Diffuse Imaging: Creating Optical Images with Unfocused Time-Resolved Illumination and Sensing

Ahmed Kirmani, Haris Jeelani, Vahid Montazerhodjat, and Vivek K Goyal, *Senior Member, IEEE*

***Abstract*—Conventional imaging uses steady-state illumination and light sensing with focusing optics; variations of the light field with time are not exploited. We develop a signal processing framework for estimating the reflectance f of a Lambertian planar surface in a known position using omnidirectional, time-varying illumination and unfocused, time-resolved sensing in place of traditional optical elements such as lenses and mirrors. Our model associates time sampling of the intensity of light incident at each sensor with a linear functional of f. The discrete-time samples are processed to obtain ℓ^2-regularized estimates of f. Improving on previous work, using non-impulsive, bandlimited light sources instead of impulsive illumination significantly improves signal-to-noise ratio (SNR) and reconstruction quality. Our simulations suggest that practical diffuse imaging applications may be realized with commercially-available temporal light intensity modulators and sensors used in standard optical communication systems.**

***Index Terms*—light transport, optical imaging, reflectance, resolution, quantization, sampling, time of flight.**

I. Introduction

IMAGING is distinguished from other forms of sensing by having the goal of producing an *image*—a representation in one-to-one spatial correspondence with an object or scene. Traditional imaging involves illumination of the scene with a light source whose intensity does not vary with time. The image, or the reflectance pattern on scene surfaces, is acquired by using a lens to focus the reflected light on a two-dimensional (2D) array of light meters. This letter expounds a *diffuse imaging* framework through which a focused image is produced without optical focus; instead, the image is reconstructed from time-resolved measurements of light intensity in response to time-varying scene illumination. Diffuse imaging opens up possibilities for forming images without lenses and mirrors and enables practical implementations of challenging new applications such as imaging occluded scenes [1].

The use of high-speed sensing in photography is associated with effectively stopping motion [2]. We use time-resolved sensing differently: to differentiate among paths of different lengths from light source to scene to sensor. Thus, like in time-of-flight range measurement systems [3], [4], we are exploiting that the speed of light is finite. However, rather than using the duration of delay to infer distances, we use temporal variation in the *intensity* of measured light to infer scene reflectance. This is achieved through an unmixing of the reflectance information that is linearly combined at the sensor because distinct optical paths may have equal path length.

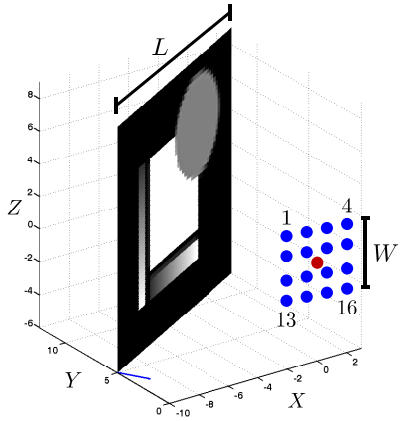


Fig. 1. An omnidirectional source (red point) illuminates a planar scene, and the reflected light is measured at a 4-by-4 array of sensors (blue points).

Diffuse imaging involves only the use of diffuse or omnidirectional light sources and omnidirectional light sensors. In conventional photography, the light source is omnidirectional but the sensor array is focused at the scene. A dual configuration is also possible in which the sensing is omnidirectional and the light source is directed, with optical focusing [5].

Sections II, III, and IV expand on [1], especially by providing a precise signal processing model. Sections V and Section VI show how to improve upon the theoretical and practical limitations of using impulsive illumination and ultrafast sensing as proposed in [1]. In particular, simulations suggest that diffuse imaging can be applied to build practical imaging systems using non-ultrafast, opto-electronic hardware used in standard optical communication systems.

