

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

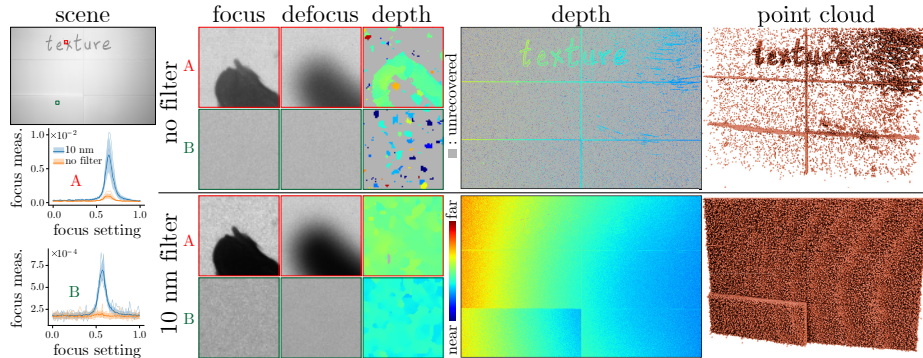


Figure 1: Computer vision traditionally considers depth recovery of scenes with sparse texture features impossible with passive methods such as depth from focus (top). We show that *simply adding a narrowband spectral filter to the camera* makes passive depth recovery of such scenes possible (bottom). 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>

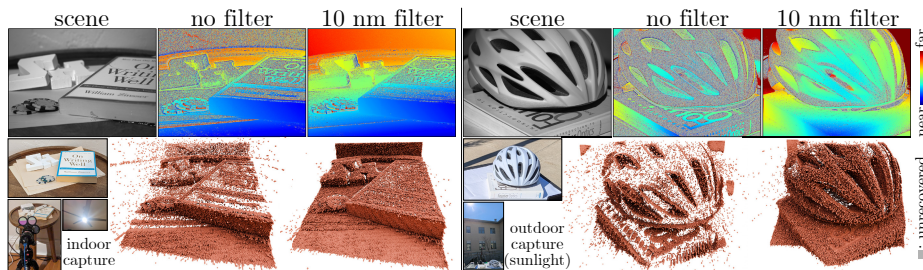


Figure 2: We demonstrate fine-scale depth recovery of objects with sparse texture features under ambient lighting indoors (ceiling lights, left) and outdoors (sunlight, right). In both cases, adding a spectral filter dramatically improves DFF performance.

[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 $j^*(\vec{p}) := \arg\max_j \tilde{c}_j^2(\vec{p})$. 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 on 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

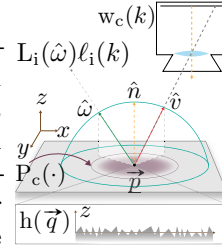


Figure 3: Problem setup and notation.

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

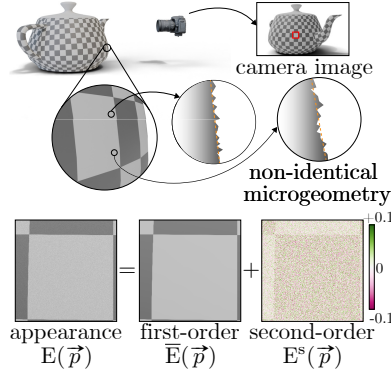


Figure 4: First-order texture (checkerboard pattern) is due to visible BRDF variations. Second-order texture (speckle) is due to invisible random microgeometry even at first-order textureless regions.

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 $f(k) = w_c(k)\ell_i(k)$):

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

spectral integration over band Δ_k

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 .

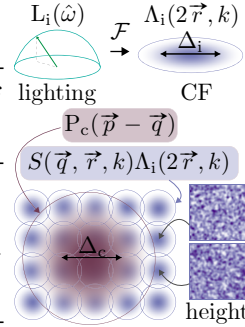


Figure 5: Coherent summation within coherence areas results in speckle (random radiance). Incoherent summation of such areas within the PSF reduces speckle contrast.

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

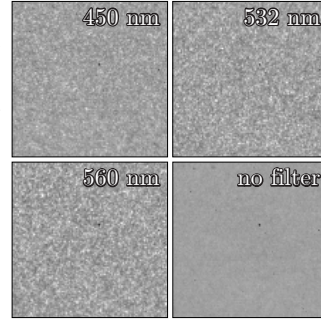


Figure 6: In-focus images of a white target using different filters (center λ) show uncorrelated speckle patterns. Their incoherent summation without a filter eliminates speckle.

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.

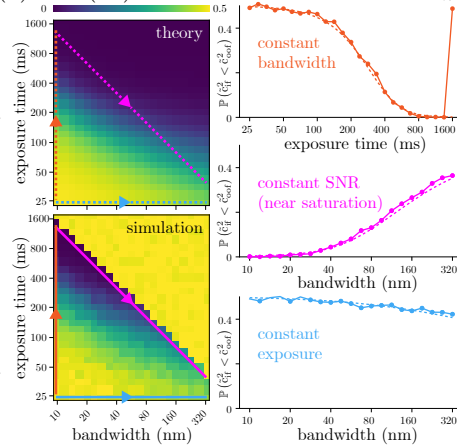


Figure 7: Error probability $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ estimation from theory and simulation.

