When Does Computational Imaging Improve Performance?

Oliver Cossairt, Mohit Gupta, and Shree K. Nayar

Abstract—A number of computational imaging techniques have been introduced to improve image quality by increasing light throughput. These techniques use optical coding to measure a stronger signal level. However, the performance of these techniques is limited by the decoding step, which amplifies noise. While it is well understood that optical coding can increase performance at low light levels, little is known about the quantitative performance advantage of computational imaging in general settings. In this paper, we derive the performance bounds for various computational imaging techniques. We then discuss the implications of these bounds for several real-world scenarios (illumination conditions, scene properties and sensor noise characteristics). Our results show that computational imaging techniques do not provide a significant performance advantage when imaging with illumination brighter than typical daylight. These results can be readily used by practitioners to design the most suitable imaging systems given the application at hand.

I. INTRODUCTION

COMPUTATIONAL Imaging (CI) techniques use optical coding followed by computational decoding. They can be classified into two categories. The first category includes techniques that provide a novel imaging *functionality*. For example, light field cameras capture 4D light fields that can be used to refocus or change perspective of images via post-processing [49], [29], [34]. Catadioptric imaging systems provide an immersive experience by capturing a wide-angle/omnidirectional field of view of the scene [53], [3], [36]. Tomographic imaging techniques recover the appearance of a 3D volume from a sequence of 2D projections [48]. Depth cameras capture scene structure using various approaches, such as and stereo [30], defocus [37], [45], [31], and diffusion [55]. These functionalities are impossible to achieve using a conventional imaging system.

The focus of this paper is on the second category of CI techniques, which are designed to improve *performance* in terms of image quality. These techniques use optical coding to increase light throughput and measure a stronger signal level. Examples include extended depth-of-field (EDOF) imaging [31], [49], [57], motion deblurring [38], [33], [11], 2D imaging [44], [7], [4], [5], [46], spectroscopy [21], [20], color imaging [2], [26], light field capture [49], [29], [34] and illumination multiplexing [41], [42], [39]. For each of these examples, there is a corresponding imaging technique that can measure the desired signal *directly without the need for any computational decoding*. For example, a shorter exposure can be used to eliminate motion blur and a stopped down aperture can be used to capture an EDOF image. Similarly, a pin-hole mask can be used to acquire light-fields and narrow-band spectral filters can be used (instead of a multiplexed

Oliver Cossairt, Mohit Gupta and Shree K. Nayar are with the Department of Computer Science, Columbia University, New York, NY, 10027 USA. E-mail: {ollie,mohitg,nayar}@cs.columbia.edu.
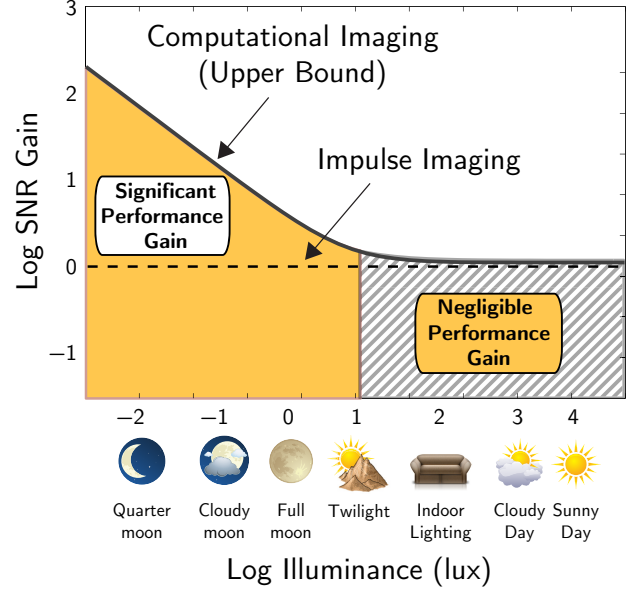
Fig. 1. **Performance of computational imaging for naturally occurring lighting conditions.** We show that CI techniques (solid curve) give a negligible performance gain over conventional (impulse) imaging (dotted line) if the illumination level is higher than that of a typical living room. This is an example plot for spectral, light field, and illumination multiplexing systems for the following scene and sensor characteristics: average scene reflectivity is .5, exposure time is $20\,ms$, aperture setting is $F/2.1$, pixel size is $1\mu m$, quantum efficiency is .5, and read noise standard deviation is $4e^-$. See Fig. 8 for similar performance plots for defocus and motion deblurring.

spectrometer) to capture multi-spectral images directly without requiring any computational decoding.

We refer to this class of imaging methods - ones whose performance we seek to improve by capturing more light - as *impulse imaging*. The term impulse is meant to convey the small amount of light captured by these methods. Impulse imaging techniques do not require computational decoding to recover the signal. Fig. 1 shows an example plot of performance for CI techniques relative to impulse imaging. Fig. 2 gives comparisons between some example CI techniques and their impulse imaging counterparts. The goal of this paper is to analyze the performance advantage of CI techniques with respect to their impulse imaging counterparts. The paper has two main contributions:

1) Theoretical performance bounds of computational imaging. Implementing a CI technique involves an additional, often significant, cost over a conventional imaging system. In order to justify the extra cost, a practitioner may ask the question: What is the performance advantage of a CI technique with respect to the corresponding impulse camera? Moreover, since CI techniques capture more light than impulse imaging, it may
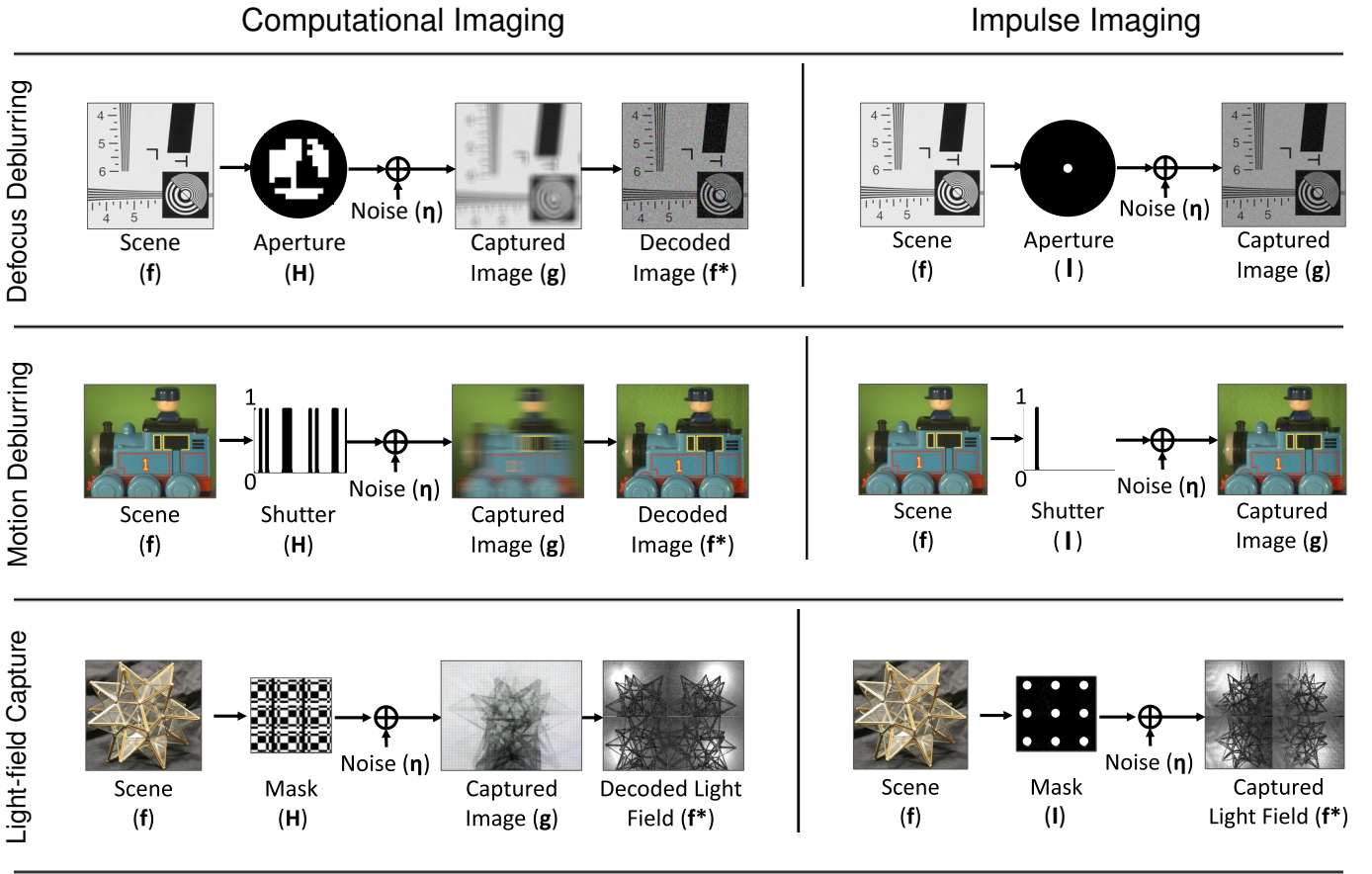
Fig. 2. **Computational versus Impulse Imaging.** (Left) All CI techniques discussed in this paper can be modeled using the linear image formation model given by Eq. 1. This includes defocus deblurring, motion deblurring, light field multiplexing, and several techniques discussed in Section III. In order to recover the desired image, these techniques require an additional decoding step, which amplifies noise. (Right) Impulse imaging techniques measure the signal directly without requiring any decoding. A stopped down aperture can be used to avoid defocus blur, a shorter exposure can be used to avoid motion blur, and a pin-hole mask can be placed near the sensor to directly measure the light field. The images in the figure are taken from [57], [38], [29].

