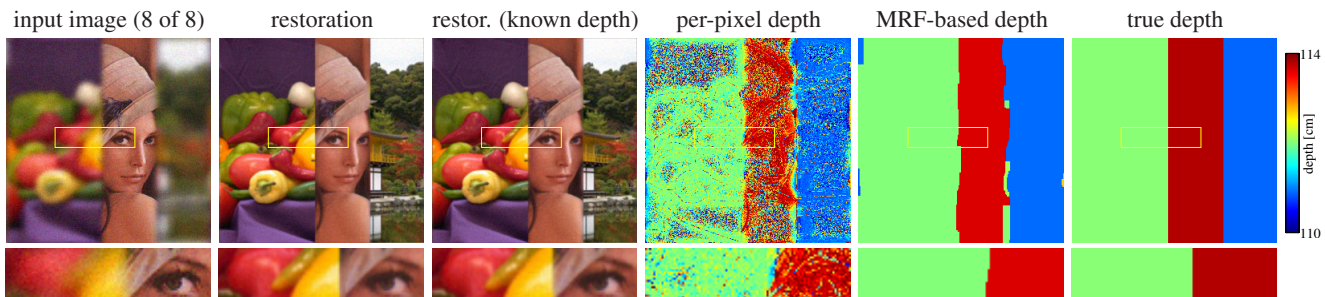

Figure 5. Image restoration with unknown depth. We simulated capturing 8 photos of a multi-layer scene using a standard camera with the configuration as in Fig. 4. We estimated depth using our Bayesian depth-from-defocus method, discretizing the DOF into 200 samples. While our method does not resolve exact per-pixel depth, the image restoration is of good quality, and its SNR (29.9 dB) is actually higher than when true depth is provided (29.5 dB). This reversal occurs because we ignore occlusion of defocus blur at depth discontinuities [9].

We evaluated the standard camera against the optimal bound of Sec. 4 and four extended-DOF cameras: focus sweep [11, 23], wavefront coding [4], coded aperture [17, 33], and (in Fig. 6 only) an idealization of the recent lattice-focal lens [19]. Note that although we simulate taking multiple shots with extended-DOF cameras, besides focus sweep, these designs may not be easily reconfigurable to handle different DOFs. See the supplementary materials [1] for implementation details; we plan to release our source code so that new designs can be evaluated as well.

5.1. Optimal time slices

Figures 6 and 7 summarize the results of our SNR analysis. These results show that despite fundamental differences in design, all cameras we considered share five basic properties with respect to the time-slice advantage:

A. N^{opt} is intermediate: When the lens DOF is small relative to the target DOF, capturing a small number of shots leads to significant blur at the worst-case depth. This causes strong frequency attenuation (Fig. 2, bottom), leading to poor image restoration and significantly reduced SNR (*e.g.*, Fig. 6, red curve in left-most plot). On the other hand, taking too many shots will reduce SNR as well—each shot incurs a penalty from additive noise but offers diminishing returns from blur reduction. By capturing the optimal number of photos, we can significantly increase SNR across all designs. Interestingly, this holds even for extended-DOF cameras designed specifically for single-shot photography
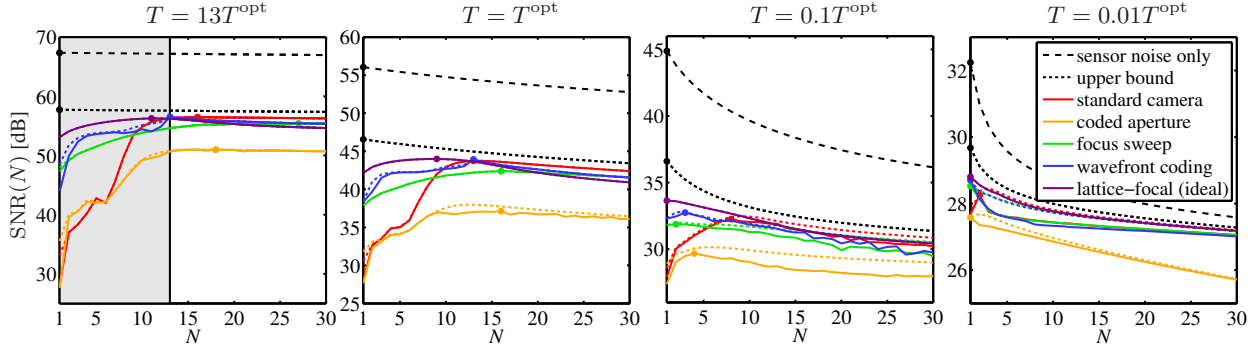
Figure 6. Restoration performance vs. number of photos for several time budgets. We use the same sensor and target DOF as in Fig. 2. The leftmost plot represents a time budget large enough to capture an ideally-exposed focal stack that spans the target DOF with the standard camera (*i.e.*, the budget for traditional extended-DOF photography [2, 10]). The shaded region corresponds to time slices that cause over-exposure and saturation. Solid curves show the restoration error assuming known depth; dotted curves show the error computed from Eq. (11), and account for uncertainty in depth estimation. For reference, the plots also include the optimal bound of Section 4 and the restoration error in absence of defocus.