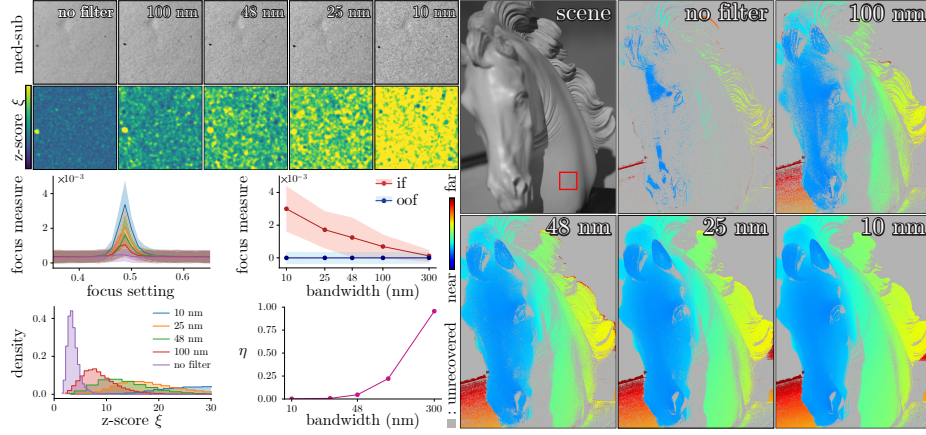


Figure 8: DFF improves with filters of decreasing bandwidth thanks to enhanced second-order texture, visible in median-subtracted crops (‘med-sub’). This enhancement also results in focus measure peaks more distinguishable from noise (top two plots), decreasing the fraction of pixels under a z-score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

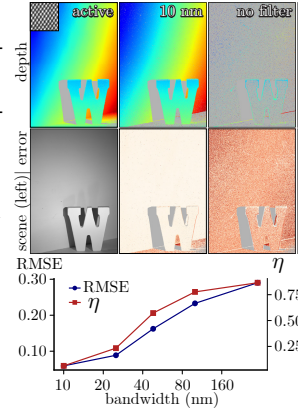


Figure 9: Validation of the η metric with active DFF.

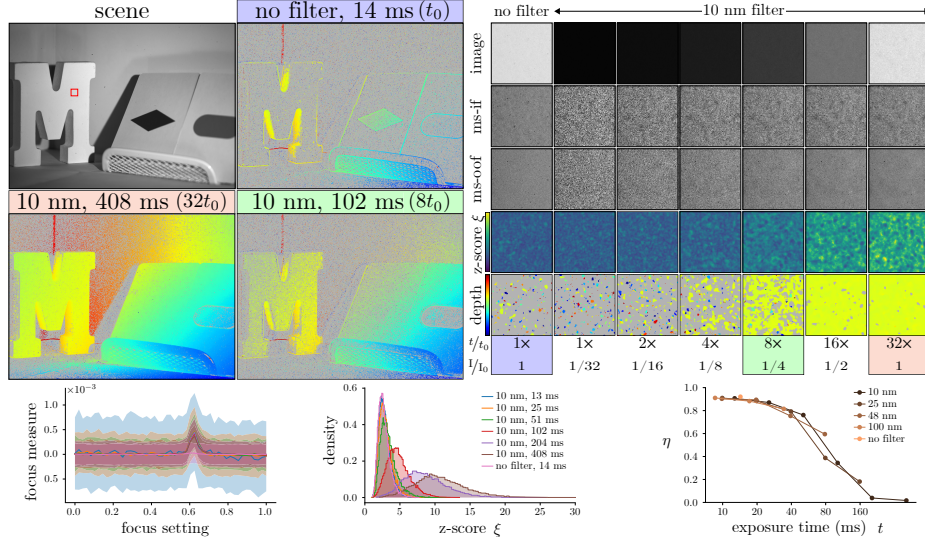


Figure 10: Using a spectral filter improves DFF even at suboptimal exposures. At $1/4$ the optimal exposure time, second-order texture is already discernible from noise at median-subtracted crops (‘ms-if’ vs ‘ms-oof’), improving focus measures and z-score statistics (bottom plots). Example captured indoors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

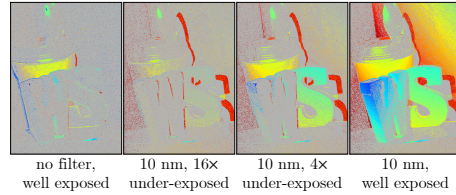


Figure 11: DFF improves with a spectral filter even at suboptimal exposures. Example captured outdoors under sunlight.

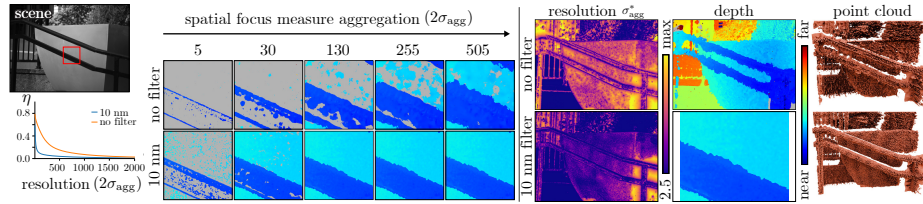


Figure 12: Aggregating focus measures over progressively larger patch sizes allows depth recovery at adaptive scales. With the filter, much finer details are recovered at smaller patch sizes (see resolution map). Example captured outdoors under sunlight.

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 footprint (*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

