

A second-order theory of texture for depth from focus

Sreekar Sai Ranganathan[✉] and Ioannis Gkioulekas[✉]

Carnegie Mellon University, Pittsburgh PA, USA
{ssrangan, igkioule}@andrew.cmu.edu

Abstract. We present a theory of textured appearance of optically rough surfaces based on wave optics, emphasizing the role of texture for passive depth from focus. Our theory shows that even surfaces that traditional computer vision would consider textureless can produce textured appearance, due to subjective speckle from surface microgeometry. We analyze the properties of this *second-order texture*, and show that we can enhance its contrast under natural ambient lighting by simply using a narrowband spectral filter. Doing so results in dramatic improvements in passive depth reconstruction of seemingly textureless scenes, as we demonstrate through extensive theory, simulations, and real-world experiments.

Keywords: passive depth · texture · wave optics · depth from focus

1 Introduction

Are textureless scenes recoverable? Sundaram and Nayar [1] posed this fundamental question to the computer vision community nearly three decades ago. Their analysis suggested that little geometrical information exists in images of a textureless scene taken by a standard camera, thus greatly hindering *passive depth sensing* of such scenes. In the succeeding years, computer vision has made great strides towards developing methods for passive depth sensing, leveraging focus or parallax cues that can be inferred from texture even under uncontrolled ambient illumination. Despite these advances, the assumption has remained that passive depth sensing of textureless scenes is difficult, if not impossible. In this paper, we challenge this assumption: we show that merely adding a narrowband spectral filter in front of an otherwise standard lens-based camera allows using depth from focus (DFF) to recover textureless objects (Figure 1).

Understanding this surprising finding requires examining another fundamental question in computer vision: *What is texture?* The conventional definition requires spatial variations of surface reflectance or normal at a scale resolvable by the camera at the working imaging magnification, manifesting as spatially varying intensities in focused images of the surface. Thus, a surface with spatially constant bidirectional reflectance distribution function (BRDF)—the macroscopic description of reflectance—would be considered “textureless”.

We challenge this definition by revisiting classical surface appearance models based on wave optics. Our analysis predicts that even such “textureless” surfaces will produce non-constant in-focus images showing *subjective speckle*—high-frequency intensity variations due to diffraction at surface microgeometry that

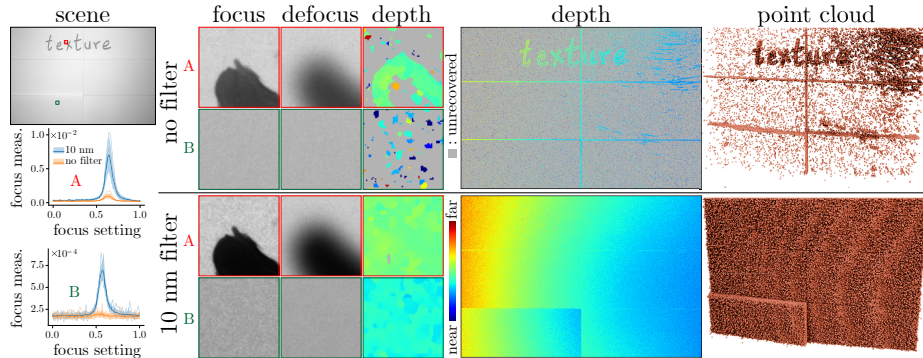


Figure 1: Computer vision traditionally considers depth recovery of scenes with sparse texture features impossible with passive methods such as depth from focus (top). We show that *simply adding a narrowband spectral filter to the camera* makes passive depth recovery of such scenes possible (bottom). We explain this surprising finding by developing a theory of *second-order texture*—a form of subjective speckle whose in-focus contrast is enhanced with decreasing spectral bandwidth (crops to the left).

the camera cannot resolve. Understanding this effect requires analysis of second-order intensity statistics, instead of just the mean (first-order moment) intensity predicted by classical appearance models. We term this effect *second-order texture*, to distinguish it from the classical *first-order texture*.

Second-order texture will be useful for DFF only if it has high-enough contrast to overcome sensor noise. Prior work assumes that significant subjective speckle is possible only under coherent illumination or at microscopic magnifications. We show that using a narrowband spectral filter (10–100 nm) results in high-contrast second-order texture under typical passive computer vision conditions—outdoor and indoor ambient lighting, macroscopic magnifications. Using such a filter also reduces incident flux, necessitating increased exposure time to maintain signal-to-noise ratio (SNR). We show that, even under severe underexposure (3–4 stops), second-order texture is discernible from noise. On the whole, using a narrowband spectral filter can dramatically improve DFF performance (Figure 2).

To present our findings, first, we provide background on DFF, and characterize depth recoverability conditions as a function of texture contrast and SNR (§ 3). Next, we develop our second-order theory of texture, and analyze the factors impacting the contrast of second-order texture (§ 4). Then, we use this analysis to assess, through theory and simulations, how DFF performance varies with filter spectral bandwidth and exposure time (§ 5). Last, we perform DFF experiments under ambient indoor and outdoor illumination that validate our findings (§ 6). The supplemental PDF includes all proofs and additional experimental results. The project website¹ includes interactive visualizations, code, and datasets.

2 Related work

Passive depth sensing in computer vision. Passive methods use *parallax* or *focus* to infer depth under ambient illumination. Parallax methods such as stereo

¹ <https://imaging.cs.cmu.edu/second-order-texture>

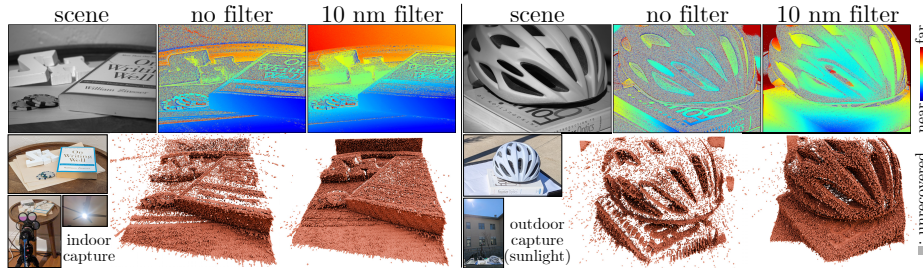


Figure 2: We demonstrate fine-scale depth recovery of objects with sparse texture features under ambient lighting indoors (ceiling lights, left) and outdoors (sunlight, right). In both cases, adding a spectral filter dramatically improves DFF performance.

[2–5] infer depth by triangulating correspondences across multi-view images. Among focus methods, depth from focus [6–11] infers depth by assessing per-pixel focus in a dense focal stack. Depth from defocus [12–16] uses two images at different focus settings to estimate defocus blur and thus depth. Other methods change both focus and aperture (confocal stereo [17]), vary focus differentially (focal flow [18, 19]), or engineer high-frequency defocus blur (coded aperture [20–24]). Focus methods can be analyzed as stereo with a baseline equal to the lens aperture [25], a relationship used in lightfield [26, 27] and dual-pixel [28–31] methods. Unfortunately, these methods fundamentally require *texture*, to either establish correspondence or assess focus, and are unreliable in textureless scenes.

Passive interferometric imaging. Several methods leverage the weak coherence of ambient illumination through *interferometry*—correlation measurements of superimposed light waves—for passive 3D imaging tasks. *Incoherent digital holography* methods take interferometric measurements to create holographic scene representations [32–37]. Other such methods can perform 3D localization for occluded imaging [38–42]. Closer to our work, interferometry has been used for passive depth sensing—through focus [43], parallax [44], or time-of-flight [45]—of even textureless scenes. Unfortunately, these methods require complex and sensitive optical systems, limiting their practical utility. Like these methods, ours achieves textureless passive depth sensing by leveraging weak coherence effects—subjective speckle from ambient illumination—but it does so through just a very simple modification to a standard depth-from-focus system—mounting a narrowband spectral filter in front of the lens.