appear that they must result in a higher signal-to-noise-ratio (SNR). However, CI involves a computational decoding step (see Fig. 2) which amplifies noise, thereby lowering the SNR.

We analyze the performance of a variety of CI techniques (e.g. EDOF imaging, motion deblurring and light-field capture), and derive a bound on their performance in terms of SNR. We show that CI techniques provide a significant performance advantage only if the average signal level is significantly lower than the sensor read noise variance. We also study the role of image priors on the decoding (CI) and denoising (impulse imaging) steps. Our empirical results show that the use of priors reduces the performance advantage of CI techniques even further.

2) Practical guidelines for computational imaging. Based on our performance bounds, we provide guidelines for when to use CI given an imaging scenario. The scenarios are defined in terms of the application (e.g., motion deblurring, defocus deblurring), real-world lighting (e.g., moonlit night or cloudy day, indoor or outdoor), scene properties (albedo, object velocities, depth range) and sensor characteristics. For example, our results show that computational imaging techniques do not provide a significant performance advantage when imaging

with illumination brighter than typical daylight. These results can be readily used by practitioners to decide whether to use CI, and if so, to design the imaging system.

A. Scope and Assumptions

Applicability: The results in this paper apply only to CI techniques that have a corresponding impulse imaging technique providing the same functionality. For example, the performance of EDOF techniques is compared with a stopped-down aperture, motion deblurring is compared with imaging using a small exposure and mask-based light field capture techniques are compared with pin-hole masks (see Figure 2).

Performance metrics: We use image quality for evaluating the performance of techniques with the same functionality. The theoretical performance bounds are derived in terms of the SNR metric. In addition, we provide empirical results for several other perceptually motivated metrics [51], [50], [43].

Imaging and noise model: The analysis in this paper deals with techniques which follow a linear imaging model. The noise is assumed to be additive Gaussian (signal independent and dependent), as discussed in Section II. We do not consider techniques which require non-linear computations for

recovering the desired image, e.g. depth estimation for EDOF imaging [22].

B. Related Work

Harvit and Sloane analyzed optically coded image acquisition in the context of spectrometers [21]. They showed that in the absence of photon noise, the optimal measurement scheme corresponds to Hadamard matrices, and provides significant performance advantage over impulse imaging. Ratner and Schechner [39], [40] extended these results to derive optimal measurement matrices in the presence of photon noise. Ihrke *et al.* [26] analyzed the noise performance of different light field cameras and color-filter array imagers. The performance gain of multiplexing, as derived in these papers, depends on the measurement matrices. Recently, Tendero [47] studied the performance gain of coded-exposure cameras [38] with respect to conventional impulse imaging, for motion deblurring. The general conclusion of the above-mentioned works was that CI techniques do not give a significant performance advantage at high light levels. Our contribution is to derive theoretical performance bounds for CI, which are independent of the optics. The bounds give the maximum performance for CI techniques as a function only of the signal level and camera read noise. We apply our results to several CI systems including EDOF imaging and motion deblurring, and provide practical guidelines for designing imaging systems.

Hasinoff *et al.* [22] show that the performance of any EDOF camera (conventional or computational) improves if multiple shots are taken with different focus settings. In a similar vein, Zhang *et al.* [54] compare the performance of acquiring multiple images versus a single image, in the context of motion deblurring and HDR imaging. These papers do not consider single shot impulse imaging. Moreover, both these papers require non-linear computations (depth estimation for EDOF and motion estimation for motion deblurring). In contrast, our focus is on linear imaging systems, and our goal is to analyze the performance gain of a wide range of CI techniques with respect to single shot impulse imaging.

II. IMAGE FORMATION MODEL

We consider CI techniques that can be expressed using a linear image formation model (see Fig. 2):

$$\mathbf{g} = H\mathbf{f} + \eta , \tag{1}$$

where $\mathbf{g}$ is the vector of measurements of size N. $\mathbf{f}$ is the vector of unknown signal values, which may represent spatial, spectral, angular, or temporal information. H is the measurement matrix. For CI techniques that take coded measurements by masking (attenuating) light, the entries of H are between 0 and 1. For CI techniques that measure the signal without masking light, either by moving the sensor during capture [23], [27], using additional refractive elements [17] or moving the camera [33] [1], the entries of H are not bounded. For impulse imaging, $H = I$, and the camera measures the signal $\mathbf{f}$ directly.

Each element of the noise vector η is assumed to be independently sampled from a zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$. We consider an affine noise model where there are two sources of noise, signal-independent *read noise*, and signal-dependent *photon noise* [2]. The photon noise can be approximated by a Gaussian with variance equal to the measured signal level J (in photons). Let the variance of the read noise be σ_r^2. The total noise variance is:

$$\sigma^2 = J + \sigma_r^2 , \tag{2}$$

An estimate of the signal can be found as:

$$\mathbf{f}^* = H^{-1}\mathbf{g} \tag{3}$$

The Mean-Squared Error (MSE) for the estimate $\mathbf{f}^*$ is given by [21]:

$$MSE = \frac{\sigma^2}{N} Tr(H^{-t}H^{-1}) , \tag{4}$$

where $Tr()$ is the matrix trace operator.

Performance Gain. In order to compute the performance gain of a CI technique, we compare the signal-to-noise-ratio (SNR) of the recovered signal with the signal captured using impulse imaging (baseline). The SNR is defined as $SNR = \frac{J}{\sqrt{MSE}}$.

Denoting σ_i^2 as the noise variance for the impulse camera, the MSE is just equal to the variance $MSE_i = \sigma_i^2$. Let σ_c^2 be the noise variance for the measurement made with the CI technique. The performance gain G is the ratio of the SNR for the CI technique to the SNR of the impulse camera:

$$G = \sqrt{\frac{MSE_i}{MSE_c}} \tag{5}$$

$$= \sqrt{\frac{N}{Tr(H^{-t}H^{-1})}} \frac{\sigma_i}{\sigma_c}. \tag{6}$$

When the noise is signal independent ($\sigma_i = \sigma_c$), the matrix that maximizes the gain for masking-based CI techniques is the S-matrix [21]. However, when the noise is signal dependent, the optimal measurement matrix, and hence the performance gain, depend on the *matrix light throughput* $C(H)$, which is the sum of elements in each row of the measurement matrix [3] H, and is a measure of the amount of light captured if H is used as the measurement matrix. For example, if H is the identity matrix, $C(H) = 1$. On the other hand, if H is the S-matrix, $C(H) \approx \frac{N}{2}$. Consequently, the S-matrix captures significantly more light. In the remainder of the paper, we drop the argument H from $C(H)$ for brevity.

Optimal Measurement Matrices. The problem of identifying optimal measurement matrices (that result in maximum gain)

---

[1] These techniques have an added benefit that they result in depth invariant blur (for defocus deblurring) and motion invariant blur (for motion deblurring), which makes the deblurring process significantly simpler.

[2] We ignore the effect of dark current noise, which is typically negligible when exposure times remain less than around one second.