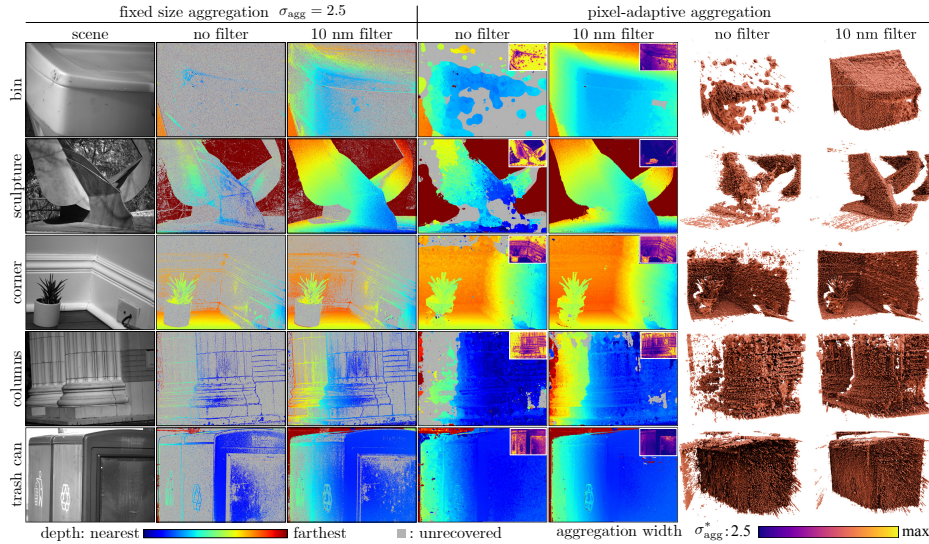


Figure 13: Examples of DFF improvement in indoor and outdoor real-world scenes. Using a spectral filter improves both fixed-size and pixel-adaptive reconstructions. The project website provides interactive visualizations of these and other scenes.

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. Punnappurath, 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. Punnappurath, 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., Boccaro, 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., Boccaro, 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)

A second-order theory of texture for depth from focus

— Supplemental Document —

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

This document is broadly organized into two parts. Appendices A–D are focused on additional results, analyses and implementation details, while Appendices E–I are dedicated to theoretical derivations including proofs of all of the main results.

Table of Contents

A	Practical conditions for second-order texture	2
A.1	Imaging: lens settings and pixel pitch	2
A.2	Illumination conditions	4
A.3	Material constraints	5
B	Interplay between texture contrast and noise	6
C	Depth recovery performance evaluation	9
C.1	Quantitative comparison	9
C.2	Validation of z-score-based metric with active DFF reference	11
D	Implementation details	11
D.1	Capture setup	11
D.2	Depth recovery algorithms	12
D.3	Monte Carlo simulation	12
E	Appearance model for second-order texture	16
E.1	Incident illumination model	16
E.2	Coherence function of the incident illumination	18
E.3	Scattering and Measurement	19
E.4	Approximating measured irradiance	20
E.5	Ensemble averaging	22
F	Characterization of second-order texture	24
F.1	Proof of exponential distribution	27
F.2	Spatial integration	28
G	Noise model	30
H	Depth from focus analysis	31
H.1	Statistics of the focus measure	32
H.2	Proof of Proposition 2	34
I	Recoverability with second-order texture	36
I.1	Attenuation due to spectral filtering	37
I.2	Gaussian spectral profile	37
I.3	Wavenumber and wavelength bandwidths	38

A Practical conditions for second-order texture

The in-focus contrast of second-order texture, and hence DFF performance, depends on practical factors spanning the imaging system, the illumination, and the surface material. The analysis and experiments presented in this section serve two purposes: 1. they validate our theoretical predictions on the nature of second-order texture in § 4.2; and 2. they demonstrate limitations of our technique, *i.e.* regimes where using a narrowband filter offers little to no benefit.

A.1 Imaging: lens settings and pixel pitch

In § 4.2, we showed that the in-focus contrast of second-order texture κ_I^2 is dependent on the in-focus PSF width $\Delta_{c,if}$ via Q —the relative size of the in-focus PSF and the coherence area. We characterize the impact of practical camera parameters—lens f -number and reproduction ratio (that together determine PSF width), and sensor pixel pitch—on the in-focus second-order texture contrast and DFF performance, and derive optimal parameter values. We excluded the impact of sensor pixel pitch in our analysis in § 4 for the sake of analytical tractability (by assuming flux is proportional to irradiance, and hence that $\kappa_I^2 = \kappa_E^2$)—the role of which will become evident below.

Theoretical analysis. We can use Equation (11) to determine how the lens’ f -number N and reproduction ratio m change second-order texture contrast, through their impact on the lens’ in-focus PSF size $\Delta_{c,if}$. At infinity focus, the *image-space* width of the diffraction-limited PSF equals λN [47, §5.2.2]. We can account for finite focus with reproduction ratio m by replacing f -number N with the *effective f -number* $N_{\text{eff}} := (m + 1)N$. [74] Lastly, we can convert to object-space widths by dividing by the reproduction ratio m , arriving at:

$$\Delta_{c,if} = \lambda N \frac{m + 1}{m}. \quad (13)$$

From Proposition 6, assuming a first-order textureless surface:

$$\kappa_I^2 \propto \frac{1}{N^2} \frac{m^2}{(m + 1)^2}, \quad (14)$$

where throughout this section we use proportionality to absorb constants that do not depend on the lens or sensor. Equation (14) suggests that the in-focus contrast of second-order texture improves with smaller f -number N (larger apertures) and larger reproduction ratio m (larger magnification). However, this analysis does not account for the finite pixel size, which imposes a bound how much we can improve in-focus contrast achieved at an optimal minimum N for any given m . We elaborate next.