Speckle imaging. Speckle has been extensively studied in optics since the invention of the laser. We refer to dedicated textbooks [46–48], and limit our discussion to literature in vision and graphics. Computer vision methods have used speckle for tamper detection [49], motion tracking [50–52], vibrometry [53, 54], imaging around corners [55], and through occluders [56–58]. All these methods are *active*, introducing coherent illumination into the scene to induce speckle. These applications have also motivated work on speckle rendering [59–63]. Lastly, speckle modeling is at the foundation of classical appearance models for rough surfaces [64–67], a foundation we revisit to develop our theory.

3 Depth from focus & recoverability conditions

We begin with background on *depth from focus* (DFF) and an analysis of recoverability conditions, focusing on the role of texture and sensor noise. DFF uses a *focal stack* of images, captured by stepping the focusing depth of the camera lens. We denote this stack as $\{\tilde{I}_j(\vec{p})\}_{j=1}^J$, where $\tilde{I}_j(\vec{p})$ is the sensor’s intensity measurement at the pixel $\vec{p}$ and focus index j —we use 2D vectors $\vec{p}$ to indicate pixel center locations, and tildes to highlight quantities impacted by sensor noise, such as the noisy intensity $\tilde{I}$ and its functionals. Depth recovery requires determining for each pixel $\vec{p}$ the (most) in-focus index $j_{\text{if}}(\vec{p})$.

DFF estimates $j_{\text{if}}(\vec{p})$ by finding the image that maximizes a measure of spatial intensity variation (or “sharpness”) for a patch $P_T(\vec{p})$ of T pixels (*e.g.*, $\sqrt{T} \times \sqrt{T}$ -square) around $\vec{p}$. Several such *focus measures* are available [68]; we consider the *squared sample coefficient of variation* $\tilde{c}_j^2(\vec{p}) := \tilde{s}_j^2(\vec{p})/\tilde{m}_j^2(\vec{p})$ computed from the *sample mean* and *sample variance* of gray patch intensities,

$$\tilde{m}_j(\vec{p}) := \frac{1}{T} \sum_{\vec{p}' \in P_T(\vec{p})} \tilde{I}_j(\vec{p}'), \quad \tilde{s}_j^2(\vec{p}) := \frac{1}{T-1} \sum_{\vec{p}' \in P_T(\vec{p})} \left(\tilde{I}_j(\vec{p}') - \tilde{m}_j(\vec{p}) \right)^2. \quad (1)$$

We use $\tilde{c}_j^2(\vec{p})$ because it simplifies analysis, but our findings extend to other common focus measures (§ 6). DFF estimates $j_{\text{if}}(\vec{p})$ as $j^*(\vec{p}) := \operatorname{argmax}_j \tilde{c}_j^2(\vec{p})$. We refer to $\tilde{c}_j^2(\vec{p})$ as simply *sample contrast*, and omit $\vec{p}$ where convenient.

The justification for estimating j_{if} as j^* is that the contrast $\tilde{c}_j^2$ will be maximal at the in-focus index because defocus acts as a low-pass filter. This justification is valid under two conditions [9, 68]:

1. *Texture condition*: The in-focus patch $P_T(\vec{p})$ has appreciable *texture*, *i.e.*, spatial intensity variations, in the absence of noise.
2. *Signal-to-noise ratio condition*: The sensor noise is sufficiently low so that intensity variations due to texture are discernible from those due to noise.

These conditions are affected by scene appearance (surface reflectance, lighting), imaging conditions (lens aperture, exposure, sensor noise), and algorithmic choices (patch size T). In the rest of this section, we characterize the impact and interplay of these factors. We first distinguish between the measured noisy intensities $\tilde{I}_j$, and the corresponding *noise-free intensities* I_j , which we define formally below. We also define the *noise-free contrast*, *sample variance*, and *mean* c_j^2 , s_j^2 , and m_j (respectively) analogously, and use them to define a notion of recoverability.

Definition 1: α -recoverability

For any $\alpha \in [0, 1]$, we say that the in-focus index j_{if} is α -recoverable when $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2) \leq \alpha$, where: 1. $\tilde{c}_{\text{if}}^2$ is the focus measure at j_{if} , and 2. $\tilde{c}_{\text{of}}^2$ is the focus measure at an out-of-focus index $j_{\text{of}} \neq j_{\text{if}}$ with sufficient defocus such that noise-free intensities are nearly identical, thus $c_{\text{of}}^2 \approx 0$.

We use α -recoverability as a weaker, but *analytically tractable*, proxy of true recoverability to formalize the interplay of texture and noise (Proposition 2).

We next characterize the statistics of noise-free and noisy intensities at j_{if} and j_{oof} —denoted as I_{if} and I_{oof} for the noise-free, and analogously for the noisy case.

Texture model. The in-focus noise-free intensities $I_{\text{if}}(\vec{p})$ are themselves random variables described by a statistical texture model. For weakly textured surfaces, as we define them in § 4.2, it suffices to consider the mean μ_I and variance σ_I^2 of their stationary pointwise distribution. From the definition of j_{oof} in Definition 1, the noise-free intensities I_{oof} will all equal μ_I and have zero variance.

Noise model. We can relate noisy to noise-free intensities at both $j \in \{j_{\text{if}}, j_{\text{oof}}\}$ using the standard Poisson–Gaussian model [69], $\tilde{I}_j = \min(I_j + n_j, I_{\text{max}})$, where: 1. I_j is proportional to the *incident flux* Φ_j and *exposure time* t , $I_j := \Phi_j t \zeta / g$, implying that $\mu_\Phi := \mathbb{E}[\Phi] = \mu_I g / t \zeta$. 2. The noise term n_j is a random variable combining photon shot (Poisson) noise and read (Gaussian) noise. It has mean $\mathbb{E}[n_j | I_j] = 0$ and signal-dependent variance $\mathbb{V}[n_j | I_j] = I_j / g + \sigma_{\text{read}}^2 / g^2$, where $\sigma_{\text{read}}^2 := \sigma_{\text{pre}}^2 + g^2 \sigma_{\text{post}}^2$. 3. The constant ζ depends on quantum efficiency, g on ISO, σ_{pre} and σ_{post} on readout circuitry, and I_{max} on the saturation limit. This model assumes for simplicity no dark current or discretization, assumptions we revisit in § 5. The marginal variance of noise n_j thus equals $\sigma_n^2 := \mathbb{V}[n_j] = \mu_I / g + \sigma_{\text{read}}^2 / g^2$, using the fact that the noise-free mean intensity equals μ_I at j_{if} and j_{oof} .

Recoverability condition. Using these texture and noise models, we can define:

$$\kappa_I^2 := \frac{\sigma_I^2}{\mu_I^2}, \quad \kappa_n^2 := \frac{\sigma_n^2}{\mu_I^2} = \frac{1}{t \zeta \mu_\Phi} \left(1 + \frac{\sigma_{\text{read}}^2}{t \zeta \mu_\Phi} \right). \quad (2)$$

The *squared texture contrast* κ_I^2 and *reciprocal squared SNR* κ_n^2 measure texture and noise strength. We can relate them to recoverability as follows.

Proposition 2: Necessary and sufficient condition for α -recoverability

For any $\alpha \in [0, 1]$, the in-focus index j_{if} is α -recoverable if and only if:

$$\mathcal{Z} \left(\overset{\text{texture}}{\kappa_I^2} \cdot \overset{\text{SNR}}{\frac{1}{\kappa_n^2}} \right) \cdot \overset{\text{patch size}}{\sqrt{\frac{T-1}{2}}} > \Phi^{-1}(1 - \alpha), \quad (3)$$

where: 1. $\mathcal{Z}(x) := x / \sqrt{1 + (x+1)^2}$ is a monotonically increasing function; and 2. $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal distribution.

Proposition 2 (proved in supplement) provides an interpretable view of the three key factors impacting recoverability: texture contrast, SNR, and patch size.