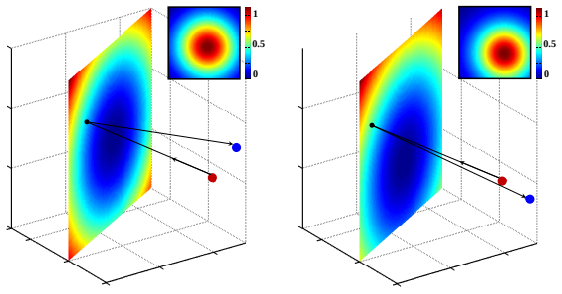
II. Scene Response to Impulsive Illumination

Consider the imaging scenario depicted in Fig. 1, with a planar surface to be imaged, a single, time-varying, monochromatic, omnidirectional illumination source, and omnidirectional time-resolved sensors indexed by $k \in \{1, 2, \ldots, K\}$. We assume that the position, orientation, and dimensions (L-by-L) of the planar surface are known. Estimation of these geometric parameters using diffuse illumination and time-resolved sensing was recently demonstrated [6]. Formation of an ideal gray scale image is the recovery of the reflectance

This material is based upon work supported in part by the National Science Foundation under Grant No. 0643836, the DARPA InPho program through the US Army Research Office award W911-NF-10-1-0404, the Texas Instruments Leadership University Program, and an Esther and Harold E. Edgerton Career Development Chair.

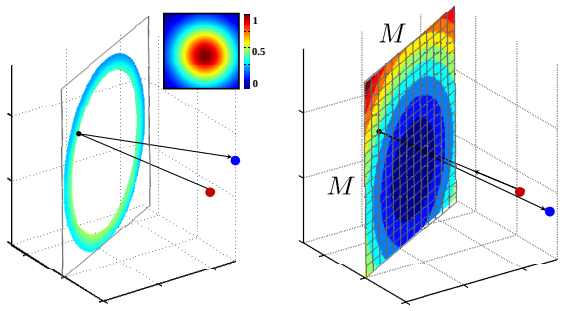
A. Kirmani, V. Montazerhodjat, and V. K. Goyal are with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: {akirmani,montazer,vgoyal}@mit.edu).

H. Jeelani was visiting MIT (e-mail: haris.jeelani@gmail.com).



(a) Sensor 4 (b) Sensor 16

Fig. 2. Main plots: Time delay from source to scene to sensor is a continuous function of the position in the scene $d_k(\mathbf{x})$. Insets: The attenuation $a_k(\mathbf{x})$ of the light is also a continuous function of the position in the scene (1).



(a) Sensor 4 (b) Sensor 16

Fig. 3. (a) A single measurement function $\varphi_{k,n}(\mathbf{x})$ when $h(t)$ is the box function (7). Inset is unchanged from Fig. 2(a). (b) When $h(t)$ is the box function (7), the measurement functions for a single sensor $\{\varphi_{k,n}(\mathbf{x})\}_{n=1}^{N}$ partition the plane into elliptical annuli. Coloring is by discretized delay using the colormap of Fig. 2. Also overlaid is the discretization of the plane of interest into an M-by-M pixel array.

pattern on the surface, which can be modeled as a 2D function $f : [0, L]^2 \to [0, 1]$. We assume the surface to be *Lambertian* so that its perceived brightness is invariant to the angle of observation [7]; incorporation of any known bidirectional reflectance distribution function would not add insight.

The light incident at Sensor k is a combination of the time-delayed reflections from all points on the planar surface. For any point $\mathbf{x} = (x_1, x_2) \in [0, L]^2$, let $d^{(1)}(\mathbf{x})$ denote the distance from illumination source to $\mathbf{x}$, and let $d_k^{(2)}(\mathbf{x})$ denote the distance from $\mathbf{x}$ to Sensor k. Then $d_k(\mathbf{x}) = d^{(1)}(\mathbf{x}) + d_k^{(2)}(\mathbf{x})$ is the total distance traveled by the contribution from $\mathbf{x}$. This contribution is attenuated by the reflectance $f(\mathbf{x})$, square-law radial fall-off, and $\cos(\theta(\mathbf{x}))$ to account for foreshortening of the surface with respect to the illumination, where $\theta(\mathbf{x})$ is the angle between the surface normal at $\mathbf{x}$ and a vector from $\mathbf{x}$ to the illumination source. Thus, when the intensity of the omnidirectional illumination is a unit impulse at time 0, denoted $s(t) = \delta(t)$, the contribution from point $\mathbf{x}$ is the light signal $a_k(\mathbf{x})\, f(\mathbf{x})\, \delta(t - d_k(\mathbf{x}))$, where we have normalized to unit speed of light and

$$a_k(\mathbf{x}) = \cos(\theta(\mathbf{x}))/\Big(d^{(1)}(\mathbf{x})\, d_k^{(2)}(\mathbf{x})\Big)^2. \tag{1}$$

Examples of distance functions and attenuation factors are shown in Fig. 2.