(Fig. 6, other curves).

B. N^{opt} increases when time budget or target DOF is large: Figure 6 shows SNR behavior for several time budgets. Larger budgets increase the exposure level, which reduces the influence of additive noise and tilts the balance toward more photos. Similarly, as the DOF grows, so does worst-case optical blur; this leads to more aggressive divisions of the time budget that accept more noise (Fig. 7, top row).

C. Camera overhead reduces N^{opt}: We explore the effect of per-shot overhead in Fig. 7(middle row). This overhead (*i.e.*, due to mechanical limits or bandwidth) can be modeled as under-exposure, reducing N^{opt}. For low time budgets, overhead may limit N^{opt} to one photo.

D. Depth estimation penalty negligible for $N \geq 2$: Our simulations indicate that (planar) depth estimation from two photos is good enough to avoid SNR loss in the all-in-focus image (Fig. 6, solid vs. dotted curves). The only exceptions are very small budgets, where the influence of additive noise increases depth confusion (*e.g.*, $T = 0.01T^{\mathrm{opt}}$ for standard and $T = 0.1T^{\mathrm{opt}}$ for coded-aperture camera in Fig. 6).

E. The time-slice advantage: If photons could be counted perfectly, *i.e.*, with zero additive noise, the optimal time slice would involve as many photos as overhead allows. As the relative level of additive noise increases, however, N^{opt} decreases because each new photo incurs an additive penalty. This behavior is explored in Fig. 7(bottom row). Compared to other designs, the standard and coded aperture cameras are less sensitive to additive noise; their relatively narrow lens DOF (Fig. 2, bottom) forces them to accept more photos to offset worst-case blur.

5.2. Camera performance evaluation

The major difference between camera designs was in the value of the optimal time slice, N^{opt}. Although the actual effect is more complicated (Fig. 2, bottom), extended-DOF cameras can be thought of as having a larger DOF than the conventional camera. Thus they need fewer photos to reduce worst-case blur or to "span" a desired DOF.

A. Multi-shot standard camera outperforms single-shot extended-DOF cameras: For the DOF used in Fig. 6, the optimal time slice with a standard camera generally achieves much higher SNR than single-shot photography with extended-DOF cameras. This advantage holds even compared to the highly idealized model we use for the lattice-focal lens. Figure 7(top) shows that this holds for large DOFs in general; computational cameras have an advantage only when the time budget is very low (*e.g.*, $T \leq 0.1T^{\mathrm{opt}}$), or overhead is high (Fig. 7, middle).

B. N^{opt} helps standard camera the most: The relative ranking of extended-DOF cameras is stable across most conditions, and agrees with recent theoretical analysis for one-shot capture [19]: the lattice-focal lens is best, wavefront coding generally performs better than focus sweep, and all perform significantly better than coded aperture. In relative terms, conventional cameras gain more by capturing N^{opt} photos. While consistently the worst for $N = 1$, they achieve SNRs on par with the best computational designs by capturing N^{opt} photos, with only a slight performance loss at low time budgets.