The green highlighted factor in Equation (3) suggests improving recoverability by increasing SNR. Doing so is possible by increasing exposure time t (Equation (2)), but only up to the saturation limit I_{max} . Further increasing SNR is possible by averaging captures from multiple sensor readouts. However, doing so increases acquisition time by more than the increase in total exposure time.

The brown highlighted factor in Equation (3) suggests improving recoverability by increasing the patch size T —representing the decreasing impact of noise in the sample contrast $\tilde{c}_{\text{if}}^2$ as samples increase. However, increasing T comes at the cost

of decreased lateral resolution and increased errors close to depth discontinuities (which are not accounted for in standard DFF).

The classical understanding of texture in computer vision suggests that we cannot improve the blue highlighted factor in Equation (3) while maintaining passive operation—*i.e.*, without active illumination to artificially texture surfaces. In what follows, we show this understanding to be incorrect: We present a *second-order* theory of texture (§ 4) based on wave-optical scattering of light, which predicts that it is indeed possible to amplify texture *by simply adding a narrow-band spectral filter in front of the camera*. We provide extensive experimental verification (§ 6) of this prediction, showing that the fundamental requirement for texture in passive depth recovery is not as severe as previously thought. Of course, adding a spectral filter also worsens the SNR factor in Equation (3), assuming fixed exposure time. We show (§ 5) that the improvement in texture contrast outweighs the worsening in SNR after only a sublinear increase in exposure time.

4 A second-order theory of texture

Further analysis of κ_I^2 in Equation (3) requires characterizing the pointwise distribution of pixel intensities $I(\vec{p})$. We assume for simplicity that the incident flux $\Phi_j(\vec{p})$ at a pixel is proportional to the irradiance $E(\vec{p})$ at its center, $\Phi_j(\vec{p}) = E_j(\vec{p})\Delta_p^2$. It follows that $\kappa_I^2 = \kappa_E^2 := \sigma_E^2/\mu_E^2$, thus it suffices to analyze κ_E^2 .

Our goal in this section is thus to characterize the irradiance $E(\vec{p})$. In § 4.1, we use wave-optical principles of light scattering to derive an expression of $E(\vec{p})$ for a camera imaging *optically rough* surfaces under ambient illumination. We then explain how this expression relates to appearance models in computer vision [64–66, 70], and why the appearance of real-world rough surfaces is *always* textured due to subjective speckle, even for surfaces that classical computer vision would consider textureless. In § 4.2 we quantify the strength of this textured appearance as a function of imaging and illumination conditions.

4.1 First and second-order texture

To simplify exposition, and without loss of generality, we assume a thin-lens camera that images a planar, frontoparallel, and optically rough surface under far-field illumination (Figure 3). Our problem setting and derivation closely adapt Goodman [46, §§5.8 & 6.3]. We use a 3D coordinate system where the z axis is parallel to the optical axis and the macroscopic surface normal $\hat{n}$. We use 3D unit-norm vectors $\hat{\omega}$ for directions, and $\hat{\omega}_{xy}$ for their projection onto the x - y plane. We use 2D vectors $\vec{p}$ on the x - y plane interchangeably for both sensor and surface points, and identify each sensor point with the surface point it maps to through its central ray $\hat{v}$ (*i.e.*, pinhole projection).

We model the camera lens using its focus-dependent *point spread function* (PSF) $P_c(\vec{p} - \vec{q})$, and use $w_c(k)$ for the spectral sensitivity function (SSF) of the sensor. We model the far-field illumination as a spectral radiance environment map that we assume separable in direction and wavenumber, $L_i(\hat{\omega}, k) := L_i(\hat{\omega})\ell_i(k)$, where $k = 2\pi/\lambda$ is the wavenumber and λ the wavelength, assuming mutually

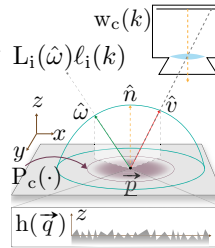


Figure 3: Problem setup and notation.

incoherent sources at different directions $\hat{\omega}$. Our goal is to compute the irradiance reaching the sensor due to the scattering of $L_i(\hat{\omega}, k)$ at the surface point $\vec{p}$.

If we zoomed in on the surface around $\vec{p}$ (Figure 4), we would observe microscopic roughness that we model in Monge form using a *random heightfield* h . This roughness has scale comparable to the visible wavelength (hence, *optically rough*) and is not resolvable by the camera, thus the surface appears macroscopically planar. Instead, the roughness is responsible for the surface’s wave-optical scattering behavior, determining its *appearance*. We model this appearance using the Fourier transform of a functional of the heightfield, as follows.

Definition 3: Sample BRDF

Given a local surface heightfield $h(\cdot)$, we define the *sample bidirectional reflectance distribution function* (sample BRDF) f_r^h as:

$$f_r^h(\vec{q}, \vec{u}, k) := \frac{1}{A} \int_{\mathbb{R}^2} S(\vec{q}, \vec{r}, k, u_z) e^{ik \vec{u}_{xy} \cdot 2\vec{r}} d\vec{r}, \quad (4)$$

where S is a functional of the heightfield h and reflection coefficient r ,

$$\begin{aligned} S(\vec{q}, \vec{r}, k, u_z) &:= s(\vec{q} + \vec{r}; k, u_z) s^*(\vec{q} - \vec{r}; k, u_z), \\ s(\vec{\rho}; k, u_z) &:= r(\vec{\rho}, k) e^{ik u_z h(\vec{\rho})}, \quad \vec{u} \in \mathbb{R}^3. \end{aligned} \quad (5)$$

We elaborate on roughness assumptions and required properties for h in the supplement. We can use f_r^h , which is random through h , to express the irradiance.²

Proposition 4: Appearance model

The irradiance received at sensor point $\vec{p}$ equals:

$$E(\vec{p}) \approx \int_{\mathbb{R}^+} \int_{\mathbb{R}^2} w_c(k) \ell_i(k) P_c(\vec{p} - \vec{q}) L_o(\vec{q}, \hat{v}, k) d\vec{q} dk, \quad \text{where the} \quad (6)$$

outgoing radiance L_o at surface point $\vec{q}$, direction $\hat{v}$, and wavenumber k is:

$$L_o(\vec{q}, \hat{v}, k) := \int_{\mathcal{H}^2(\hat{n})} f_r^h(\vec{q}, \hat{\omega} + \hat{v}, k) L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) d\sigma(\hat{\omega}). \quad (7)$$

Relationship to classical appearance model. Equation (6) is equivalent to the standard PSF-based image formation model for a thin lens. Equation (7) is *almost* the standard reflectance equation in computer vision, except it uses the sample BRDF f_r^h instead of the standard BRDF $\bar{f}_r$ for rough surfaces as defined in, *e.g.*, Stam [65]. This difference is the basis for our texture theory: f_r^h , and thus also E , are random variables due to their dependence on h . Classical appearance models eliminate this randomness by using *ensemble averages* $\mathbb{E}_h[f_r^h]$, $\mathbb{E}_h[E]$ over all microgeometry realizations. We make the link to classical models explicit.

² f_r^h is also a scaled Wigner transform of $s(\cdot; k, u_z)$.

Proposition 5: Ensemble averaging

Assuming ergodicity for h , $\mathbb{E}_h[f_r^h(\vec{q}, \hat{\omega} + \hat{v}, k)] = \bar{f}_r(\vec{q}, \hat{\omega} + \hat{v}, k)$.

From linearity of expectation, replacing f_r^h with $\bar{f}_r$ in Equations (6) and (7) produces the classical radiometric equations for the ensemble-averaged irradiance.

Subjective speckle as second-order texture. We use the notation:

$$\bar{E}(\vec{p}) := \mathbb{E}_h[E(\vec{p})], \quad E^s(\vec{p}) := E(\vec{p}) - \bar{E}(\vec{p}). \quad (8)$$

The difference E^s between irradiance and its ensemble-average is known in optics as *subjective speckle* [46, 62]. E^s arises because of wavelength-scale surface microgeometry that *cannot be resolved* by the camera at the working magnification. Resolvable changes in surface geometry or material, including visible surface roughness at high magnifications [9], are instead represented in the BRDF $\bar{f}_r$ (which includes foreshortening), and thus $\bar{E}$.