[3] We consider matrices for which all the rows have the same sum.

for masking-based CI techniques was explored by Ratner *et al.* [39], [40]. They found an analytic expression for the lower bound of the trace term:

$$Tr(H^{-t}H^{-1}) \geq \frac{N(CN - 2C + 1)}{(N - C)C^2} \, . \tag{7}$$

Suppose the average signal level for the impulse camera is J, so that the total noise $\sigma_i^2 = J + \sigma_r^2$. The CI technique captures C times more light. Hence, the total noise is $\sigma_c^2 = C\,J + \sigma_r^2$. Substituting these and Eq. 7 in Eq. 6, we get an expression for the maximum SNR gain G for matrices with a given light throughput C:

$$G(C) \leq \sqrt{\underbrace{\frac{(N - C)C^2}{CN - 2C + 1}}_{\text{Decoding Term}} \quad \underbrace{\frac{J + \sigma_r^2}{C\,J + \sigma_r^2}}_{\text{Noise Term}}}. \tag{8}$$

The right hand side of Eq. 8 consists of two competing terms. As light throughput is increased, the noise dependent term decreases while the decoding dependent term increases. There is an optimal light throughput C_{max} for which the SNR gain achieves the maximum value of G_{max}.

III. COMPUTATIONAL VERSUS IMPULSE IMAGING: EXAMPLES

There are several CI techniques that follow the linear imaging model of Eq. 1 and which have a corresponding impulse camera. The following are a few examples.

Defocus Blur. Coded aperture masks have been used to deblur defocused images [31], [49], [57], [56]. There are also several techniques that extend depth of field (DOF) by producing a depth-independent blur that can be inverted without the need for depth estimation [10], [17], [35], [23], [27], [14], [13], [19]. Assuming the blur kernel to be shift invariant, the measurement matrix H is a circulant matrix, where each row of the matrix encodes the defocus blur kernel. $\mathbf{g}$ is the captured blurred image and $\mathbf{f}$ is the EDOF image. The corresponding impulse imaging technique is to capture images with a stopped down aperture (H is equal to the identity matrix I).

Motion Blur. Temporal shuttering has been used to remove motion blur from images [38]. Methods have also been proposed that create motion-invariant blur that can be removed without prior knowledge of object speed [33], [11]. Similar to EDOF imaging, the measurement matrix H is a circulant matrix, where each row of the matrix encodes the motion blur kernel. $\mathbf{g}$ is the captured blurred image and $\mathbf{f}$ is the blur-free image. In contrast, impulse imaging avoids motion blur by simply capturing images with a short exposure ($H = I$).

Multiplexed Light Fields. CI techniques for capturing light fields include placing a transmissive mask either at the lens aperture [34], or near the sensor [49], [29]. In this case also, the measurement matrix H is block circulant. Each pixel measures a linear combination of ray intensities ($\mathbf{g}$) and the light field ($\mathbf{f}$) must be recovered by demultiplexing the captured data. In contrast, a light field camera can also be built by placing a

mask consisting of an array of pinholes near the sensor [25], or by capturing a sequence of images with a shifting pinhole in the aperture [34]. These techniques are the impulse imaging counterparts of multiplexed light field capture.

Multiplexing Color and Spectrum. Mask-based Hadamard multiplexing is used for point [21] and imaging [20] spectrometers. Here, H is the spectral mixing matrix, $\mathbf{g}$ is the vector of multiplexed spectral samples and $\mathbf{f}$ is the vector of narrow-band spectral samples (desired). The impulse imaging counterpart is capturing narrow-band spectral samples [8]. Color Filter Arrays (CFAs) that multiplex color have been proposed to capture three color images with more light throughput than RGB Bayer filters [2], [26] (impulse imaging).

Multiplexed Illumination. Measuring the appearance of a scene under varying illumination is useful for scene relighting and estimating depth. These measurements can be multiplexed by measuring the appearance of a scene when illuminated by linear combinations of light sources [42], [39], [40]. Here, H is the measurement ensemble, $\mathbf{g}$ is the vector of acquired (multiplexed) image intensities and $\mathbf{f}$ is the vector of intensities corresponding to only single sources. Here, the impulse imaging counterpart is capturing images by turning on only one light source at a time.

IV. OPTICS INDEPENDENT PERFORMANCE BOUNDS

In this section, we derive a performance bound for CI that is independent of the signal size N and the measurement matrix H (defined by the optics). As a result, this bound allows us to analyze a wide range of CI techniques. In comparison, the previous result (Eq. 8) gives the maximum SNR gain G for coding matrices with a light throughput C and signal of size N. We first derive an upper bound on the decoding term in Eq. 8:

$$
\begin{aligned}
\frac{C^2(N - C)}{NC - 2C + 1} &= \frac{(N - C)C^2}{NC - C^2 + C^2 - 2C + 1} \\
&= \frac{(N - C)C^2}{C(N - C) + (C - 1)^2} \\
&\leq \frac{(N - C)C^2}{C(N - C)} \\
&\leq C. \tag{9}
\end{aligned}
$$

Next, we derive an upper bound on the noise term:

$$
\begin{aligned}
\frac{J + \sigma_r^2}{C\,J + \sigma_r^2} &= \frac{1}{C}\,\frac{C\,J + C\,\sigma_r^2}{C\,J + \sigma_r^2} \\
&\leq \frac{1}{C}\,\frac{C\,J + \sigma_r^2 + C\,\sigma_r^2}{C\,J + \sigma_r^2} \\
&\leq \frac{1}{C}\left(1 + \frac{C\,\sigma_r^2}{C\,J + \sigma_r^2}\right) \\
&\leq \frac{1}{C}\left(1 + \frac{\sigma_r^2}{J}\right) \tag{10}
\end{aligned}
$$

By substituting the bounds in Eqs. 9 and 10 in Eq. 8, we get the upper bound on the performance gain of masking-based CI techniques:

(a) Performance for defocus blur   (b) Performance for motion blur   (c) Performance for light field multiplexing
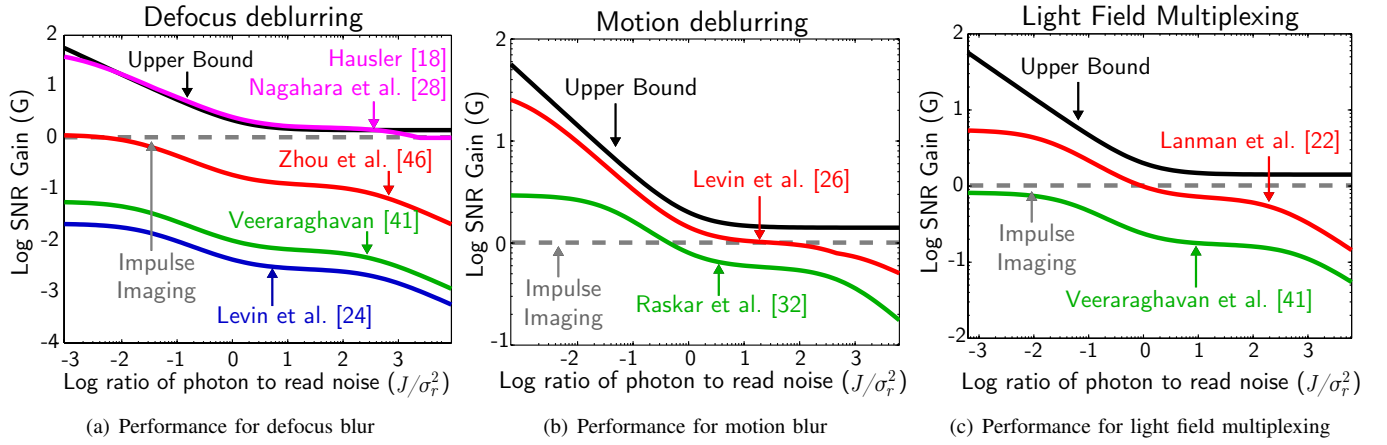
Fig. 3. **Verification of the performance bound using simulations.** (a) Performance for several defocus deblurring cameras [31], [49], [57], [23], [27]. (b) Motion deblurring performance for the flutter shutter [38] and motion-invariant cameras [33]. (c) Performance for the sum-of-sinusoids [49] and MURA [29] light field multiplexing cameras. All techniques perform at or below the performance bound given by Eq. 12. The plots in 3(c) are similar to those given by Ihrke *et al.* [26]. However, our plots are different for higher signal levels because we reduce camera sensitivity to avoid saturation.

$$G < \sqrt{1 + \frac{\sigma_r^2}{J}}. \tag{11}$$

In Appendix A, we derive the performance bound for the CI techniques discussed in this paper that do not mask light [10], [17], [35], [23], [27], [14], [13], [19], [33], [11]:

$$G < \sqrt{2 \left(1 + \frac{\sigma_r^2}{J} \right)}. \tag{12}$$

Eqs. 11 and 12 are noteworthy because they provide the maximum possible SNR gain for all CI techniques mentioned in Section III.