We denote by $\Delta_{p,1}$ the sensor’s pixel pitch (physical width of pixels on the sensor), and by $\Delta_{p,0} := \Delta_{p,1}/m$ the *magnified pixel pitch* (width of the surface area imaged by a pixel after accounting for magnification). Pixels optically blur incident irradiance with a rect kernel of width $\Delta_{p,0}$. This additional blur is an incoherent summation process that will reduce the contrast of subjective speckle when incident irradiance (in object surface-space) varies considerably

within the $\Delta_{p,0}$ -sized spatial extent of a pixel—equivalently when the in-focus PSF width $\Delta_{c,\text{if}}$ is significantly smaller than the magnified pixel pitch $\Delta_{p,0}$. These considerations suggest that, to maximize the in-focus contrast of second-order texture at reproduction ratio m , we should set the f -number so that $\Delta_{c,\text{if}} \approx \Delta_{p,0}$. From Equations (13) and (14), the optimal in-focus texture contrast and corresponding optimal f -number become:

$$\kappa_I^{*2}(m) \propto \frac{m^2}{\Delta_{p,1}^2} \quad \text{achieved at} \quad N^*(m) := \frac{\Delta_{p,1}}{\lambda(m+1)}. \quad (15)$$

Further decreasing the f -number results in reduced contrast:

$$\kappa_I^2 = \kappa_I^{*2}(m) \frac{\Delta_{c,\text{if}}^2}{\Delta_{p,1}^2} \propto \kappa_I^{*2}(m) \frac{N^2(m+1)^2}{\Delta_{p,1}^2}, \quad \text{if} \quad N < N^*(m). \quad (16)$$

Therefore, Equation (15) shows that using sensors with smaller pixel pitch can improve in-focus contrast of second-order texture, and provides guidance on how to optimally set f -number N at a given target reproduction ratio m .

Lastly, from Proposition 2, characterizing impact on DFF performance requires accounting for both texture contrast and SNR. We follow § 5 and do so by considering how to maximize the texture contrast–SNR product $\Theta^2 = \kappa_I^2/\kappa_n^2$ for a given target reproduction ratio m . We distinguish two scenarios:

1. *Exposure time is not constrained:* We achieve optimal Θ^2 by setting f -number $N = N^*(m)$ (Equation (15)) to maximize texture contrast κ_I^2 , and exposure time t to saturation (Proposition 7) to simultaneously maximize SNR $1/\kappa_n^2$.
2. *Exposure time is fixed:* Under Poisson noise-limited conditions, $1/\kappa_n^2 \propto 1/N^2$ (Equation (2)). Thus $\Theta^2 \propto m^2/N^4(m+1)^2$ for $N \geq N^*(m)$ (Equation (14)), and $\Theta^2 \propto \kappa_I^{*2}(m)(m+1)^2/\Delta_{p,1}^2$ for $N < N^*(m)$ (Equation (16)). Therefore, we achieve optimal Θ^2 by setting f -number N to the minimum value before saturation, collecting as much light as possible through the lens aperture.

Experimental validation. Figure 14 shows experimental results for an indoor scene, using filter bandwidth $\Delta_\lambda = 10$ nm and reproduction ratio $m \approx 1/30$. We captured focal stacks using different f -number settings, adjusting exposure time to achieve proper exposure at each setting, then visualized the focus measure $\tilde{c}_j^2$ at a textureless patch as a function of f -number. The in-focus contrast of second-order texture, both relative to out-of-focus contrast and in absolute terms, becomes worse as we use more suboptimal settings— f number too large, or too small (pixel pitch limit). The results closely match our theoretical predictions in Equations (14)–(16).

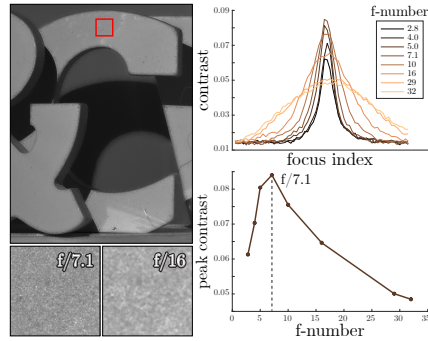


Figure 14: Experimental evaluation of second-order texture contrast in a textureless region (marked, top left) as a function of f -number. The crops (bottom left) are representative of high and low-contrast cases.

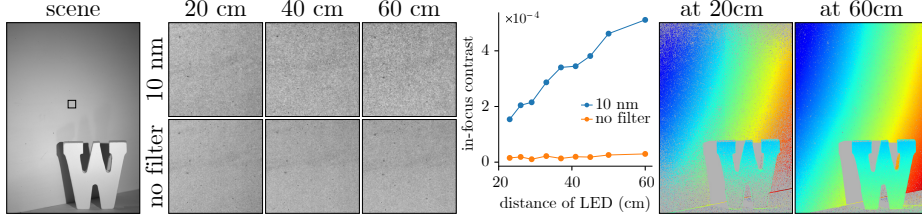


Figure 15: Impact of illumination distance on DFF performance. As the LED moves closer to the scene, the angular bandwidth increases and the coherence area decreases, leading to a reduction in in-focus contrast and DFF performance.

A.2 Illumination conditions

In § 4.2, we also pointed out that the in-focus contrast of second-order texture κ_I^2 is dependent on the coherence area Δ_i^2 of the illumination (via Q). Our theory therefore predicts that the in-focus contrast of second-order texture, and hence DFF performance, will be worse under *wider* angular footprints of illumination, that result in smaller coherence areas. This is the only section where we use artificial light sources instead of pre-existing ambient illumination, to facilitate a controlled experiment assessing the impact of coherence area.

Validation of theory. Figure 15 shows an experiment to validate this prediction. To emulate increasing angular bandwidth while keeping all other conditions constant, we moved an LED light source progressively closer to a scene (Figure 16), and captured a full focal stack at each position. As a translation stage moves an LED closer to the scene, the effective angular footprint of incident illumination increases, and the coherence area decreases.

The crops show that the in-focus contrast of second-order texture (amplified by introducing a spectral filter) is much weaker when the LED is closer to the scene (20 cm), compared to when the LED is far (60 cm). As our theory predicts, in-focus contrast decreases, and DFF performance worsens.

This experiment is complementary to the effect of decreasing *spectral bandwidth* already presented in Figure 8. These two pieces of evidence together provide confidence that the improvements we see are indeed due to diffractive (wave-optical) effects, and therefore can be attributed to the phenomenon of second-order texture that we describe in our theory.