Returning to DFF, classical computer vision theory defines “texture” as spatial variations of the ensemble-averaged irradiance $\bar{E}$ (Figure 4). However, the presence of subjective speckle E^s suggests that even a surface that would classically be considered *textureless*—random microgeometry even at *i.e.*, one with spatially constant BRDF—can show irradiance variations due to non-resolvable microgeometry. We use the terms *first-order* and *second-order texture* to distinguish the two types of variations, as one considers only the first moment of E (ensemble average $\bar{E}$), whereas the other considers also second moments³ of E (variance of subjective speckle E^s).

When is ensemble averaging accurate? The ensemble averaging assumption of classical computer vision is accurate under geometric camera models (*e.g.*, perspective, orthographic) with an infinitesimal aperture and thus infinite PSF in Equation (6). Most previous derivations of appearance models for rough surfaces [64–66] assume an infinite PSF, making them inaccurate for lens-based cameras under large-aperture or close-focus conditions—and thus DFF. Our derivation of Proposition 4 assumes finite PSFs to study implications of deviations from ensemble averaging for DFF.

4.2 Characterizing second-order texture contrast

Our next goal is to derive an expression for the coefficient of variation κ_E^2 —and thus κ_I^2 —that accounts for both first-order and second-order texture. We can

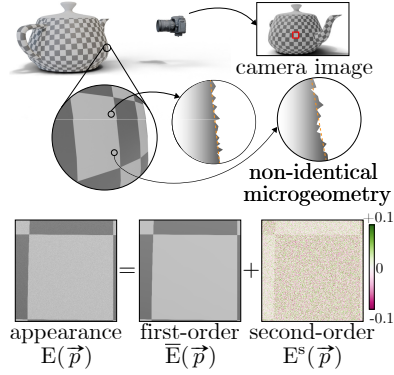


Figure 4: First-order texture (checkerboard pattern) is due to visible BRDF variations. Second-order texture (speckle) is due to invisible random microgeometry even at first-order textureless regions.

³ E^s captures all higher-order moments; we focus on second-order moments in the context of Proposition 2 for DFF.

then use this expression to assess the practicality of second-order texture for DFF, and characterize recoverability improvements. We define

$$\Lambda_i(2\vec{r}, k) := \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) e^{jk2\vec{r} \cdot \hat{\omega}_{xy}} d\sigma(\hat{\omega}) \quad (9)$$

as the *coherence function* (CF) of the far-field illumination L_i , *i.e.* its (projected) Fourier transform. The effective support size of Λ_i is the *coherence area* Δ_i^2 [66, 71], which increases as the illumination's directional footprint decreases. Combining Equations (4), (7) and (9) gives $L_o(\vec{q}, \hat{v}, k) = \mathcal{F}_{\vec{r}}\{S \cdot \Lambda_i\}(\hat{v})$, which we use with Proposition 4 to rewrite $E(\vec{p})$ as (denoting $f(k) = w_c(k)\ell_i(k)$):
 spectral integration over band Δ_k

$$\int_{\mathbb{R}^+} f(k) \int_{\mathbb{R}^2} P_c(\vec{p} - \vec{q}) \int_{\mathbb{R}^2} S(\vec{q}, \vec{r}, k) \Lambda_i(2\vec{r}, k) e^{jk2\vec{r} \cdot \hat{v}_{xy}} d\vec{r} d\vec{q} dk. \quad (10)$$

spatial integration over area Δ_c^2
 spatial integration over area Δ_i^2

We can interpret this expression as follows (Figure 5):

1. The innermost integral equals $L_o(\vec{q}, \hat{v}, k)$, and aggregates complex and random (due to the dependence of S on h) values over points $\vec{r}$ in an area of size Δ_i^2 controlled by the CF Λ_i . As the integrand is the product of conjugate-symmetric functions, the integral is real-valued⁴ for all $\vec{q}$.
2. The intermediate integral equals the spectral density of irradiance $E(\vec{p}, k)$, and aggregates real random $L_o(\vec{q}, \hat{v}, k)$ values over points $\vec{q}$ in an area of focus-dependent size Δ_c^2 controlled by the PSF P_c .
3. The outermost integral aggregates real random $E(\vec{p}, k)$ values over wavenumbers k in a spectral range Δ_k controlled by the illuminant ℓ_i and SSF w_c .

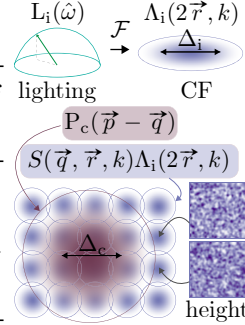


Figure 5: Coherent summation within coherence areas results in speckle (random radiance). Incoherent summation of such areas within the PSF reduces speckle contrast.

Using optics terminology, the innermost integral performs *coherent summation* within each coherence area—resulting in random $L_o(\vec{q}, \hat{v}, k)$ values manifesting as speckle—whereas the other two perform *incoherent summation* within the PSF area and spectral bandwidth. Assuming h has sufficiently fast variation within each coherence area, the random $L_o(\vec{q}, \hat{v}, k)$ values are approximately independent for points $\vec{q}$ separated by more than Δ_i (*i.e.*, non-overlapping Δ_i^2 -sized integration areas), or for sufficiently separated wavenumbers k —manifesting as uncorrelated speckle. Then, we can understand $E(\vec{p})$ as averaging independent real-valued random radiances whose number scales proportionally to $\Delta_k \Delta_c^2 / \Delta_i^2$. Consequently, the variance of $E(\vec{p})$ is inversely proportional to this number, tending to 0 as

⁴ L_o can take negative values, similar to forms of a generalized radiance [72, 73].

defocus (and thus Δ_c^2) or bandwidth (and thus Δ_k) increase—reducing speckle contrast and converging to ensemble averaging (Figure 6)—while attaining its maximal value when in focus. We formalize this discussion next.

Proposition 6: Pointwise coefficient of variation of irradiance

In focus, the coefficient of variation of intensity and irradiance equal

$$\kappa_I^2 = \kappa_E^2 \approx \frac{1}{MQ} (1 + Q\kappa_E^2) + \kappa_E^2, \quad (11)$$

where: 1. κ_E^2 is the in-focus coefficient of variation for first-order texture $\bar{E}$; 2. $Q := \pi\Delta_{c,if}^2/\Delta_I^2$ is the relative size of the in-focus PSF and coherence area; 3. $M := G_1\Delta_k$ is proportional to the spectral bandwidth, through a constant G_1 that depends on statistical properties of h and the mean wavevector.

Notably, we consider a surface without first-order texture, meaning constant ensemble-averaged irradiance $\bar{E}$, and thus $\kappa_E^2 = 0$. Such a surface would be classically described as “textureless.” Yet, Equation (11) predicts non-zero in-focus texture contrast $\kappa_I^2 = 1/MQ$, due to second-order texture. Therefore, from Proposition 2, even this *first-order textureless* surface is potentially recoverable, so long as the contrast is stronger than noise. Equation (11) further predicts that we can increase this contrast by reducing the bandwidth Δ_k and thus M —*e.g.*, by using a spectral filter.⁵ Assessing recoverability requires considering the impact on SNR, as we do in § 5.

Practicality for computer vision. The factor Q in Equation (11) helps understand under what conditions second-order texture is practical for DFF.