C. Blocking light reduces SNR: The coded-aperture camera performs poorly because it sacrifices half the available light. Despite this handicap (up to 6.0 dB), it still achieves SNR on par with a standard camera in single-shot, large-DOF conditions.

D. Performance converges with increasing N: As the target DOF becomes finely divided, lens optics have less influence and power is more evenly spread. As a result, for large, suboptimal N, the performance of all designs (except the coded aperture) converges to the upper bound.

6. Experimental Validation

To test our predictions, we ran preliminary experiments with a high-end digital camera. We used a fronto-parallel

scene and carried out a wide range of capture strategies, collecting 345 images to evaluate 13 different candidates for N, for each of 2 time budgets (without taking camera overhead into account). This required calibrating the camera for focusing depth, image magnification over focus, and noise properties [5], then adjusting radiometry to account for lighting variations. As shown in Fig. 8(left), the restorations compare favorably to the ground truth. To confirm the time-slice advantage, we computed the actual SNR of the restored all-in-focus image for each run, and compared it to those from our model. These comparisons are shown in Fig. 8(right), and reveal good agreement. Moreover, our analysis correctly predicted the peaks of these curves, *i.e.*, the optimal time slice, for both time budgets we tested.

7. Concluding Remarks

We believe that one-shot photography provides an incomplete characterization of a given camera design, and of its imaging abilities. Our initial study of the time-slice advantage suggests that splitting an exposure budget across many photos can confer a significant quality advantage—for both conventional and computational cameras. Toward this goal, this paper provides the basic tools for (1) making comparisons across camera designs, and (2) answering the practical question of how many shots to take to optimize image quality within an exposure and DOF budget.

As in previous work on extended-DOF cameras [4, 23, 17], we pose image formation using a local convolution model. While this model may suffice to support our conclusions and to obtain good restorations for simple scenes (Fig. 5), a more sophisticated approach would explicitly model occlusion at depth discontinuities [9, 18]. Moreover, the question of how image priors [17] affect the time-slice advantage is still open. Our intuition is that stronger priors would lead to more photos, since they would allow us to denoise underexposed (but well-focused) photos more aggressively.

References

[1] http://csail.mit.edu/~hasinoff/timecon/.

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

[3] Y. Bando, B.-Y. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. In *SIGGRAPH Asia*, pages 1–9, 2008.

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

[5] R. N. Clark. Digital camera sensor performance summary, http://clarkvision.com/, 2008.

[6] H. Farid and E. P. Simoncelli. Range estimation by optical differentiation. *JOSA A*, 15(7):1777–1786, 1998.

[7] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *TPAMI*, 27(3):406–417, Mar. 2005.

[8] S. W. Hasinoff. *Variable-Aperture Photography*. PhD thesis, University of Toronto, Dept. of Computer Science, 2008.

[9] S. W. Hasinoff and K. N. Kutulakos. A layer-based restoration framework for variable-aperture photography. In *ICCV*, pages 1–8, 2007.

[10] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. In *ECCV*, volume 4, pages 45–59, 2008.

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

[12] G. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *TPAMI*, 16(3):267–276, 1994.

[13] 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.

[14] B. K. P. Horn. Focusing. Technical Report AIM-160, Massachusetts Institute of Technology, 1968.

[15] ISO 2721:1982. Photography—Cameras—Automatic controls of exposure, 1982.

[16] E. Krotkov. Focusing. *IJCV*, 1(3):223–237, 1987.

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

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

[19] 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.

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

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

[22] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *TPAMI*, 30(2):299–314, Feb. 2008.

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

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

[25] A. P. Pentland. A new sense for depth of field. *TPAMI*, 9(4):523–531, July 1987.

[26] R. J. Pieper and A. Korpel. Image processing for extended depth of field. *App. Optics*, 22(10):1449–1453, May 1983.