Limitation. This experiment also highlights a limitation of our method in practice: when the angular footprint of illumination is too wide, the texture contrast κ_I^2 is still too small after spectral filtering that it still cannot overcome the noise level of the camera. Figure 17 shows experiments on a scene comprising objects of different levels of geometric complexity (in ascending order: walls, letters, meshed pen holder, hair wig). We illuminate this scene using two artificial sources of illumination, an LED spot light and an LED ring light: The spot light has a small directional bandwidth, and thus large coherence area Δ_i^2 ; the ring light has a much larger directional bandwidth, and thus much smaller coherence

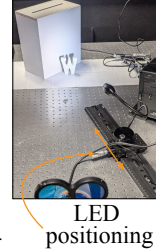


Figure 16: Capture setup with LED positioned on a translation stage.

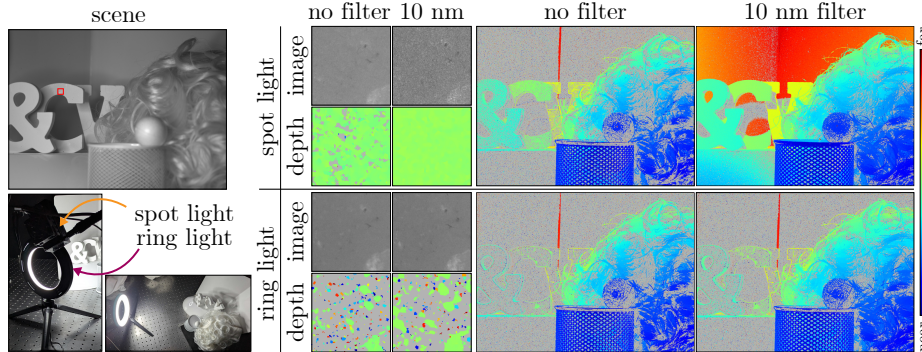


Figure 17: Using a narrowband spectral filter provides negligible improvements on this scene when we illuminate it with a nearby large ring light. The reason is the excessively small coherence area of the illumination, which results in second-order texture of very low contrast—even when spectrally filtered—relative to the noise level. By contrast, using the filter results in drastic improvements under spot-light illumination, which has a much larger coherence area.

area Δ_i^2 (Equation (9)). The crops show that the in-focus contrast of second-order texture is much weaker under the ring light than under the spot light. As a result, our method provides only minor performance improvements under the ring light, compared to the much more marked improvements under the spot light.

Impact of shadows. In many of our results, shadowed regions do not show as much improvement as non-shadowed regions (*e.g.* the shadow of the letter in Figure 17). One obvious reason is poor SNR in these regions. Another reason is that shadowed regions are primarily illuminated by indirect light, which typically has a much larger angular footprint than direct light, and thus much smaller coherence area. As a result, second order texture contrast is much weaker in these region, offering little to no improvement in DFF performance.

A.3 Material constraints

While we considered the physics of *surface scattering*, our theory does not account for possible subsurface scattering in the materials of the scene. Subsurface scattering is a phenomenon where light penetrates the surface of a material, scatters internally, and exits at a different location. This can significantly reduce the contrast of second-order texture, as it effectively results in the incoherent superposition of far more speckle patterns⁶ than would be present in a purely surface-scattering material.

Figure 18 shows experiments on a scene comprising objects made from different materials, including materials with varying levels of subsurface scattering (plastic bottle and wax candle on the back right, ceramic cups on the front right). The experiment shows that the performance improvements from our method gradually decrease with increasing amount of scattering (*e.g.*, ceramic cups, crop A), until

⁶ due to the finite spectral bandwidth, i.e. smaller temporal coherence length than the scattering distance within the object. Refer [46]

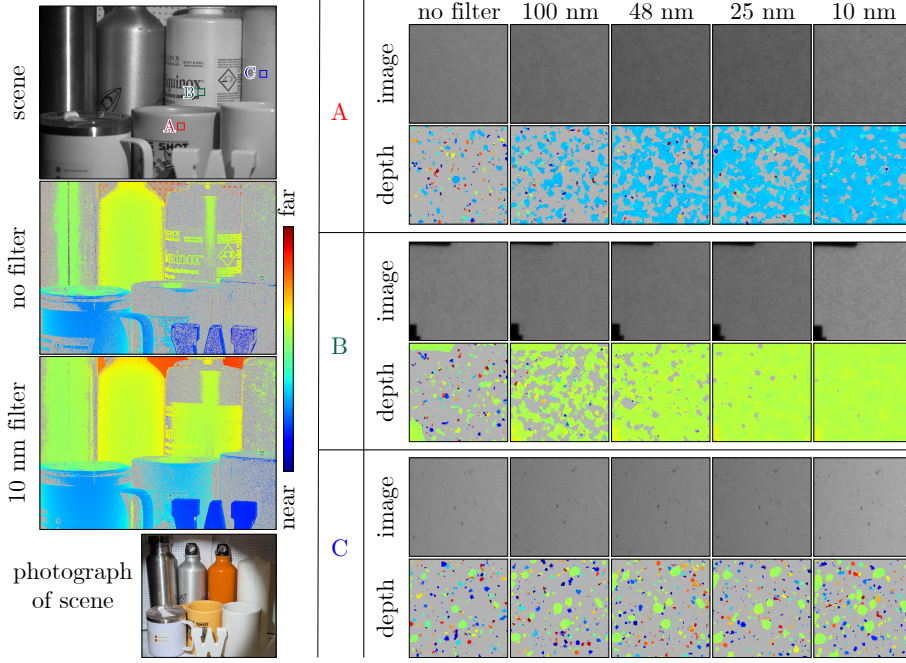


Figure 18: Performance improvements from using a narrowband filter will vary for different optical materials. Crop B, which corresponds to a nearly opaque material (label) shows the strongest improvement. Crop A, which corresponds to a weakly translucent material (ceramic cup) shows reduced but still significant improvement. Crop C, which corresponds to a strongly translucent material (wax) shows negligible improvement. The underlying cause is that strong subsurface scattering in translucent materials (*e.g.*, wax) greatly reduces the contrast of second-order texture (subjective speckle).

eventually they become negligible in very translucent materials (*e.g.*, wax candle, crop C). By contrast, our method provides strong performance improvements at parts of the scene that are largely opaque (*e.g.*, the label on the plastic bottle, crop B). This behavior is expected, as subsurface scattering strongly reduces the contrast of subjective speckle—an effect extensively documented and analyzed in optics [46, §6.4.3].