Its in-focus contrast improves with larger coherence areas Δ_I^2 due to illumination dominated by concentrated far-field sources (*e.g.*, outdoor sunlight, indoor small ceiling lights), and worsens with more directionally uniform illumination (*e.g.*, outdoor overcast conditions, indoor large ceiling panels). It also improves with smaller in-focus PSF sizes $\Delta_{c,if}^2$, *i.e.*, smaller f-numbers or reproduction ratios [74]. In § 6 we show that, using bandwidths $\Delta_\lambda = 10$ –100 nm, second-order texture improves DFF in settings typical for passive computer vision: 1. both outdoor and indoor ambient lighting ($\Delta_I \approx 50$ –100 μm [45]); 2. reproduction ratios $1/50$ – $1/10$ at f-numbers up to 6 ($\Delta_{c,if} \approx 30$ –150 μm). Second-order contrast reduces also with

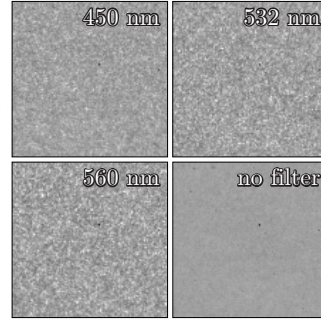


Figure 6: In-focus images of a white target using different filters (center λ) show uncorrelated speckle patterns. Their incoherent summation without a filter eliminates speckle.

⁵ The same principle applies to using a *polarizer* in front of the camera to improve texture contrast, as the two orthogonal (incoherent) polarization states from unpolarized light generally produce uncorrelated speckle [46, §4.5].

larger pixel sizes [44, 75], favoring smaller pixel pitches (3 μm in our experiments). We elaborate in the supplement.

5 Recoverability with second-order texture

We focus on recoverability for a first-order textureless surface, *i.e.*, $\kappa_E^2 = 0$. From Proposition 2, it suffices to express the *texture contrast-SNR product* Θ^2 as a function of bandwidth Δ_k and exposure time t . Equation (11) already provides an expression for κ_I^2 ; whereas from Equation (2), getting an expression for κ_n^2 requires relating the mean flux μ_Φ to Δ_k . We prove the following.

Proposition 7: Texture contrast-SNR product

In focus, the texture contrast-SNR product $\Theta^2 := \kappa_I^2/\kappa_n^2$ equals:

$$\Theta^2(t, \Delta_k) \approx \frac{1}{Q} \frac{G_2}{G_1} \frac{t}{\left(1 + \frac{\sigma_{\text{read}}^2}{G_2} \frac{1}{\Delta_k t}\right)}, \quad (12)$$

where the exposure time $t \in [0, g_{\text{I max}}/G_2 \Delta_k)$ has a saturation limit dependent on spectral bandwidth Δ_k , and the constant G_2 depends on spectral reflectance.

Analysis and simulation. In Figure 7, we visualize the probability of error $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ predicted from Equations (3) and (12) for different bandwidths Δ_k and exposure times t . We also use Monte Carlo simulation to numerically estimate $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ and assess the accuracy of our theory. Assuming a first-order textureless surface, we sampled focal stacks of noisy speckle patterns, then computed sample contrast $\tilde{c}^2$. We sampled speckle using a statistical model detailed in the supplement, and sensor noise using the model of § 3. The supplement lists all parameters (illumination, surface, sensor, lens) and other details of the visualization and simulation. The results show closely matching theory and simulation predictions for $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ (lower is better). We also observe:

1. For constant bandwidth (a vertical line), the optimal t is the saturation limit. As texture contrast κ_I^2 is fixed, maximizing t maximizes SNR $1/\kappa_n^2$ and Θ^2 .
2. For constant exposure time (a horizontal line), the optimal Δ_k is again determined by the saturation limit, *i.e.*, maximizing Δ_k till intensities saturate. The decrease in κ_I^2 is outweighed by the increase in $1/\kappa_n^2$, improving Θ^2 .
3. For constant bandwidth-exposure time product (a diagonal line, corresponding to constant SNR), optimizing Θ^2 requires minimizing Δ_k while proportionally increasing t , to stay close to the saturation limit.

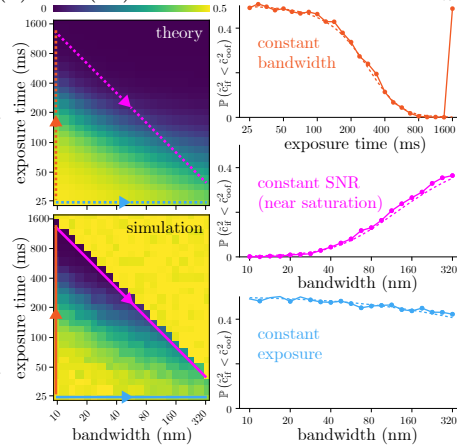


Figure 7: Error probability $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ estimation from theory and simulation.

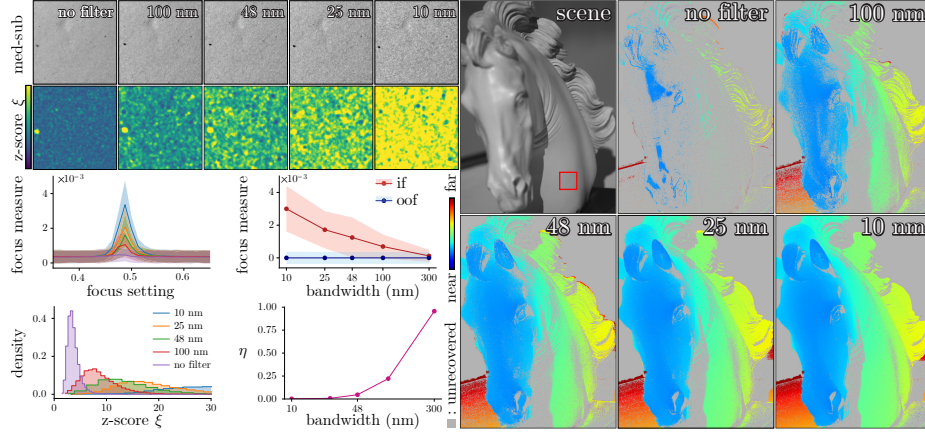


Figure 8: DFF improves with filters of decreasing bandwidth thanks to enhanced second-order texture, visible in median-subtracted crops (‘med-sub’). This enhancement also results in focus measure peaks more distinguishable from noise (top two plots), decreasing the fraction of pixels under a z-score threshold (bottom two plots, gray pixels in depth maps). Example captured outdoors under sunlight.

4. When decreasing Δ_k , it is also possible to improve Θ^2 by increasing t *sublinearly*, *i.e.*, by less than the amount required to maintain constant SNR.
5. Due to the saturation limit for a single capture (diagonal discontinuity in simulated results), lowering Δ_k with a filter enables much reduced probability of error (top-left corner) than without a filter (bottom-right corner).

In the next section, we validate these observations through extensive experiments.

6 Experiments & limitations

We perform DFF experiments across multiple scenes using *only pre-existing ambient light*—either sunlight, or standard ceiling lights. The supplemental PDF and project website include more experiments, visualizations, code, and data.

Experimental setup. Our setup uses a machine vision camera (FLIR BFS-U3-122S6M-C), a motorized lens mount for focus control, a photographic lens (Canon EF-S 60mm f/2.8 Macro USM), and narrow-band spectral filters (Edmund Optics, different FWHM bandwidths, center wavelength ≈ 530 nm).

Implementation details. Given a focal stack $\{I_j(\vec{p})\}_{j=1}^J$, we use a focus measure that aggregates 3×3 Laplacian values over a Gaussian kernel of width $\sigma_{\text{agg}} = 5/2$ [68]. Empirically, this measure performed better than the sample contrast. For simplicity, we continue to use $\tilde{c}^2$ for this measure.

As we do not have access to ground-truth depth in general, we quantify DFF performance using the per-pixel *robust z-score* $\xi := |\tilde{c}_*^2 - \text{median}(\tilde{c}_j^2)| / \text{MAD}(\tilde{c}_j^2)$, where: 1. $\tilde{c}_*^2$ is the max focus measure at j^* ; 2. $\text{median}(\tilde{c}_j^2)$ and

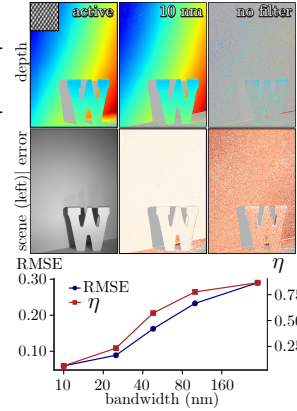


Figure 9: Validation of the η metric with active DFF.