Combining contributions over the plane, the total light incident at Sensor k is

$$g_k(t) = \int_0^L \int_0^L a_k(\mathbf{x})\, f(\mathbf{x})\, \delta(t - d_k(\mathbf{x}))\, dx_1\, dx_2. \tag{2}$$

Thus, evaluating $g_k(t)$ at a fixed t amounts to integrating over $\mathbf{x} \in \mathbb{R}^2$ with $d_k(\mathbf{x}) = t$. Define the isochronal curve $C_k^t = \{\mathbf{x} : d_k(\mathbf{x}) = t\}$. Then

$$g_k(t) = \int a_k(\mathbf{x}(k, u)) f(\mathbf{x}(k, u))\, du \tag{3}$$

where $\mathbf{x}(k, u)$ is a parameterization of C_k^t with unit speed. The intensity $g_k(t)$ thus contains the contour integrals over C_k^t's of the desired function f. Each C_k^t is a level curve of $d_k(\mathbf{x})$; as illustrated in Fig. 2, these are ellipses.

III. Sampling the Scene Response

A digital system can use only samples of $g_k(t)$ rather than the continuous-time function itself. We now see how uniform sampling of $g_k(t)$ with a linear time-invariant (LTI) prefilter relates to linear functional measurements of f. This establishes the foundations of a Hilbert space view of diffuse imaging.

Suppose discrete samples are obtained at Sensor k with sampling prefilter $h_k(t)$ and sampling interval T_k:

$$y_k[n] = \big(g_k(t) * h_k(t)\big)\big|_{t=nT_k}, \qquad n = 1, 2, \ldots, N. \tag{4}$$

A sample $y_k[n]$ can be seen as a standard $\mathcal{L}^2(\mathbb{R})$ inner product between g_k and a time-reversed and shifted h_k [8]:

$$y_k[n] = \langle g_k(t),\, h_k(nT_k - t)\rangle. \tag{5}$$

Using (2), we can express (5) in terms of f using the standard $\mathcal{L}^2([0, L]^2)$ inner product:

$$y_k[n] = \langle f,\, \varphi_{k,n}\rangle \qquad \text{where} \tag{6a}$$

$$\varphi_{k,n}(\mathbf{x}) = a_k(\mathbf{x})\, h_k(nT_k - d_k(\mathbf{x})). \tag{6b}$$

Over a set of sensors and sample times, $\{\varphi_{k,n}\}$ will span a subspace of $\mathcal{L}^2([0, L]^2)$, and a sensible goal is to form a good approximation of f in that subspace.

For ease of illustration and interpretation, let $T_k = T$ and

$$h_k(t) = h(t) = \begin{cases} 1, & \text{for } 0 \le t \le T; \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

for all k, which corresponds to "integrate and dump" sampling. Now since $h_k(t)$ is nonzero only for $t \in [0, T]$, by (4) or (5), the sample $y_k[n]$ is the integral of $g_k(t)$ over $t \in [(n-1)T, nT]$. Thus, by (3), $y_k[n]$ is an a-weighted integral of f between the contours $C_k^{(n-1)T}$ and C_k^{nT}. To interpret this as an inner product with f as in (6), we see that $\varphi_{k,n}(x)$ is $a_k(\mathbf{x})$ between $C_k^{(n-1)T}$ and C_k^{nT} and zero otherwise. Fig. 3(a) shows a single representative $\varphi_{k,n}$. The functions $\{\varphi_{k,n}\}_{n\in\mathbb{Z}}$ for a single sensor have disjoint supports; their partitioning of the domain $[0, L]^2$ is illustrated in Fig. 3(b).

IV. Image Recovery using Linear Backprojection

To express an estimate $\hat{f}$ of the reflectance f, it is convenient to fix an orthonormal basis for a subspace of $\mathcal{L}^2([0, L]^2)$ and estimate the expansion coefficients in that basis. For an M-by-M pixel representation, let

$$\psi_{i,j}(\mathbf{x}) = \begin{cases} M/L, & \text{for } (i-1)L/M \le x_1 < iL/M, \\ & \quad (j-1)L/M \le x_2 < jL/M; \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

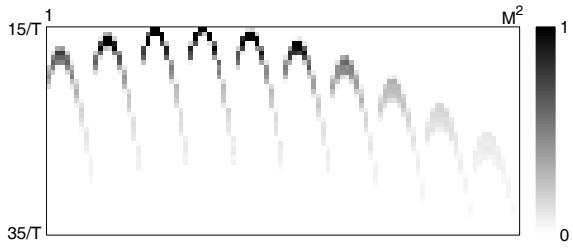


Fig. 4. Visualizing linear system representation A in (10). Contribution from Sensor 1 is shown. The matrix has 100 columns (since $M = 10$) and 40 rows (through choice of T). All-zero rows arising from times prior to first arrival of light and after the last arrival of light are omitted.

so that $\hat{f} = \sum_{i=1}^{M}\sum_{j=1}^{M} c_{i,j}\psi_{i,j}$ in the span of $\{\psi_{i,j}\}$ is constant on Δ-by-Δ patches, where $\Delta = L/M$.