Simulations to Verify the Bounds. In Fig. 3, we show the simulated performance of several of the techniques discussed in Section III. The SNR gain of each technique is calculated using Eq. 6, and the result is plotted against the ratio of photon to read noise variance (J/σ_r^2). Fig. 3(a) shows performance for several previously proposed defocus deblurring cameras [31], [49], [57], as well as the focal sweep camera [23], [27]. Fig. 3(b) shows motion deblurring performance for the flutter shutter [38] and motion-invariant [33] cameras. For focal sweep and motion-invariant techniques, the coding was optimized for different signal levels. Fig. 3(c) shows performance for the sum-of-sinusoids [49] and MURA [29] light field multiplexing cameras. The masks for both light field cameras were generated with a period of 11×11 pixels. As expected, all the techniques perform at or below the performance bound given by Eq. 12.

Implication of the Bounds. The bounds in Eqs. 11 and 12 imply that the *performance gain for computational imaging is significant only when the average signal level J is considerably smaller than the read noise variance σ_r^2*. The read noise variance for currently available sensors ranges from less than one grey level (on a scale of $[0 - 255]$) for high quality DSLR cameras to approximately 5 grey levels for low-quality

machine-vision cameras [42]. Typical daylight illumination results in signal strengths that are considerably greater than these read noise variance values.

V. ROLE OF IMAGE PRIORS

Thus far, we have not considered the role of image priors. Priors can be used to improve image quality, both for computational and impulse imaging [31], [57]. The improvement depends on the type of priors and image coding used. In addition, our analysis thus far used MSE as the quality metric for images because it makes the derivation of performance bounds tractable. However, a number of metrics have been introduced which measure the perceived image quality [51], [43], [50], [18]. In this section we analyze the effect of various priors and metrics on performance.

Image Priors. We can think of the estimate $\mathbf{f}^*$ given in Section II as the Maximum Likelihood (ML) estimate of the following optimization problem:

$$\mathbf{f}_{ML}^* = \arg \max_{\mathbf{f}} P(\mathbf{f}|\mathbf{g}) \tag{13}$$

$$= \arg \min_{\mathbf{f}} ||\mathbf{f} - H\mathbf{g}||^2, \tag{14}$$

where $P(\mathbf{f}|\mathbf{g})$ is the probability of the unknown image $\mathbf{f}$ given the measurement $\mathbf{g}$, which is Gaussian due to the properties of the measurement noise vector η. If we have knowledge of the probability distribution of our unknown image $P(\mathbf{f})$, then we can improve performance by finding the Maximum A Posteriori (MAP) estimate:

$$\mathbf{f}_{MAP}^* = \arg \max_{\mathbf{f}} P(\mathbf{g}|\mathbf{f})P(\mathbf{f}) \tag{15}$$

$$= \arg \min_{\mathbf{f}} ||\mathbf{f} - H\mathbf{g}||^2 + \lambda \log(P(\mathbf{f})), \tag{16}$$

where the constant λ determines how much significance to attach to the prior. The image prior can essentially be thought of as a way to coerce the optimization problem to produce

more probable estimates of the unknown image. In this section, we consider three image priors. Firstly, we consider a Gaussian prior on the distribution of gradients in the image [31], [22] of the form :

$$\log(P_{Gauss}(\mathbf{f})) = -||\nabla \mathbf{f}||_2^2, \qquad (17)$$

where ∇ is the first order finite difference matrix, which calculates a discrete approximation to the gradient operator. In this case, because the prior is Gaussian, the MAP estimate can be calculated directly as:

$$\mathbf{f}_{MAP}^* = (\nabla^t \nabla + H^t H)^{-1} H^t \mathbf{g}. \qquad (18)$$
$$(19)$$

We also consider the Total Variation (TV) prior [6], [9] which has the form:

$$\log(P_{TV}(\mathbf{f})) = -||\nabla \mathbf{f}||_1. \qquad (20)$$

In this case, there is no direct way to calculate the MAP estimate, and an iterative algorithm must be used. We use the TwIST algorithm [6] to solve Eqn. 16 and find a MAP estimate.

Lastly, we consider the current state-of-the-art BM3D algorithm [15]. Although it is hard to write a single linear expression for this prior (because of non-linear shrinkage in the 3D spectrum domain), it is possible to calculate a MAP estimate numerically.

Image Quality Metrics. In Section II, we establish MSE as the metric for evaluating images. We now extend this formalization to more general metrics. We can define a general form for any similarity metric $S(\mathbf{f}, \mathbf{f}^*)$ that measures how close our estimate $\mathbf{f}^*$ is to the actual image $\mathbf{f}$. The MSE metric can then be defined as

$$S^{MSE}(\mathbf{f}, \mathbf{f}^*) = \frac{1}{||\mathbf{f} - \mathbf{f}^*||_2^2}. \qquad (21)$$

The interpretation here is that the smaller the MSE, the more similar the estimate $\mathbf{f}^*$ relative to the true image $\mathbf{f}$. The performance gain for any metric can then be written as

$$G = \sqrt{\frac{S_c(\mathbf{f}, \mathbf{f}^*)}{S_i(\mathbf{f}, \mathbf{f}^*)}}, \qquad (22)$$

where S_c is the metric applied to the computational camera, and S_i is the metric applied to the corresponding impulse camera. For the MSE metric, the definition remains the same as the expression given in Eq. 6. However, with this general definition in place, we are now free to use other metrics to evaluate performance. Note that because we have defined SNR gain in terms of a similarity metric instead of an error metric, the term for the computational camera appears in the numerator instead of the denominator. In addition to MSE, we use the following image quality metrics to measure performance: Structural Similarity (SSIM)[51], Visual Information Fidelity

(VIF)[43], and Universal Quality Index (UQI)[50]. We use the MeTriX MuX Visual Quality Assessment package to calculate performance using these metrics [18].

Simulations. Fig. 4 shows the simulated performance of focal sweep and flutter shutter cameras. Since the performance of reconstruction algorithms can be image depend, we report the performance averaged over a large dataset of images. For this simulation, we use the Caltech 101 image database of 9140 different images [28]. For each image, we simulate the performance under ten different photon to read noise ratios (J/σ_r^2). Moving from left to right columns, performance is shown for the MSE, SSIM, VIF, and UQI metrics. The top row shows performance for the focal sweep camera, and the bottom row shows performance for the flutter shutter camera. For each plot, the performance gain G is plotted on a log scale. Thus, a value of zero corresponds to a performance gain of $G = 1$, meaning that both computational and impulse imaging have the same performance (dotted line). The black line corresponds to the performance bound expressed by Eqn. 12. The magenta lines correspond to performance gain using direct linear inversion (i.e. estimating the image using Eqn. 3). The red, green, and blue curves correspond to reconstructions using Gaussian, TV, and BM3D priors, respectively.

There are two interesting observations to be made from these plots. First, in most of the cases, image priors boost the performance of impulse imaging more than computational imaging. As a result, the performance advantage of CI techniques over impulse imaging is reduced even further, especially at low light levels. Thus, the performance bound expressed by Eqn. 12 is the tightest when no prior is used.

The second observation is that the bound derived using linear inversion and the MSE metric (black curve) appears to be an upper bound for performance across all metrics and priors. This is surprising because it is well known that MSE does not accurately measure perceived image quality. Nonetheless, the upper bound expressed by Eqn. 12 does appear to provide a consistent upper bound on performance regardless of the image quality metric used.