B Interplay between texture contrast and noise

§ 5 provided a high-level overview of the tradeoff between second-order texture contrast and SNR for depth recovery performance in terms of the probability of error $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{oof}}^2)$. This section provides an additional layer of detail about what happens under the hood—showing how the texture-contrast-SNR product, and the statistics of the focus measure (mean and standard devi-

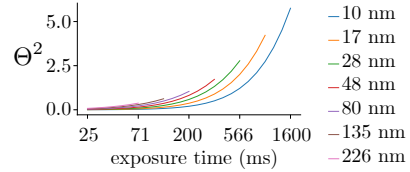


Figure 19: Plot of $\Theta^2(t, \Delta_k)$. Parameters match those in in Figures 7 and 20.

ation, derived in Appendix H) are impacted by spectral bandwidth and exposure time, and how that, in turn, results in the behaviour of the probability of error shown in Figure 7.

Texture Contrast-SNR product. Just like we visualized the probability of error as a grid as in Figure 7, we can also visualize the texture contrast-SNR product $\Theta^2 = \kappa_I^2/\kappa_n^2$ defined in § 5 directly. An increase in Θ^2 directly maps to a decrease in the probability of error through Lemma 12. Figure 19 shows Θ^2 as a function of exposure time t for different values of spectral bandwidth. Each line stops at the maximum exposure time t that can be used for a given spectral bandwidth, before the camera saturates.

Focus measure statistics. Lemma 11 (from Appendix H) provides approximate expressions for the mean and variance of the sample contrast focus measure $\tilde{c}^2$. We have the mean and standard deviation of the in-focus and out-of-focus focus measures:

$$\mathbb{E}[\tilde{c}_{\text{if}}^2] \approx \kappa_I^2 + \kappa_n^2, \quad (17)$$

$$\mathbb{E}[\tilde{c}_{\text{oof}}^2] \approx \kappa_n^2, \quad (18)$$

$$\begin{aligned} \sqrt{\mathbb{V}[\tilde{c}_{\text{if}}^2]} &\approx \sqrt{\frac{2}{T-1}}(\kappa_I^2 + \kappa_n^2), \\ \sqrt{\mathbb{V}[\tilde{c}_{\text{oof}}^2]} &\approx \sqrt{\frac{2}{T-1}}(\kappa_n^2). \end{aligned} \quad (19)$$

As a consequence, the mean *separation* between the (noisy) in-focus and out-of-focus sample contrast focus measures

$$\mathbb{E}[\tilde{c}_{\text{if}}^2] - \mathbb{E}[\tilde{c}_{\text{oof}}^2] \approx \kappa_I^2 \quad (20)$$

In addition, we know from Proposition 6 that for a first-order textureless surface ($\kappa_{\text{E}}^2 = 0$)

$$\kappa_I^2 = \frac{1}{MQ} \quad (21)$$

and from Equation (2)

$$\kappa_n^2 = \frac{1}{t\zeta\mu_\Phi} \left(1 + \frac{\sigma_{\text{read}}^2}{t\zeta\mu_\Phi} \right) \quad (22)$$

Thus, given the parameters of the capture system, we can plug in these values to compute the mean and standard deviation of the in-focus and out-of-focus focus measures. The bottom row in Figure 20 plots these quantities as an extension of Figure 7. The ‘theory’ curves show the predicted separation $\mathbb{E}[\tilde{c}_j^2] - \kappa_n^2$ for $j = j_{\text{if}}$ and $j = j_{\text{oof}}$ (the latter is uniformly zero, as evident from Equation (18)), also labeling a region $\pm 1 \times$ standard deviations surrounding it. These regions give us a sense of the width of the distributions of $\tilde{c}_{\text{if}}^2$ and $\tilde{c}_{\text{oof}}^2$. The *separation* between the in-focus and out-of-focus focus measure distributions provides immediate insight

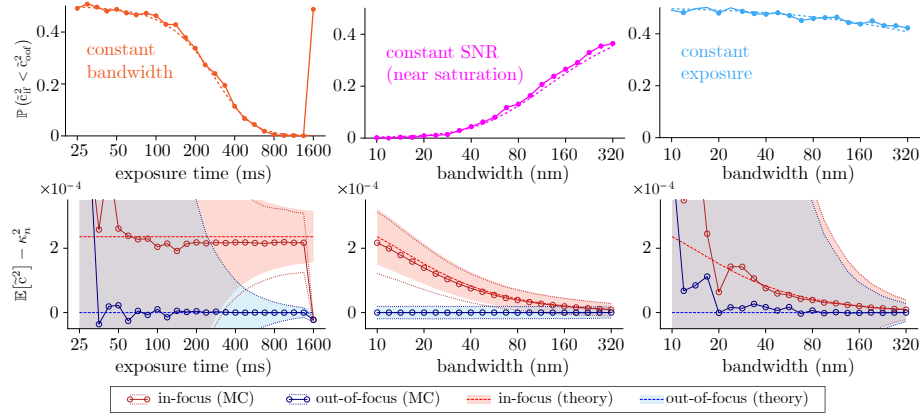


Figure 20: Visualization of the mean and standard deviation of the focus measure (specifically, $\mathbb{E}[\tilde{c}^2] - \kappa_n^2$, and $\sqrt{\mathbb{V}[\tilde{c}^2]}$) along the three paths shown in Figure 7. ‘theory’ corresponds to direct analytical evaluation using Equations (17)–(19), while ‘MC’ corresponds to Monte-Carlo estimates of the same quantities using samples from simulation. The MC estimates (using 10k samples per point) have higher variance in regimes of very low exposure.