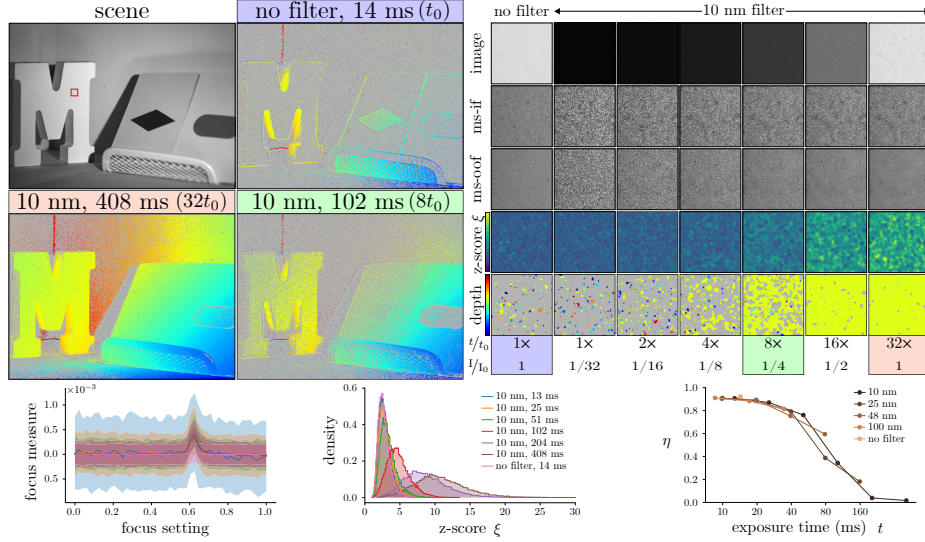


Figure 10: Using a spectral filter improves DFF even at suboptimal exposures. At $1/4$ the optimal exposure time, second-order texture is already discernible from noise at median-subtracted crops (‘ms-if’ vs ‘ms-oof’), improving focus measures and z-score statistics (bottom plots). Example captured indoors under ceiling lights.

$\text{MAD}_j(\tilde{c}_j^2) := \text{median}_j|\tilde{c}_j^2 - \text{median}(\tilde{c}_j^2)|$ are the median and median absolute deviation (MAD) of the focus measure curve. Thus, ξ quantifies confidence that the focus measure peak corresponds to the true depth.

We use ξ to filter out noisy depth estimates by thresholding at some level ξ_{th} (shown as gray pixels in depth maps), and to compute an error metric $\eta \in [0, 1]$ equal to the *fraction of unrecovered pixels* (pixels with $\xi(\vec{p}) < \xi_{\text{th}}$). We use η as a metric to compare DFF performance, as a proxy of $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{oof}}^2)$. In Figure 9, we validate this metric by comparing it to the root-mean-square error (RMSE) with respect to reference depth obtained with *active* DFF (*i.e.*, using a projector to create artificial texture). The figure shows that RMSE and η are strongly correlated as expected—wrong depth estimates arise from flat focus curves (low texture or high noise), hence have low ξ values. We elaborate in the supplement.

Contrast enhancement. Figure 8 shows enhanced in-focus texture and drastically improved depth using filters of decreasing bandwidth, while adjusting exposure time to maintain exposure. Median-subtracted images of a weakly textured patch help visualize the contrast enhancement. The plots show focus curves and z-score statistics. Figures 1, 2 and 13 show more scenes.

Exposure impact. Figures 10 and 11 assess the tradeoff between exposure time and spectral bandwidth. In Figure 10, adding a filter of bandwidth

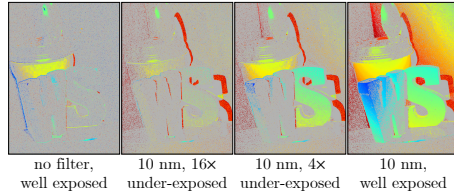


Figure 11: DFF improves with a spectral filter even at suboptimal exposures. Example captured outdoors under sunlight.

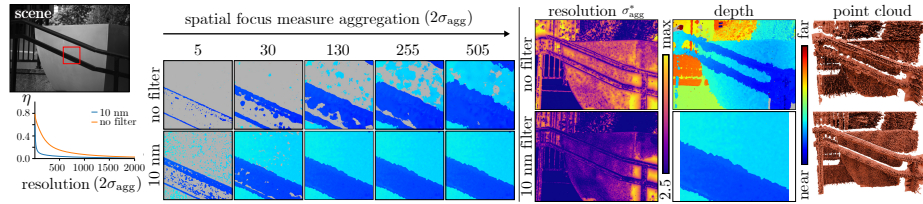


Figure 12: Aggregating focus measures over progressively larger patch sizes allows depth recovery at adaptive scales. With the filter, much finer details are recovered at smaller patch sizes (see resolution map). Example captured outdoors under sunlight.

sure time by $32\times$ to maintain proper exposure. Yet DFF performance already improves when we increase exposure time by just $4\times$ (*i.e.*, underexposure by 3 stops), with further improvements closer to proper exposure. The crops and plots show the increase in contrast and z-scores, at a weakly textured patch—where second-order texture gradually overcomes noise—and in aggregate. The same behavior holds for other bandwidths (bottom right) and outdoors (Figure 11).

Pixel-adaptive aggregation. So far we considered DFF with a fixed focus measure aggregation width σ_{agg} (or effectively, fixed patch size T) at all pixels. However, the optimal aggregation width can vary per pixel, depending on local texture contrast and noise level. We use a simple adaptive method that, at each pixel, selects the *minimum* aggregation width σ_{agg}^* that achieves z-score above a threshold ξ_{th} . We can then use σ_{agg}^* as a metric to assess improvements in lateral resolution due to second-order texture. Figure 12 shows that using a spectral filter indeed results in depth maps of significantly higher lateral resolution.

Limitations. The supplement shows experiments representative of the two main cases where using a spectral filter may not improve DFF performance: The first is translucent surfaces, as subsurface scattering greatly diminishes speckle contrast [46, §6.4.3]. The second is illumination with large directional bandwidth (*e.g.*, indoor lights that are very large or very close to the scene, overcast outdoor conditions, or strong indirect light), making Δ_i much smaller than $\Delta_{\text{c,if}}$. We show that the deterioration of contrast, and thus DFF performance, with increasing directional bandwidth is consistent with our theory (§ 4.2).

Moreover, improving DFF performance with a spectral filter requires increasing exposure time—even if sublinearly. Figures 7 and 10 show that a filter bandwidth $\Delta_\lambda = 10\text{ nm}$ requires a $4\times$ exposure time increase before performance starts to improve. Our results should help assess whether different application settings justify this tradeoff between depth accuracy and acquisition time.