For $\hat{f}$ to be consistent with the value measured by Sensor k at time n, we must have

$$y_k[n] = \langle \hat{f},\, \varphi_{k,n}\rangle = \sum_{i=1}^{M}\sum_{j=1}^{M} c_{i,j}\langle \psi_{i,j},\, \varphi_{k,n}\rangle. \quad (9)$$

Note that the inner products $\{\langle \psi_{i,j},\, \varphi_{k,n}\rangle\}$ exclusively depend on Δ, the positions of illumination and sensors, the plane geometry, the sampling prefilters $\{h_k\}_{k=1}^{K}$, and the sampling intervals $\{T_k\}_{k=1}^{K}$—not on the unknown reflectance of interest f. Hence, we have a system of linear equations to solve for the coefficients $\{c_{i,j}\}$. (In the case of basis (8), the coefficients are directly the pixel values.)

When we specialize to the box sensor impulse response (7) and basis (8), many inner products $\langle \psi_{i,j},\, \varphi_{k,n}\rangle$ are zero so the linear system is sparse. The inner product $\langle \psi_{i,j},\, \varphi_{k,n}\rangle$ is nonzero when reflection from the (i,j) pixel affects the light intensity at Sensor k within time interval $[(n-1)T, nT]$. Thus, for a nonzero inner product the (i,j) pixel must intersect the elliptical annulus between $C_k^{(n-1)T}$ and C_k^{nT}. With reference to Fig. 3(a), this occurs for a small number of (i,j) pairs unless M is small or T is large. The value of a nonzero inner product depends on the fraction of the square pixel that overlaps with the elliptical annulus and the attenuation factor $a_k(\mathbf{x})$.

To express (9) with a matrix multiplication, replace double indexes with single indexes (i.e., vectorize, or reshape) as

$$\mathbf{y} = A\,\mathbf{c} \quad (10)$$

where $\mathbf{y} \in \mathbb{R}^{KN}$ contains the data samples $\{y_k[n]\}$, the first N from Sensor 1, the next N from Sensor 2, etc.; and $\mathbf{c} \in \mathbb{R}^{M^2}$ contains the coefficients $\{c_{i,j}\}$, varying i first and then j. Then $\langle \psi_{i,j},\, \varphi_{k,n}\rangle$ appears in row $(k-1)N+n$, column $(i-1)M+j$ of $A \in \mathbb{R}^{KN\times M^2}$. Fig. 4 illustrates an example of the portion of A corresponding to Sensor 1 for the scene in Fig. 1.

Assuming that A has a left inverse (i.e., $\operatorname{rank}(A) = M^2$), one can form an image by solving (10). The portion of A from one sensor cannot have full column rank because of the collapse of information along elliptical annuli depicted in Fig. 3(a). Full rank can be achieved with an adequate number of sensors, noting that sensor positions must differ to increase rank, and—as demonstrated in Section VI—greater distance between sensor positions improves conditioning.

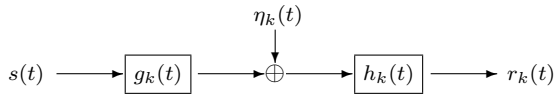
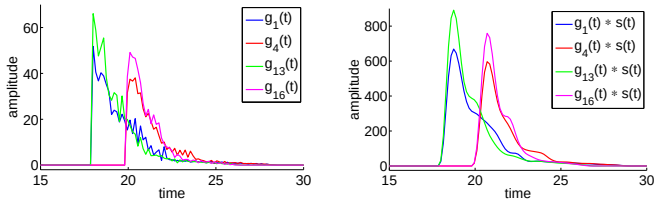


Fig. 5. Block diagram abstraction for signal sampled at Sensor k.