into DFF performance. When the distributions of $\tilde{c}_{\text{if}}^2$ and $\tilde{c}_{\text{oof}}^2$ are well-separated, the chances of a patch being misclassified as in-focus or out-of-focus is low, and hence $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{oof}}^2)$ is low. When they are close together, the opposite is true (worst case: 50-50 chance).

This plots provide a clearer picture of how capture parameters (exposure time and spectral bandwidth) individually influence the statistics of the focus measures in and out of focus with more granularity.

1. If the spectral bandwidth is reduced at a fixed exposure time, the mean separation between the in-focus and out-of-focus focus measures increases in accordance with Equation (20) (right sections).
2. When the exposure time is increased (left section), the mean separation is unaffected, but the standard deviation of the focus measures decreases, reducing overlap between the distributions and thus reducing the probability of error.
3. If the spectral bandwidth is reduced, while also increasing the exposure time to maintain proper exposure (middle section), the mean separation increases, while the in-focus standard deviation increases slightly (in accordance with Equation (19)).

Match with Monte-Carlo simulation. Overlaid on the theory-based curves and shaded regions (from Equations (19)–(20)) are the corresponding Monte-Carlo (MC) estimates from simulation. The close match between theory and simulation (for both focus measure statistics, and probability of error) shows that our theoretical analysis is accurate, and that the assumptions made in Appendix H are reasonable. In particular, we assumed that the distributions of $\tilde{c}_{\text{if}}^2$ and $\tilde{c}_{\text{oof}}^2$ are approximately Gaussian, which is validated by the close match

Table 1: Fraction of unrecovered pixels η with $\xi > \xi_{\text{th}} (= 4.0)$ for all scenes in our results (main paper and supplement). Dashes indicate unavailable (not captured) datasets. Visit <https://imaging.cs.cmu.edu/second-order-texture> for interactive visualizations of all scenes

	scene	no filter	100 nm	48 nm	25 nm	10 nm
	illusion	0.834	-	-	-	0.004
	side table	0.560	0.469	-	-	0.243
	helmet	0.742	-	-	0.461	0.313
	knight	0.866	0.532	0.427	0.388	0.362
	set cards	0.842	0.732	0.621	0.460	0.287
	bin	0.853	-	-	0.461	0.583
	sculpture	0.552	-	-	-	0.113
	corner	0.715	0.645	0.615	0.613	0.644
	columns	0.823	0.664	0.668	0.631	0.642
	trash can	0.633	0.493	0.501	0.462	0.216
	hair & mesh	0.561	0.423	0.359	0.261	0.141
	bottles & mugs	0.534	0.415	0.371	0.348	0.295
	cup	0.893	-	-	-	0.518

between theory and simulation. One may notice that the theoretical curves consistently overestimate the mean contrast in the high-contrast regimes (left and middle plots). This can be directly attributed to an assumption we make in the derivation of Lemma 11, *i.e.* we assume that the intensities of pixels are independent. See Appendix H for details.

C Depth recovery performance evaluation

We report quantitative metrics across all our scenes, and validate our z-score-based evaluation metric. Visit our webpage (<https://imaging.cs.cmu.edu/second-order-texture>) for interactive visualizations of all scenes in the main paper and supplement, including depth maps, surface renders, and comparisons.

C.1 Quantitative comparison

Table 1 aggregates DFF performance metrics across all captured scenes we show in both the main paper and the supplement. As we do not have ground-truth depth, to assess performance, we use the fraction of unrecovered pixels η we defined in § 6, which serves as a measurable proxy for the probability of error $\Pr(\tilde{c}_{\text{if}}^2 < \tilde{c}_{\text{of}}^2)$ (thus lower values of η represent better DFF performance). The table shows that, consistently across all scenes, performance improves as spectral bandwidth decreases—and while adjusting exposure time to maintain proper exposure—in agreement with our theoretical predictions. The main paper shows quantitative comparisons also for cases of underexposure (Figure 10).