7 Conclusion

We developed a second-order theory of texture, with significant implications for the definition of texture and its role in passive depth from focus. Our theory shows that: 1. Even surfaces that would traditionally be considered textureless can produce textured appearance, thanks to subjective speckle due to surface microgeometry. 2. Using a narrowband spectral filter enhances the in-focus contrast of this texture. 3. This enhancement can bring about dramatic improvements in

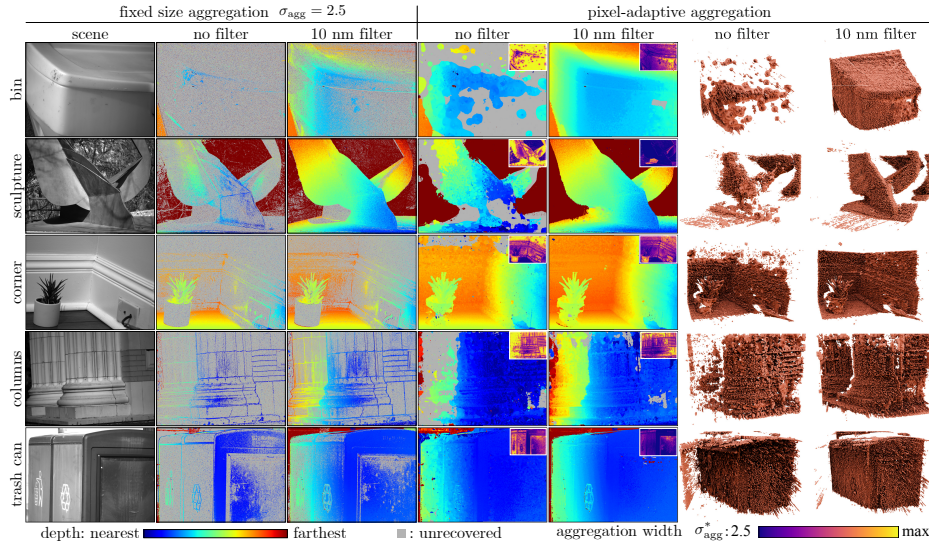


Figure 13: Examples of DFF improvement in indoor and outdoor real-world scenes. Using a spectral filter improves both fixed-size and pixel-adaptive reconstructions. The project website provides interactive visualizations of these and other scenes.

depth accuracy under conditions typical for passive computer vision. We validated this theory through extensive simulations and experiments.

Our findings invite a broader reexamination of texture and passive depth sensing in computer vision: Though we developed our theory in the context of depth from focus, second-order texture can likely benefit other passive methods using focus or parallax. Further development of algorithms (confocal constancy [17] and focal flow [18] of subjective speckle) and theory (multi-view correlations of subjective speckle [76] and memory effect [57]) can realize these benefits. Additionally, the different statistics of speckle and sensor noise hint at the possibility of using learning-based methods to detect speckle even under very low SNR. Future research can adapt related successful approaches in other speckle imaging applications [55, 56]. Lastly, both our work and recent work on passive interferometric depth sensing [43–45] rely on the weak coherence of ambient light to relax texture requirements. Further research should help understand the relative merits of interferometric and non-interferometric methods.

Acknowledgements

We thank Dorian Chan, Aswin Sankaranarayanan, and Matthew O’Toole for helpful discussions about speckle, and Mian Wei for feedback on writing. This work was supported by NSF award 2047341, and a Sloan Research Fellowship.

References

1. Sundaram, H., Nayar, S.: Are textureless scenes recoverable? In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 814–820, IEEE (1997)

2. Barnard, S.T., Thompson, W.B.: Disparity analysis of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**(4), 333–340 (1980)
3. Nalpantidis, L., Sirakoulis, G., Gasteratos, A.: Rev. of stereo vision algorithms: Software to hardware. *Int. J. Optomechatron.* **2**(4), 435–462 (2008)
4. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2004)
5. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**(1), 7–42 (2002)
6. Suwajanakorn, S., Hernandez, C., Seitz, S.M.: Depth from focus with your mobile phone. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3497–3506 (2015)
7. Hazirbaş, C., Soyer, S.G., Staab, M.C., Leal-Taixé, L., Cremers, D.: Deep depth from focus. In: *Asian Conf. Comput. Vis.*, pp. 525–541 (2018)
8. Grossmann, P.: Depth from focus. *Pattern Recognit. Lett.* **5**(1), 63–69 (1987)
9. Nayar, S.K., Nakagawa, Y.: Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(8), 824–831 (1994)
10. Nayar, S.K., Watanabe, M., Noguchi, M.: Real-time focus range sensor. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(12), 1186–1198 (1996)
11. Subbarao, M., Choi, T.: Accurate recovery of three-dimensional shape from image focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(3), 266–274 (1995)
12. Subbarao, M., Surya, G.: Depth from defocus: A spatial domain approach. *Int. J. Comput. Vis.* **13**(3), 271–294 (1994)
13. Favaro, P.: Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1133–1140 (2010)
14. Pentland, A.P.: A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.* **9**(4), 523–531 (1987)
15. Tang, H., Cohen, S., Price, B., Schiller, S., Kutulakos, K.N.: Depth from defocus in the wild. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2740–2748 (2017)
16. Watanabe, M., Nayar, S.K.: Rational filters for passive depth from defocus. *Int. J. Comput. Vis.* **27**(3), 203–225 (1998)
17. Hasinoff, S.W., Kutulakos, K.N.: Confocal stereo. *Int. J. Comput. Vis.* **81**, 82–104 (2009)
18. Alexander, E., Guo, Q., Koppal, S., Gortler, S., Zickler, T.: Focal flow: Measuring distance and velocity with defocus and differential motion. In: *Eur. Conf. Comput. Vis.*, pp. 667–682, Springer (2016)
19. Guo, Q., Alexander, E., Zickler, T.: Focal track: Depth and accommodation with oscillating lens deformation. In: *Int. Conf. Comput. Vis.*, pp. 966–974 (2017)
20. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* **26**(3), 70 (2007)
21. Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.* **26**(3), 69 (2007)

22. Zhou, C., Nayar, S.K.: What are good apertures for defocus deblurring? In: IEEE Int. Conf. Comput. Photography, pp. 1–8 (2009)
23. Zhou, C., Lin, S., Nayar, S.K.: Coded aperture pairs for depth from defocus. In: Int. Conf. Comput. Vis., pp. 325–332 (2009)
24. Chakrabarti, A., Zickler, T.: Depth and deblurring from a spectrally-varying depth-of-field. In: Eur. Conf. Comput. Vis., pp. 648–661, Springer (2012)
25. Schechner, Y.Y., Kiryati, N.: Depth from defocus vs. stereo: How different really are they? Int. J. Comput. Vis. **39**, 141–162 (2000)
26. Bolles, R.C., Baker, H.H., Marimont, D.H.: Epipolar-plane image analysis: An approach to determining structure from motion. Int. J. Comput. Vis. **1**(1), 7–55 (1987)
27. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.H.: Scene reconstruction from high spatio-angular resolution light fields. ACM Trans. Graph. **32**(4), 73:1–73:12 (2013)
28. Punnapurath, A., Abuolaim, A., Afifi, M., Brown, M.S.: Modeling defocus-disparity in dual-pixel sensors. In: IEEE Int. Conf. Comput. Photography, pp. 1–12 (2020)
29. Punnapurath, A., Brown, M.S.: Reflection removal using a dual-pixel sensor. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 1556–1565 (2019)
30. Garg, R., Wadhwa, N., Ansari, S., Barron, J.T.: Learning single camera depth estimation using dual-pixels. In: Int. Conf. Comput. Vis., pp. 7628–7637 (2019)
31. Xin, S., Wadhwa, N., Xue, T., Barron, J.T., Srinivasan, P.P., Chen, J., Gkioulekas, I., Garg, R.: Defocus map estimation and deblurring from a single dual-pixel image. In: Int. Conf. Comput. Vis., pp. 2228–2238 (2021)
32. Rosen, J., Vijayakumar, A., Kumar, M., Rai, M.R., Kelner, R., Kashter, Y., Bulbul, A., Mukherjee, S.: Recent advances in self-interference incoherent digital holography. Adv. Opt. Photonics **11**(1), 1–66 (2019)
33. Liu, J.P., Tahara, T., Hayasaki, Y., Poon, T.C.: Incoherent digital holography: a review. Appl. Sci. **8**(1), 143 (2018)
34. Tahara, T., Zhang, Y., Rosen, J., Anand, V., Cao, L., Wu, J., Koujin, T., Matsuda, A., Ishii, A., Kozawa, Y., Okamoto, R., Oi, R., Nobukawa, T., Choi, K., Imbe, M., Poon, T.C.: Roadmap of incoherent digital holography. Appl. Phys. B **128**(11), 193 (2022)
35. Tahara, T., Kanno, T., Arai, Y., Ozawa, T.: Single-shot phase-shifting incoherent digital holography. J. Opt. **19**(6), 065705 (2017)
36. Tahara, T., Kozawa, Y., Ishii, A., Okamoto, R.: Palm-sized single-shot phase-shifting incoherent digital holography system with birefringent materials. In: Laser Science, pp. JT5B–52, Optica Publishing Group (2022)
37. Muroi, T., Nobukawa, T., Katano, Y., Hagiwara, K.: Capturing videos at 60 frames per second using incoherent digital holography. Opt. Contin. **2**(11), 2409–2420 (2023)
38. Davy, M., Fink, M., de Rosny, J.: Green’s function retrieval and passive imaging from correlations of wideband thermal radiations. Phys. Rev. Lett. **110**, 203901 (May 2013)