(a) (b)

Fig. 6. (a) Responses to impulsive illumination $g_k(t)$. (b) Responses to lowpass illumination $g_k(t) * s(t)$ are stronger and smoother.

V. Illumination Design

A Dirac impulse illumination is an abstraction that cannot be realized in practice. One can use expensive, ultrafast optical lasers that achieve Terahertz bandwidth as an approximation to impulsive illumination, as in experimental verification of our framework [1]. We show that non-impulsive illuminations are not only more practical for implementation, but they improve upon impulsive illumination for typical scenes and sensors.

Light transport is linear and time invariant. Hence, one can consider the effect of a general illumination intensity waveform $s(t)$ as the superposition of effects of constant illuminations over infinitesimal intervals. This superposition changes the light incident at Sensor k from $g_k(t)$ in (2) to $g_k(t) * s(t)$. Thus, the block diagram in Fig. 5 represents the signal at Sensor k, including its sampling prefilter and photodetector noise represented by $\eta_k(t)$. Except at very low flux, $\eta_k(t)$ is modeled well as single-independent, zero-mean, white and Gaussian. The noise variance σ^2 depends on the device physics and assembly; our simulations use $\sigma = 0.1$.

A typical natural scene f has a good bandlimited approximation. Integration over elliptical contours C_k^t further smooths the signal. Plotted in Fig. 6(a) are continuous-time responses $g_k(t)$ corresponding to Fig. 1. Since these have sharp decay with frequency, $s(t)$ is best chosen to be lowpass to put signal energy at frequencies most present in $g_k(t)$.

VI. Simulations

In [1], high-bandwidth illumination and sampling were used under the assumption that these would lead to the highest reconstruction quality. However, impulsive illumination severely limits the illumination energy, leading to poor SNR, especially due to the radial fall-off attenuations in (1).

Here we compare impulsive and lowpass illumination. All image reconstructions are obtained with (10) regularized by the ℓ^2 norm of the discrete Laplacian, with regularization parameter optimized for ℓ^2 error. This conventional technique for backprojection [9] mildly promotes smoothness of the reconstruction; additional prior information, such as sparsity with suitable $\{\psi_{i,j}\}$, is beneficial but would obscure the present contribution. Results are for $M = 50$ and several values of sampling period T and sensor array extent W.

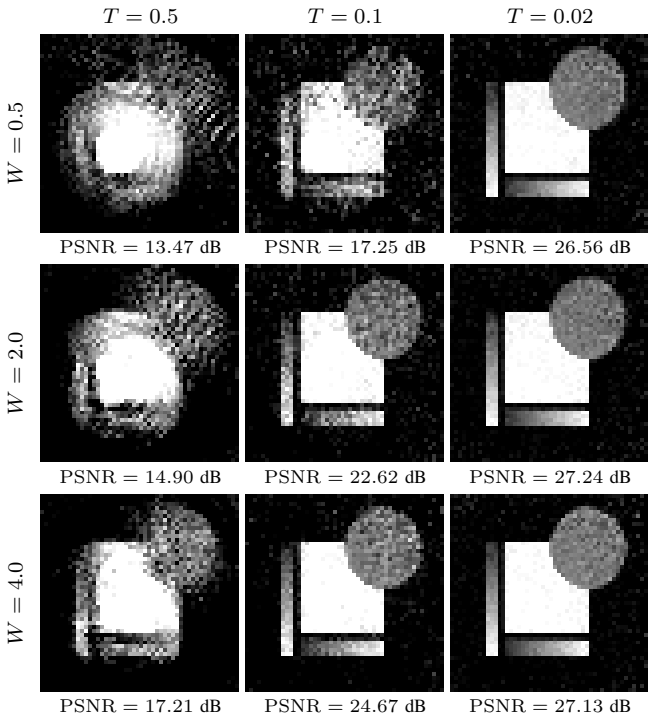


Fig. 7. Simulated reconstructions with impulsive illumination and 4-by-4 array of sensors and true scene as in Fig. 1.