VI. WHEN TO USE COMPUTATIONAL IMAGING

Eqs. 11 and 12 provide performance bounds for CI techniques in terms of the sensor read noise σ_r and the average signal level J of the impulse image. In order to determine when CI is advantageous, we have derived an expression for the signal level J in terms of the scene and sensor dependent parameters (see Appendix B for a derivation):

$$J = 10^{15} \underbrace{(F/\#)^{-2} \, t \, I_{src} \, R}_{\substack{\text{Scene} \\ \text{Dependent}}} \underbrace{q \, \Delta^2}_{\substack{\text{Sensor} \\ \text{Dependent}}}, \qquad (23)$$

where $F/\#$ is the ratio of focal length to aperture size of the lens, t is the exposure time, I_{src} is the incident illuminance given in lux, R is the average reflectivity of the scene, q is the quantum efficiency of the sensor, and Δ is the pixel size in meters. In Fig. 5, we give values of J corresponding to several commonly encountered lighting conditions.
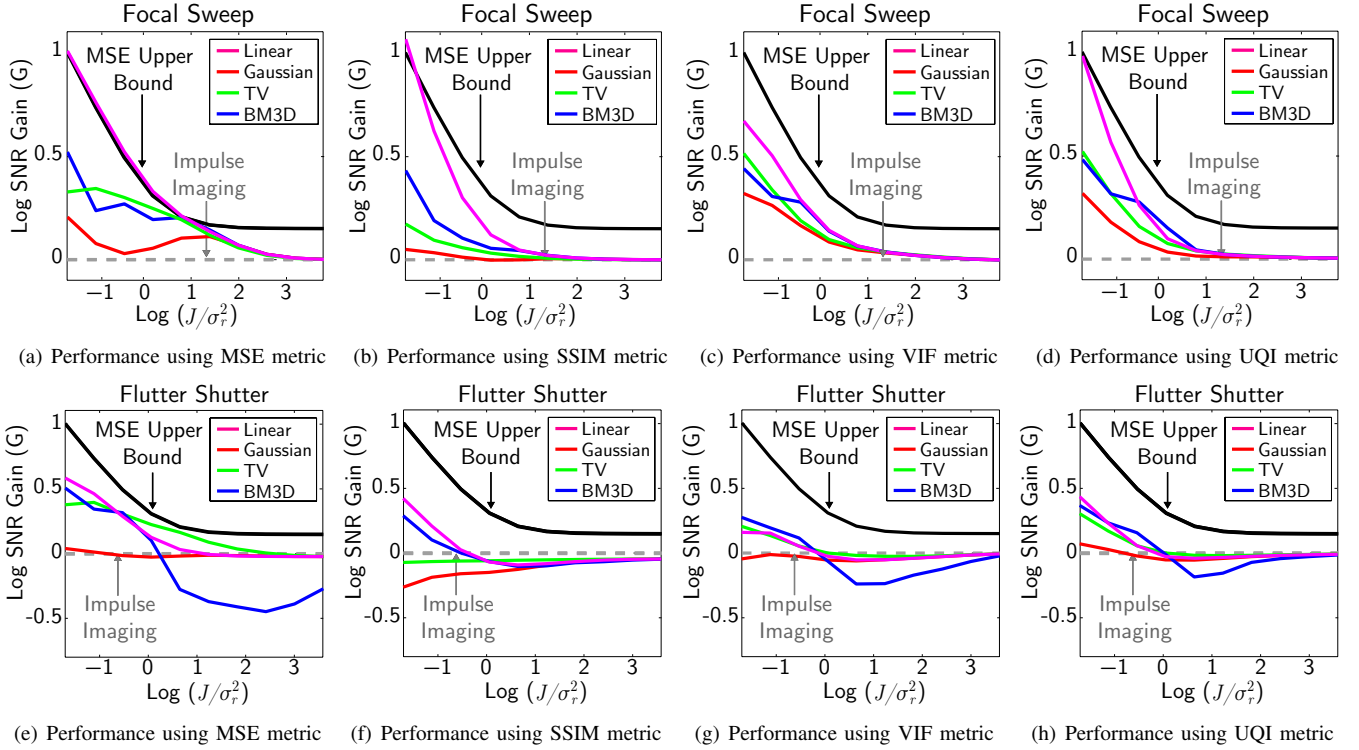
Fig. 4. **Simulated performance for a focal sweep (top row) and flutter shutter (bottom row) cameras using various priors and metrics.** (a)(e) Performance using the Mean-Squared Error metric. (b)(f) Performance using the Structural Similarity Metric. (c)(g) Performance using the Visual Information Fidelity Metric. (d)(h) Performance using the Universal Quality Index metric. Several reconstruction techniques are applied, including Linear inversion (magenta), Gaussian prior on derivatives (red), Total Variation prior (green), and BM3D prior (blue). The SNR gain is always less than the bound given by Eq. 12, regardless of the prior or metric used.

Scene Dependent Parameters. For defocus deblurring systems, the $F/\#$ of the camera depends on the depth range of the scene. A larger depth range will require the impulse camera to stop down to smaller apertures to reduce defocus blur to within one pixel. Similarly, for motion deblurring systems, the exposure time t depends on the range of scene velocities. Higher velocities will require the impulse camera to have a small exposure to reduce motion blur to within one pixel. Finally, the signal level is directly proportional to the illumination brightness I_{src} and the object albedo R.

Sensor Dependent Parameters. Different sensors have different quantum efficiencies and pixel sizes. Quantum efficiency for commercially available sensors is quite high, usually greater than $q > .5$. For today's sensors, the size of pixels Δ has a wide range, from 1 micron for small cell phone sensors to nearly 10 microns for large format sensors.

A. Rule of Thumb

When using one of today's commercial grade image sensors (DSLRs, cell-phone cameras, machine vision sensors), computational imaging will only yield significant performance benefits when the illuminance is less than $125\,lux$ (typical living room lighting).

For EDOF imaging and motion deblurring, this implies that when the illuminance is high (more than typical living room lighting), it is better to capture the impulse image without any blur (using a small aperture and exposure, respectively). Similarly, for light field and spectral acquisition, if the illuminance

	Starry night	Quarter moon	Full moon	Twilight	Indoor Lighting	Cloudy Day	Sunny Day
$I_{src}(lux)$	2×10^{-3}	10^{-2}	1	10	10^2	10^3	10^4
$J\,(e^-)$	7×10^{-4}	4×10^{-3}	.39	3.85	38.49	384.9	3,849

Fig. 5. **Relating lighting levels to average photon counts.** The top row shows typical illuminance values in lux [52]. The bottom row shows the photon counts calculated using Eq. 23 assuming an average reflectivity of $R = .5$, quantum efficiency of $q = .5$, and exposure time of $t = 1/50\,seconds$, aperture setting of $F/2.1$, and pixel size of $\Delta = 1\mu m$.

is high, it is better to capture the light field and the spectral samples directly without multiplexing (using pin-hole masks and narrow-band filters, respectively).

We support this rule of thumb with several example scenarios. Each scenario consists of an application, lighting condition (e.g., moonlit night or cloudy day, indoor or outdoor) and scene properties (albedo, speed, range of object velocities). In all our examples, we assume an average reflectivity of $R = 0.5$, quantum efficiency of $q = 0.5$, and read noise of $\sigma_r = 4e^-$, which is typical for today's CMOS sensors [12].