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

[28] Y. Schechner, S. Nayar, and P. Belhumeur. Multiplexing for optimal lighting. *TPAMI*, 29(8):1339–1354, Aug. 2007.

[29] W. J. Smith. *Modern Optical Engineering*. McGraw-Hill, New York, 3rd edition, 2000.

[30] J. Telleen, A. Sullivan, J. Yee, P. Gunawardane, O. Wang, I. Collins, and J. Davis. Synthetic shutter speed imaging. In *Eurographics*, pages 591–598, 2007.

[31] T. Treibitz and Y. Y. Schechner. Polarization: Beneficial for visibility enhancement? In *CVPR*, 2009.

[32] V. Vaish, R. Szeliski, C. L. Zitnick, and S. B. Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *CVPR*, volume 2, pages 2331–2338, 2006.

[33] 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.
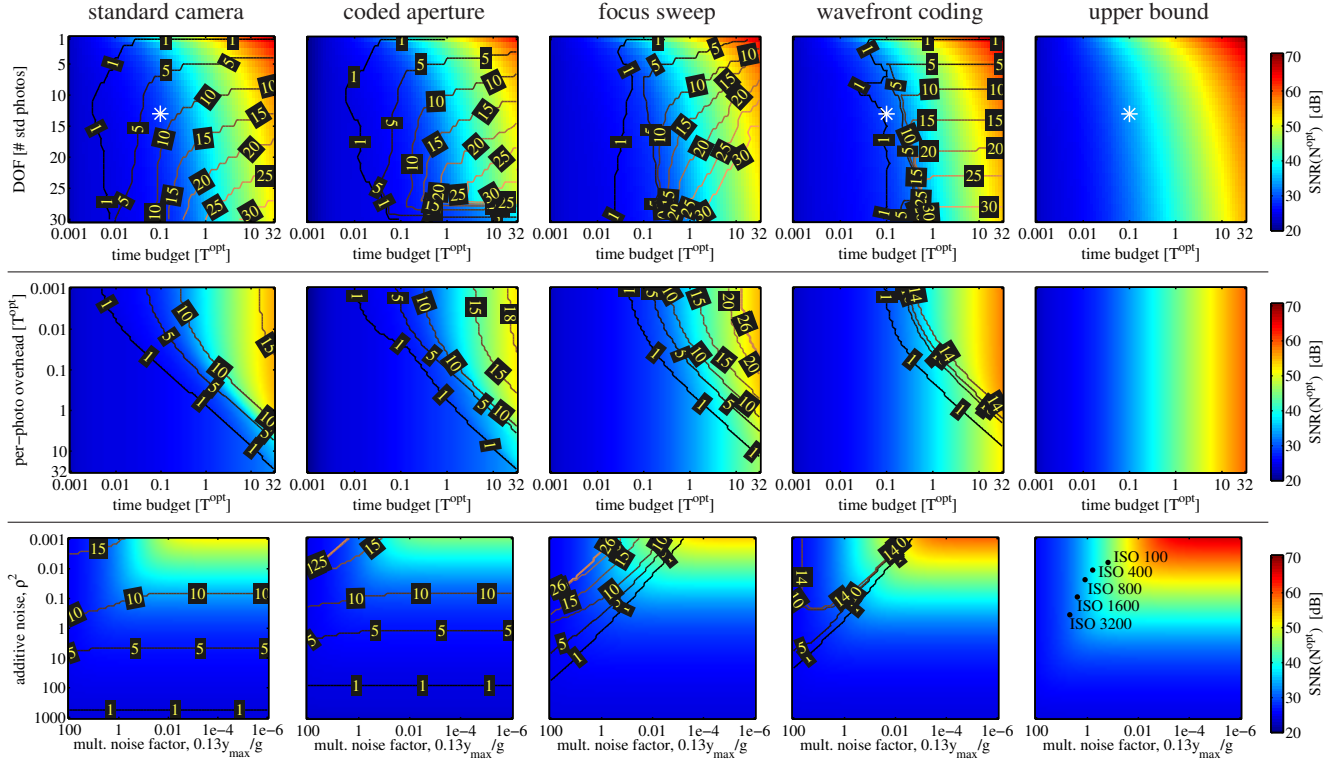
Figure 7. Performance analysis of optimal time-constrained photography. We use the same sensor as in Figs. 2–6. Each point in the plots represents a specific photography scenario, with more challenging ones going down and left in each plot. Color indicates SNR achieved by the optimal time slice and iso-contours indicate N^{opt}. No contours are shown for the optimal bound since it corresponds to one photo. **Top row:** *varying target DOF size and time budget*. We express DOF size by the size of the focal stack required to span it, using the standard camera of Fig. 2. White stars represent the scenario in Figs. 4–5, and the target DOF in Fig. 6. **Middle row:** *varying per-photo overhead and time budget*. The plots consider