39. Badon, A., Lerosey, G., Boccara, A.C., Fink, M., Aubry, A.: Retrieving time-dependent Green’s functions in optics with low-coherence interferometry. *Phys. Rev. Lett.* **114**, 023901 (Jan 2015)
40. Badon, A., Li, D., Lerosey, G., Boccara, A.C., Fink, M., Aubry, A.: Spatio-temporal imaging of light transport in highly scattering media under white light illumination. *Optica* **3**(11), 1160–1166 (Nov 2016)
41. Boger-Lombard, J., Katz, O.: Passive optical time-of-flight for non line-of-sight localization. *Nat. Commun.* **10**(1), 3343 (Jul 2019)
42. Batarseh, M., Sukhov, S., Shen, Z., Gemar, H., Rezvani, R., Dogariu, A.: Passive sensing around the corner using spatial coherence. *Nat. Commun.* **9**(1), 3629 (Sep 2018)
43. Cossairt, O., Matsuda, N., Gupta, M.: Digital refocusing with incoherent holography. In: *IEEE Int. Conf. Comput. Photography*, pp. 1–9, IEEE (2014)
44. Chen, W.Y., Sankaranarayanan, A.C., Levin, A., O’Toole, M.: Coherence as texture—passive textureless 3D reconstruction by self-interference. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 25058–25066 (2024)
45. Kotwal, A., Levin, A., Gkioulekas, I.: Passive micron-scale time-of-flight with sunlight interferometry. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 4139–4149 (2023)
46. Goodman, J.W.: *Speckle phenomena in optics: theory and applications*. SPIE (2020)
47. Mertz, J.: *Introduction to optical microscopy*. Cambridge University Press (2019)
48. Akkermans, E., Montambaux, G.: *Mesoscopic physics of electrons and photons*. Cambridge university press (2007)
49. Shih, Y.C., Davis, A., Hasinoff, S.W., Durand, F., Freeman, W.T.: Laser speckle photography for surface tampering detection. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 33–40, IEEE (2012)
50. Jo, K., Gupta, M., Nayar, S.K.: Spedo: 6 dof ego-motion sensor using speckle defocus imaging. In: *Int. Conf. Comput. Vis.*, pp. 4319–4327 (2015)
51. Smith, B.M., Desai, P., Agarwal, V., Gupta, M.: CoLux: Multi-object 3D micro-motion analysis using speckle imaging. *ACM Trans. Graph.* **36**(4), 1–12 (2017)
52. Smith, B.M., O’Toole, M., Gupta, M.: Tracking multiple objects outside the line of sight using speckle imaging. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 6258–6266 (2018)
53. Kichler, M., Bagon, S., Sheinin, M.: Learning to see inside opaque liquid containers using speckle vibrometry. In: *Int. Conf. Comput. Vis.*, pp. 9466–9476 (2025)
54. Sheinin, M., Chan, D., O’Toole, M., Narasimhan, S.G.: Dual-shutter optical vibration sensing. In: *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 16324–16333 (2022)
55. Metzler, C.A., Heide, F., Rangarajan, P., Balaji, M.M., Viswanath, A., Veeraraghavan, A., Baraniuk, R.G.: Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging. *Optica* **7**(1), 63–71 (2020)

56. Xie, M., Guo, H., Feng, B.Y., Jin, L., Veeraraghavan, A., Metzler, C.A.: WaveMo: learning wavefront modulations to see through scattering. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 25276–25285 (2024)
57. Alterman, M., Bar, C., Gkioulekas, I., Levin, A.: Imaging with local speckle intensity correlations: theory and practice. *ACM Trans. Graph.* **40**(3), 1–22 (2021)
58. Boniface, A., Blochet, B., Dong, J., Gigan, S.: Noninvasive light focusing in scattering media using speckle variance optimization. *Optica* **6**(11), 1381–1385 (2019)
59. Kim, J., Benko, C., Wrenninge, M., Villemin, R., Barber, Z., Jarosz, W., Pediredla, A.: A monte carlo rendering framework for simulating optical heterodyne detection. *ACM Trans. Graph.* **44**(4), 1–19 (2025)
60. Bar, C., Gkioulekas, I., Levin, A.: Rendering near-field speckle statistics in scattering media. *ACM Trans. Graph.* **39**(6), 1–18 (2020)
61. Bar, C., Alterman, M., Gkioulekas, I., Levin, A.: A monte carlo framework for rendering speckle statistics in scattering media. *ACM Trans. Graph.* **38**(4), 1–22 (2019)
62. Steinberg, S., Yan, L.Q.: Rendering of subjective speckle formed by rough statistical surfaces. *ACM Trans. Graph.* **41**(1), 2:1–2:23 (Feb 2022)
63. Liu, Z., Huo, Y., Peng, Y., Wang, R.: A fully-statistical wave scattering model for heterogeneous surfaces. *ACM Trans. Graph.* **44**(4), 1–17 (2025)
64. Beckmann, P., Spizzichino, A.: The scattering of electromagnetic waves from rough surfaces. Artech House (1987)
65. Stam, J.: Diffraction shaders. In: ACM SIGGRAPH, pp. 101–110 (1999)
66. Levin, A., Glasner, D., Xiong, Y., Durand, F., Freeman, W., Matusik, W., Zickler, T.: Fabricating BRDFs at high spatial resolution using wave optics. *ACM Trans. Graph.* **32**(4), 1–14 (2013)
67. Dong, Z., Walter, B., Marschner, S., Greenberg, D.P.: Predicting appearance from measured microgeometry of metal surfaces. *ACM Trans. Graph.* **35**(1), 1–13 (2015)
68. Subbarao, M., Tyan, J.K.: Selecting the optimal focus measure for autofocusing and depth-from-focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 864–870 (1998)
69. Hasinoff, S.W., Durand, F., Freeman, W.T.: Noise-optimal capture for high dynamic range photography. In: IEEE Conf. Comput. Vis. Pattern Recog., pp. 553–560, IEEE (2010)
70. Cuypers, T., Haber, T., Bekaert, P., Oh, S.B., Raskar, R.: Reflectance model for diffraction. *ACM Trans. Graph.* **31**(5), 1–11 (2012)
71. Wolf, E.: Introduction to the Theory of Coherence and Polarization of Light. Cambridge university press (2007)
72. Testorf, M.E., Hennelly, B.M., Ojeda-Castañeda, J.: Phase-space optics: fundamentals and applications. The McGraw-Hill Companies, Inc. (2010)
73. Steinberg, S., Pharr, M.: Wave tracing: Generalizing the path integral to wave optics. *Comput. Graph. Forum* p. e70322 (2026)
74. Kingslake, R.: The effective aperture of a photographic objective. *J. Opt. Soc. Am.* **35**(8), 518–520 (1945)

- 75. Gkioulekas, I., Levin, A., Durand, F., Zickler, T.: Micron-scale light transport decomposition using interferometry. *ACM Trans. Graph.* **34**(4), 37:1–37:14 (Jul 2015)
- 76. Fienup, J.R., Idell, P.S.: Imaging correlography with sparse arrays of detectors. *Opt. Eng.* **27**(9), 778–784 (1988)