Motion Deblurring. For this case, we used a pixel size of $\Delta = 5\mu m$, aperture setting of $F/20$ and the impulse camera exposure time $t = \frac{1}{50}\,s$. For flutter shutter camera, we use
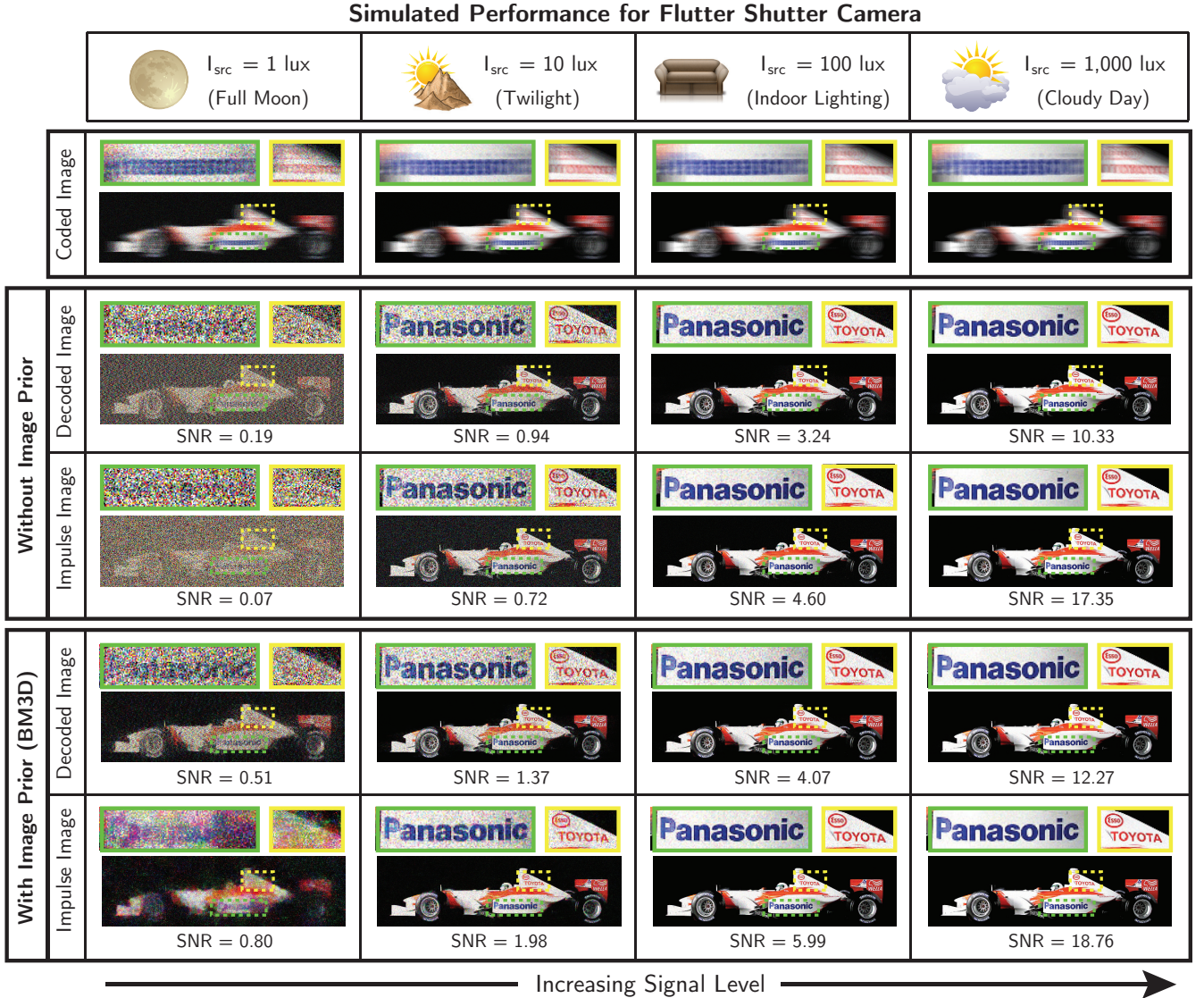
Fig. 6. **Simulated performance for the flutter shutter camera.** The parameters for the simulation are given in Section VI-A. The top row shows an image blurred by the flutter sequence given in [38]. The second row shows the results after linear deblurring. The third row shows the results from the impulse camera (shorter exposure). The fourth row shows the results after deblurring the images in the first row with the BM3D algorithm [15]. The last row shows the results for denoising the images in the third row with the BM3D algorithm. Gaussian distributed read noise and Poisson distributed photon noise is added to each image. The illumination I_{src} increases from left to right. The flutter shutter camera has higher SNR when $I_{src} < 100\,lux$.

the 52 digit long sequence given in [38]. We simulated the effect of photon and read noise, and decoded the captured image using Eq. 3. In Fig. 6, we show the simulated images. The flutter shutter performance is lower than that of impulse imaging when the illuminance is greater than $100\,lux$.

In Fig. 8(a), we show a contour plot of the SNR gain bound [4] versus the illuminance (I_{src}) and the exposure time of the impulse camera (t). Note that the bound is independent of the particular flutter sequence and the exposure time of the flutter camera. As the maximum object speed increases, the exposure time of the impulse camera must be reduced to avoid motion blur. We can observe that CI never gives an SNR gain greater

---

[4]Although the SNR bound is derived assuming no priors and MSE metric, we have observed empirically that it bounds the performance of CI techniques irrespective of the prior and the image quality metric (see Section V).

---

than 2 when the illuminance is greater than $83\,lux$.

EDOF Imaging. For this case, we used a pixel size of $\Delta = 5\mu m$, the camera exposure time $t = \frac{1}{50}\,s$ and the impulse aperture setting of $F/20$. The aperture setting for the focal sweep camera was set to $F/1$. In Fig. 7, we show simulated images. The figure shows captured (coded), decoded, and impulse images with read noise and varying amounts of photon noise added. Images are decoded using Eq. 3. The performance of focal sweep is always greater than impulse imaging (i.e. stopping down the camera), but the increase in performance is negligible when the illuminance is greater than $100\,lux$.

In Fig. 8(b), we show a contour plot of the SNR gain bound (independent of the focal sweep aperture setting) versus the illuminance and the $F/\#$ of the impulse camera. As the scene depth increases, $F/\#$ of the impulse camera must also
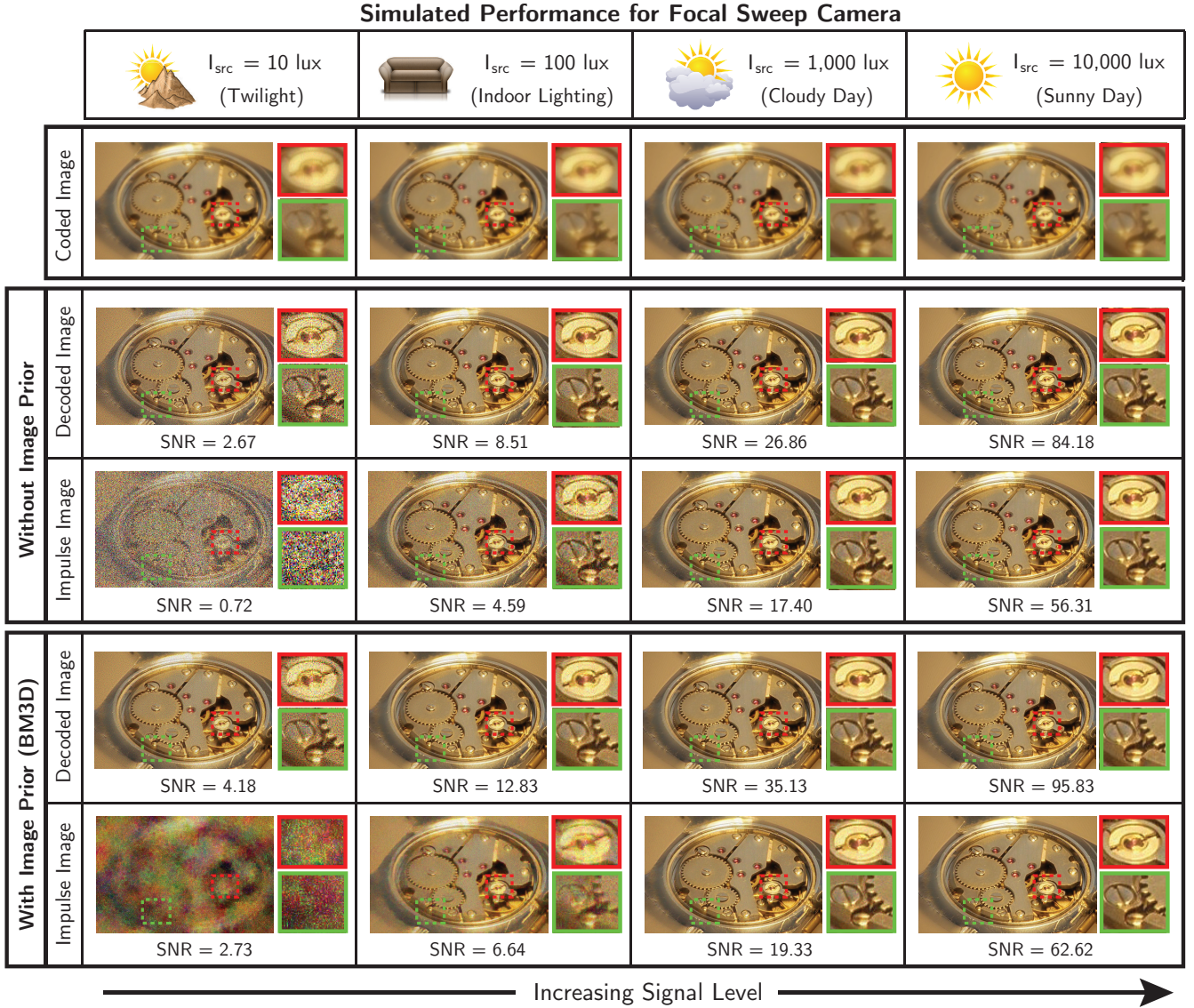
Fig. 7. **Simulated performance for the focal sweep camera.** The parameters for the simulation are given in Section VI-A. The top row shows an image blurred by a focal sweep PSF. The second row shows the results after linear deblurring. The third row shows the results from the impulse camera (stopped down aperture). The fourth row shows the results after deblurring the images in the first row with the BM3D algorithm [15]. The last row shows the results for denoising the images in the third row with the BM3D algorithm. Gaussian distributed read noise and Poisson distributed photon noise is added to each image. The illumination I_{src} increases from left to right. The focal sweep camera always has a higher SNR than impulse imaging, but the improvement becomes negligible when $I_{src} > 100\,lux$.

increase to avoid defocus blur. As we can notice, CI does not give an SNR gain greater than 2 when the illumination is more than $125\,lux$.