conditions identical to those of Fig. 6, with the addition of overhead. **Bottom row:** *varying noise properties*. We consider the same conditions as in Fig. 6, with the time budget fixed to $T = T^{\text{opt}}$, and vary the read noise and photon noise components of the noise variance. These components are expressed relative to a maximum pixel value of $y_{\max} = 255$. The points on the rightmost plot show values of sensor noise for the 1D Mark II at different ISO settings.
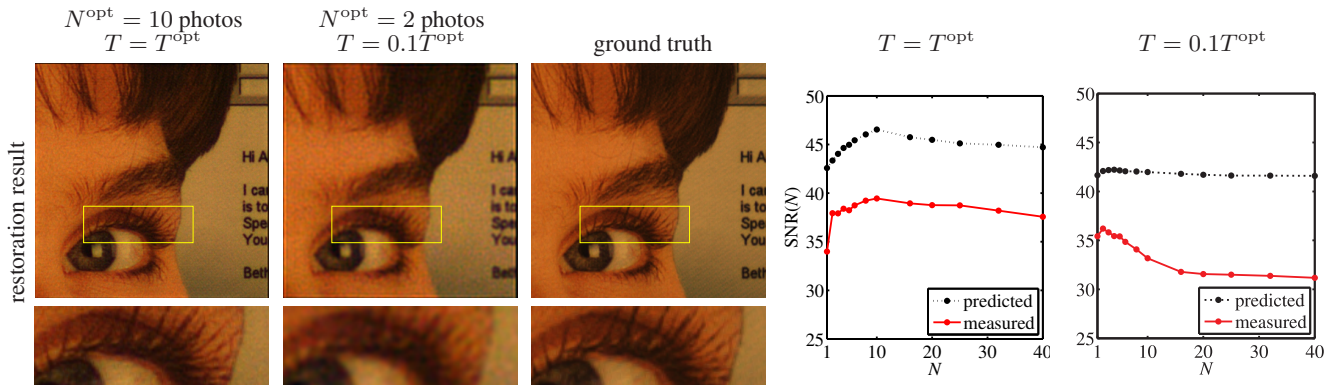

Figure 8. Experimental results for photos captured with a Canon 1D Mark III and an $85\,\text{mm}$ f/1.2 lens. Photos were captured at full ($10\,\text{MP}$) resolution in raw mode, with focus adjusted programmatically. The target DOF was $[95, 98]\,\text{cm}$, corresponding to about 13 shots at $f/1.2$. A scene (a printed advertisement) was placed at the worst-case depth ($d = 95\,\text{cm}$). For ground truth, we captured an ideally-focused and ideally-exposed image with $T^{\text{opt}} = 1/20\,\text{s}$. We then tested various time slices for two time budgets. Actual exposure times ranged from $1/20\,\text{s}$ to $1/8000\,\text{s}$. **Left:** Restorations for the predicted (and confirmed) optimal time slices for each time budget. **Right:** SNR predicted by our model, compared to SNR measured against ground truth. The gap between the curves is mainly the result of demosaicking artifacts, which affects a small number of high contrast edges, but does not reduce perceived image quality.