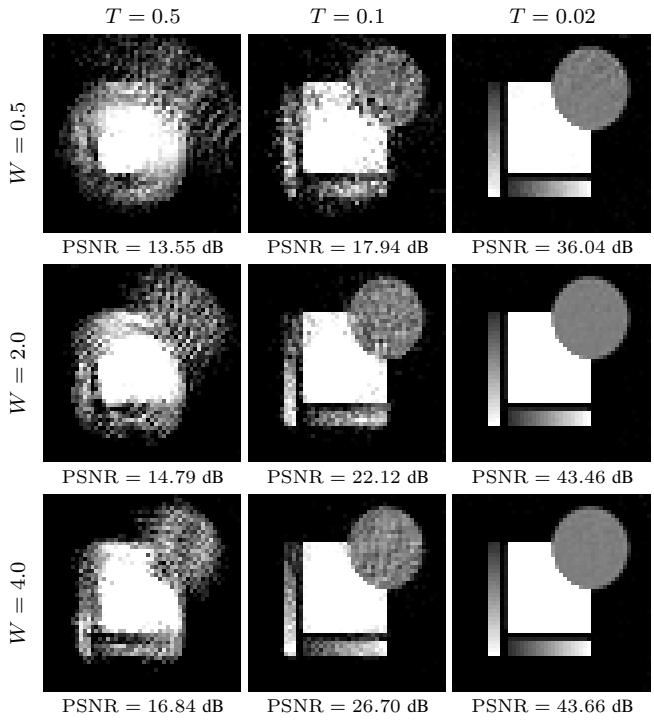


Fig. 8. Simulated reconstructions with lowpass illumination and 4-by-4 array of sensors and true scene as in Fig. 1.

In direct analogy with [1], we simulated short-pulsed, high-bandwidth illumination using a square-wave source with unit amplitude and time width equal to one-fifth of T. Fig. 7 shows the results. Reconstruction with good spatial resolution is indeed possible, and the conditioning improves as W increases and as T decreases.

Using non-impulsive, low-bandwidth illumination, we show that high SNR can be achieved while improving the reconstruction resolution. We chose $s(t)$ to be the truncated impulse response of a third-order Butterworth filter, with again a unit amplitude. As illustrated in Fig. 6(b), this choice of $s(t)$ produces a much stronger scene reflection and hence improves the SNR at the detector. Note that the critical lowpass portion of $g_k(t)$ is preserved and furthermore amplified. Fig. 8 shows the resulting improvements in reconstructed images; we can infer that the improvement in SNR is coming without excessive loss of matrix conditioning. Hence, the choice of a non-impulsive illumination is not only practical but demonstrably better in terms of image reconstruction quality.

VII. Summary

The proof-of-concept experiments in [1] show that diffuse imaging can succeed in forming image reconstructions. In this letter we have used signal processing abstractions to show that using lowpass time-varying illumination instead of impulsive sources greatly improves the SNR and reconstruction quality. Our work inspires detailed study of how pixel size, sensor locations, sampling prefilters, and sampling rates affect the conditioning of the inverse problem.

Assigning dimensions to our simulation enables the specification of required device capabilities. Suppose the 50-by-50 pixel image reconstruction with $W = 4$ and $T = 0.1$ shown in Fig. 8 corresponds to a physical planar scene of edge length 15 m imaged from 10 m away using an array with 4 m extent. Then the illumination bandwidth is about 375 MHz and the sensor sampling frequency is about 750 MHz. The total energy output of the source is about 44 mJ. Compared to the 40 THz bandwidth laser (10.5 nJ per pulse) and 500 GHz streak camera used in [1], our simulations show that diffuse imaging can be implemented with practical opto-electronic hardware used in optical communication.

References

[1] A. Kirmani, A. Velten, T. Hutchison, M. E. Lawson, V. K. Goyal, M. Bawendi, and R. Raskar, "Reconstructing an image on a hidden plane using ultrafast imaging of diffuse reflections," May 2011, submitted.

[2] H. E. Edgerton and J. R. Killian, Jr., *Flash! Seeing the Unseen by Ultra High-Speed Photography*. Boston, MA: Hale, Cushman and Flint, 1939.

[3] B. Schwarz, "LIDAR: Mapping the world in 3D," *Nature Photonics*, vol. 4, no. 7, pp. 429–430, Jul. 2010.

[4] S. Foix, G. Alenyà, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.

[5] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, "Dual photography," *ACM Trans. Graphics*, vol. 24, no. 3, pp. 745–755, Jul. 2005.

[6] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using transient imaging," in *Proc. IEEE 12th Int. Conf. on Computer Vision*, Kyoto, Japan, Sep.–Oct. 2009.

[7] M. Oren and S. K. Nayar, "Generalization of the Lambertian model and implications for machine vision," *Int. J. Comput. Vis.*, vol. 14, no. 3, pp. 227–251, Apr. 1995.

[8] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[9] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.