Spectral and Light Field Acquisition. In this example, we consider the performance of spectral and light field cameras. We use an exposure time of $t = 1/50$ seconds, pixel size of $\Delta = 1\mu m$, and an aperture setting of $F/2.1$. In Fig. 1 we plot the SNR gain bound against illuminance. In this case the illuminance must be less than $18\,lux$ in order for CI to give an SNR gain greater than 2.

VII. DISCUSSION

The Role of Sensor Quality. High quality sensors are carefully engineered to have low read noise, albeit with an added cost. In applications where low quality (high read noise) sensors are used, CI can enhance performance even when illuminance is greater than $125\,lux$. In these situations, however, the additional cost required to implement coding should be weighed against the cost of improving performance by simply switching to a high-quality sensor.

Effects of Diffraction: Defocus blur is a purely geometrical phenomenon that depends only on object depth, aperture size, and the focal length of the lens. However, lenses also exhibit some amount of blur due to the diffraction of light from the aperture. While defocus blur is directly proportional to aperture size, diffractive blur is inversely proportional to the aperture size. Therefore, any attempt to remove one type of blur will increase the other.

(a) When to use motion deblurring.
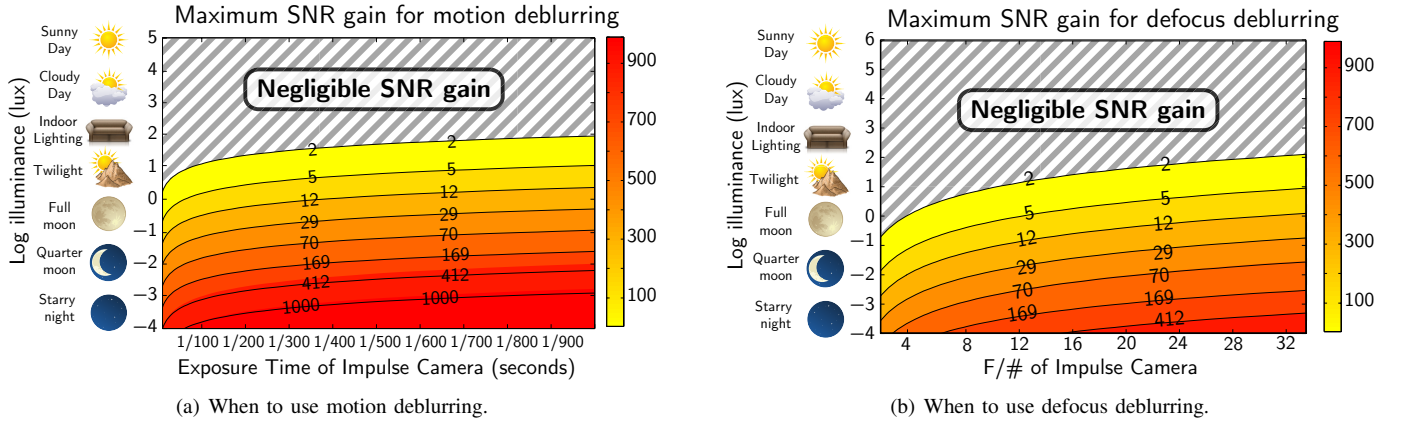

(b) When to use defocus deblurring.

Fig. 8. **Performance of motion and defocus deblurring.** The average signal level is calculated using Eq. 23 with the parameters outlined in Section VI-A. (a) Contour plot of the SNR gain bound versus the illuminance and the exposure time of the impulse camera. (b) Contour plot of the SNR gain bound versus the illuminance and the $F/\#$ of the impulse camera. For both motion and defocus deblurring, the SNR gain is always negligible when the illuminance is greater than $125\,lux$ (typical indoor lighting).

For the impulse camera, when diffraction blur size is larger than a pixel, the only option for removing blur is to use an EDOF technique. This is a particularly important problem when considering cameras with small pixel sizes. In this case, EDOF techniques provide a functionality which cannot be achieved using impulse imaging.

Task-Specific Imaging: Our aim in this paper is to analyze the performance of CI insofar as the final goal is to capture high quality images. If a further task is to be performed on the captured images (e.g., tracking, face recognition, intrusion detection), reconstruction algorithms can benefit from task-specific priors (as opposed to priors based on natural image statistics). Moreover, in this case, the performance should be evaluated in terms of task-specific metrics. While such task-specific priors and image quality metrics are beyond the scope of this paper, we believe they form an interesting direction for future research.

VIII. ACKNOWLEDGEMENTS

APPENDIX A
CODING FOR INVARIANCE

For shift-invariant systems, the image formation of Eq. 1 can be described in the Fourier domain as

$$G(\omega_x, \omega_y) = H(\omega_x, \omega_y)F(\omega_x, \omega_y) + \Psi(\omega_x, \omega_y), \quad (24)$$

where ω_x, ω_y are continuous valued spatial frequencies given in units of $1/pixels$, F is the focused image, H is the optical transfer function (OTF) of the camera, Ψ is the noise, and G is the captured image. An estimate of the focused image can be found as

$$F^*(\omega_x, \omega_y) = \frac{G(\omega_x, \omega_y)}{H(\omega_x, \omega_y)}, \quad (25)$$

and the expected MSE can be written as

$$MSE = E\left[||F^*(\omega_x, \omega_y) - F(\omega_x, \omega_y)||^2\right] \quad (26)$$

$$= \sigma^2 \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \frac{1}{||H(\omega_x, \omega_y)||^2} d\omega_x d\omega_y, \quad (27)$$

where E denotes expectation w.r.t. the noise Ψ.

Motion Invariant Blur. Let S_m be the maximum speed of objects in the scene measured in $pixels/sec$. An impulse camera will remove motion blur by setting the exposure time T so that the maximum blur size is equal to one pixel and $S_m T = 1$. Cho *et al.* [11] derived an upper bound on the best possible motion-invariant MTF that can be achieved for 2D motion [5]:

$$||H_m(\omega_x)||^2 \leq \frac{C}{2S_m T \sqrt{\omega_x^2 + \omega_y^2}}. \quad (28)$$

Substituting Eq. 28 into Eq. 27 gives a lower bound on the MSE for any motion-invariant camera:

$$MSE_m \geq \frac{\sigma_m^2 \left(\sqrt{2} + \operatorname{asinh}(1)\right)}{3C} \geq \frac{\sigma_m^2}{2C}. \quad (29)$$

Substituting Eq. 29 into Eq. 5 results in the bound expressed by Eq. 12.

Defocus Invariant Blur. In the same way that motion blur depends on the speed S_m, defocus blur depends on the defocus parameter S_d. Following the notation in Levin *et al.* [32], $S_d = 2\frac{d_{max} - d_{min}}{d_{max} + d_{min}}$, where d_{max} and d_{min} are the maximum and minimum depths of objects in the scene, respectively.

---

[5]The bound derived by Cho *et al.* assumes $H_m(0,0) = T$. We assume $H_m(0,0) = C$. Both conditions give the same SNR gain.

An impulse camera will remove defocus blur by reducing the aperture width A (measured in pixels) so that the maximum defocus blur size is equal to one pixel and $S_d A = 1$.

Baek [1] showed that the focal sweep camera is nearly optimal at simultaneously maximizing transfer efficiency and minimizing depth-invariance. Thus, the performance of focal sweep will closely approximate the best possible performance of any technique that produces depth-invariant blur. Levin et al. derived an approximate expression for the focal sweep MTF [32] [6]

$$\|H_{fs}(\omega_x, \omega_y)\|^2 \approx \frac{C \max(\omega_x^2, \omega_y^2)}{A^2 S_d^2} \qquad (30)$$

Substituting Eq. 30 into Eq. 27 gives a lower bound on the MSE for any depth-invariant camera:

$$MSE_d \geq \frac{\sigma_m^2}{2C}. \qquad (31)$$

Substituting Eq. 31 into Eq. 5 results in the bound expressed by Eq. 12.

APPENDIX B
LIGHTING CONDITIONS

For a Lambertian scene with average reflectance R that is lit by a source with illuminance I_{src} (given in units of lux), the average illuminance falling on the detector (also in units of lux) is [24]

$$I_{det} = \frac{1}{4} \frac{1}{F/\#^2} E_{src} R. \qquad (32)$$

Given a quantum efficiency q, an exposure time of $t \, seconds$, and a pixel size of $\Delta \, meters$, the average energy in $joules$ collected by a pixel is [16]

$$E = K \frac{1}{4} \frac{1}{F/\#^2} I_{src} R q \Delta^2 t, \qquad (33)$$

where $K = 1/680 \, watts/lumen$ is the conversion factor between photometric and radiometric units when the detector spectral response is matched to the photopic spectral response of the standard human observer. The energy in $joules$ of a single photon is given by $\hbar c / \lambda$, where $\hbar$ is Planck's constant, and c is the speed of light. The average number of photons collected by a pixel is then

$$J = K \frac{\lambda}{4 \hbar c} \frac{1}{F/\#^2} I_{src} R \, q \, \Delta^2 \, t. \qquad (34)$$

Assuming a mean wavelength of $\lambda = .55 \mu m$, the average number of photons becomes

$$J = 10^{15} \frac{1}{F/\#^2} I_{src} R \, q \, \Delta^2 \, t. \qquad (35)$$

---

[6]The derivation by Levin *et al.* assumes $H_{fs}(0,0) = A^2$. We assume $H_{fs}(0,0) = C$. Both conditions give the same SNR gain.

REFERENCES

[1] J. Baek. Transfer Efficiency and Depth Invariance in Computational Cameras. In *ICCP*, 2010.
[2] R. Baer, W. Holland, J. Holm, and P. Vora. A comparison of primary and complementary color filters for ccd-based digital photography. In *SPIE Electronic Imaging Conference*. Citeseer, 1999.
[3] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *IJCV*, 35, 1999.
[4] M. Ben-Ezra, A. Zomet, and S. Nayar. Jitter camera: high resolution video from a low resolution detector. In *CVPR*, 2004.
[5] M. Ben-Ezra, A. Zomet, and S. Nayar. Video super-resolution using controlled subpixel detector shifts. *PAMI*, 2005.
[6] J. Bioucas-Dias and M. Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *TIP*, 16(12), 2007.
[7] E. Caroli, J. Stephen, G. Cocco, L. Natalucci, and A. Spizzichino. Coded aperture imaging in x-and gamma-ray astronomy. *Space Science Reviews*, 45(3), 1987.
[8] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *CVPR*, 2011.
[9] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1), 2004.
[10] W. Chi and N. George. Electronic imaging using a logarithmic asphere. *Optics Letters*, 2001.
[11] T. Cho, A. Levin, F. Durand, and W. Freeman. Motion blur removal with orthogonal parabolic exposures. In *ICCP*, 2010.
[12] R. N. Clark. Digital camera sensor performance summary. "http://www.clarkvision.com/articles/digital.sensor.performance.summary/#model", 2011.
[13] O. Cossairt and S. K. Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *ICCP*, 2010.
[14] O. Cossairt, C. Zhou, and S. K. Nayar. Diffusion Coding Photography for Extended Depth of Field. In *SIGGRAPH*, 2010.
[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *TIP*, 16(8), 2007.
[16] C. DeCusatis. *Handbook of applied photometry*. Amer Inst of Physics, 1997.
[17] J. E. R. Dowski and W. T. Cathey. Extended depth of field through wave-front coding. *Applied Optics*, 1995.
[18] M. Gaubatz. Metrix mux visual quality assessment package. "http://foulard.ece.cornell.edu/gaubatz/metrix_mux/", 2011.
[19] F. Guichard, H. Nguyen, R. Tessières, M. Pyanet, I. Tarchouna, and F. Cao. Extended depth-of-field using sharpness transport across color channels. In *Digital Photography V*, volume 7250. SPIE, 2009.
[20] Q. Hanley, P. Verveer, and T. Jovin. Spectral imaging in a programmable array microscope by hadamard transform fluorescence spectroscopy. *Applied Spectroscopy*, 53(1), 1999.
[21] M. Harwit and N. Sloane. Hadamard transform optics. *New York: Academic Press*, 1979.
[22] S. Hasinoff, K. Kutulakos, F. Durand, and W. Freeman. Time-constrained photography. In *ICCV*, 2009.
[23] G. Häusler. A method to increase the depth of focus by two step image processing. *Optics Communications*, 1972.
[24] B. Horn. *Robot vision*. The MIT Press, 1986.
[25] R. Horstmeyer, G. W. Euliss, R. A. Athale, and M. Levoy. Flexible multimodal camera using a light field architecture. In *ICCP*, 2009.
[26] I. Ihrke, G. Wetzstein, and W. Heidrich. A Theory of Plenoptic Multiplexing. In *CVPR*, 2010.
[27] S. Kuthirummal, H. Nagahara, C. Zhou, and S. K. Nayar. Flexible Depth of Field Photography. In *PAMI*, 2010.
[28] R. F. L. Fei-Fei and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In *CVPR*, 2004.
[29] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin. Shield fields: modeling and capturing 3d occluders. In *SIGGRAPH*, 2008.
[30] D. H. Lee, I. S. Kweon, and R. Cipolla. Single lens stereo with a biprism. In *IAPR International Workshop on Machine Vision and Applications*, 1998.
[31] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*, 2007.
[32] A. Levin, S. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4d frequency analysis of computational cameras for depth of field extension. In *SIGGRAPH*, 2009.
[33] A. Levin, P. Sand, T. Cho, F. Durand, and W. Freeman. Motion-invariant photography. In *SIGGRAPH*, 2008.
[34] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *SIGGRAPH*, 2008.
[35] J. Ojeda-Castaneda, J. E. A. Landgrave, and H. M. Escamilla. Annular phase-only mask for high focal depth. *Optics Letters*, 2005.

[36] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *PAMI*, 23, 2001.
[37] A. Pentland. A new sense for depth of field. *PAMI*, 1987.
[38] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *SIGGRAPH*, 2006.
[39] N. Ratner and Y. Schechner. Illumination multiplexing within fundamental limits. In *CVPR*, 2007.
[40] N. Ratner, Y. Schechner, and F. Goldberg. Optimal multiplexed sensing: bounds, conditions and a graph theory link. *Optics Express*, 15, 2007.
[41] Y. Schechner, S. Nayar, and P. Belhumeur. A theory of multiplexed illumination. In *ICCV*, 2003.
[42] Y. Schechner, S. Nayar, and P. Belhumeur. Multiplexing for optimal lighting. *PAMI*, 2007.
[43] H. Sheikh and A. Bovik. Image information and visual quality. *TIP*, 15(2), 2006.
[44] G. Skinner. X-ray imaging with coded masks. *Scientific American;(USA)*, 259(2), 1988.
[45] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *CVPR*, 1988.
[46] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (tombo): concept and experimental verification. *Applied Optics*, 40(11), 2001.
[47] Y. Tendero. *Mathematical Theory of the Flutter Shutter*. PhD thesis, Ecole normale suprieure de Cachan.
[48] B. Trifonov, D. Bradley, and W. Heidrich. Tomographic reconstruction of transparent objects. In *EGSR*, 2006.
[49] 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.
[50] Z. Wang and A. Bovik. A universal image quality index. *Signal Processing Letters*, 9(3), 2002.
[51] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *TIP*, 13(4), 2004.
[52] Wikipedia. lux. "http://en.wikipedia.org/wiki/Lux", 2011.
[53] K. Yamazawa, Y. Yagi, and M. Yachida. Omnidirectional imaging with hyperboloidal projection. In *International Conference on Robots and Systems*, 1993.
[54] L. Zhang, A. Deshpande, and X. Chen. Denoising versus Deblurring: HDR techniques using moving cameras. In *CVPR*, 2010.
[55] C. Zhou, O. Cossairt, and S. K. Nayar. Depth from diffusion. In *CVPR*, 2010.
[56] C. Zhou, S. Lin, and S. Nayar. Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. *IJCV*, 2011.
[57] C. Zhou and S. Nayar. What are Good Apertures for Defocus Deblurring? In *ICCP*, 2009.