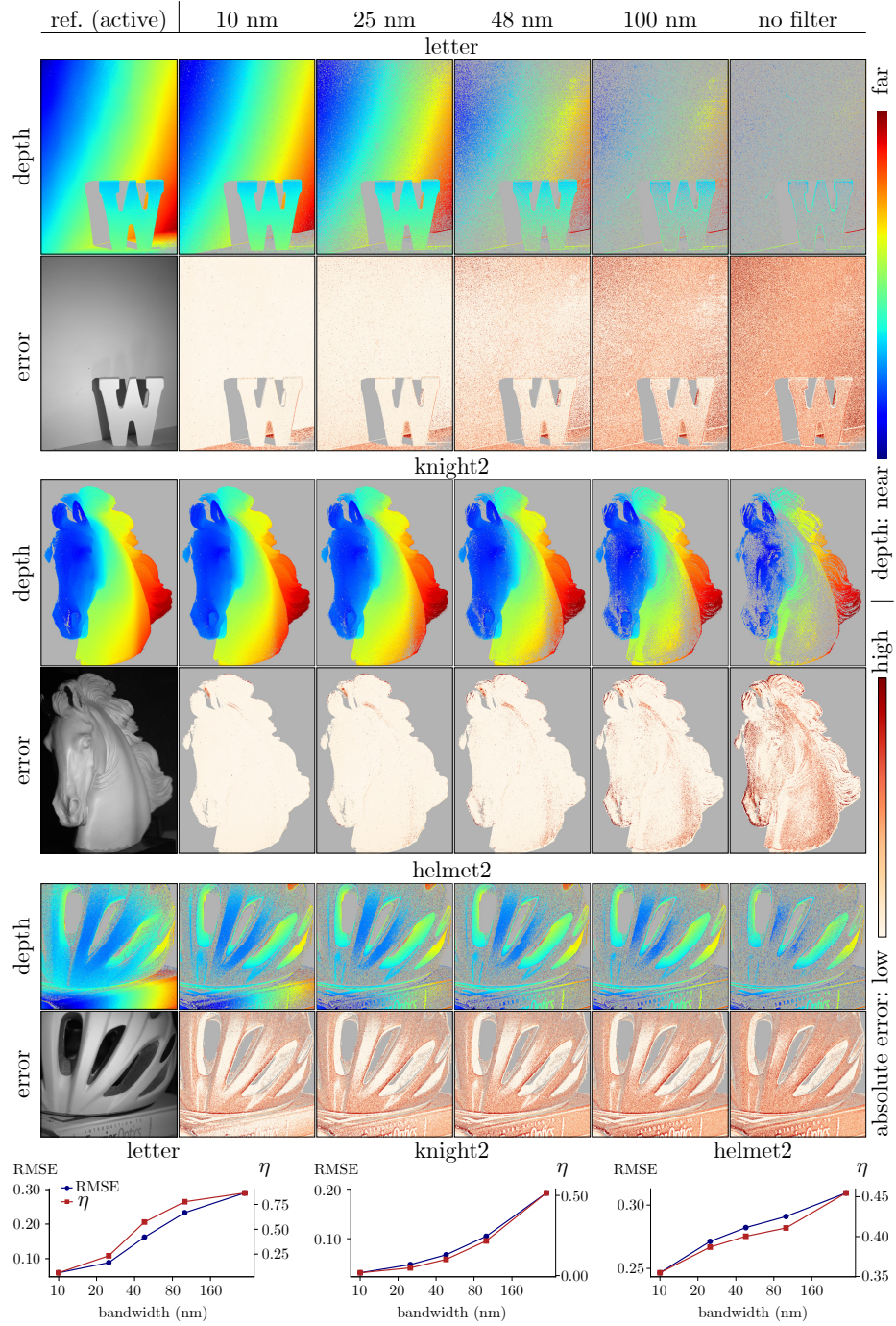


Figure 21: Validation of the η metric using RMSE. (scenes captured indoors). For the reference depth map, a projector was used to illuminate the scene with a fine checkerboard pattern to create artificial texture.

Table 2: Spectral filters we use for experiments.

filter model	center wavelength (nm)	full-width half-max (FWHM) (nm)
EO #65-216	532	10
EO #84-115	534	25
EO #67-045	534.5	48
EO #33-331	550	100

C.2 Validation of z-score-based metric with active DFF reference

Figure 21 shows comparisons with a reference from *active* DFF—a projector illuminates the scene with a high-frequency pattern to create artificial texture. We compute RMSE relative to this reference depth, excluding pixels with low-confidence reference.

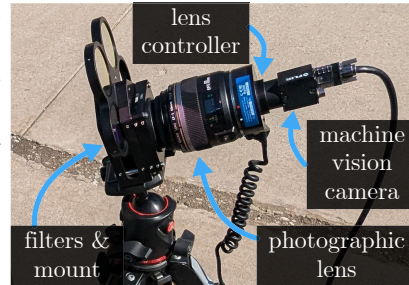
We observe: 1. RMSE decreases as spectral bandwidth narrows, validating our theory. 2. RMSE is strongly correlated with our η metric (fraction of unrecovered pixels). This strong correlation is expected: From DFF theory [9, 68], wrong depth estimates with high z-score are *extremely unlikely* (a patch that, when out of focus, serendipitously has far higher noise than at other focus settings). Essentially all wrong depth estimates arise from flat focus-measure curves (low texture or high noise), and hence have low z-score. Therefore, our η metric quantifies DFF performance as accurately as RMSE, while being much more practical for outdoor experiments. Capturing active DFF for reference is challenging under sunlight.

D Implementation details

We provide details for simulations and experiments in the main paper and supplement.

D.1 Capture setup

Our setup comprises only off-the-shelf components for standard depth from focus. We use a machine vision camera (FLIR Black-Fly S BFS-U3-122S6M-C) with a monochromatic sensor (Sony IMX304, 1.4 cm diagonal, 3.5 μ m pixel pitch, 12 bits). The camera provides access to RAW measurements. We equip the camera with an electronic lens mount (ISSI Canon EF lens controller) for programmatic focus and aperture control, and a photographic lens (Canon EF-S 60mm f/2.8 Macro USM) for imaging. We use standard filter rings and an optical filter wheel to mount and easily rotate filters at the front of the lens. Lastly, we use narrowband spectral filters of 2 inch diameter (Edmund Optics, different FWHM bandwidths, center wavelength $\approx$ 530 nm). Figure 22 shows a photograph, and Table 2 lists all filters we use in our experiments.

**Figure 22:** Capture setup.

Algorithm 1 Depth recovery using single-step aggregation.

```

1: procedure SINGLESTEPDFF( $\{\tilde{\mathbf{I}}_j\}_{j=1}^J$ )
2:   for  $j \in \{1, \dots, J\}$  do
3:      $\tilde{c}_j^2 \leftarrow \text{LOCALFM}(\tilde{\mathbf{I}}_j)$  ▷ focus measure stack
4:      $\tilde{c}_j^2 \leftarrow \tilde{c}_j^2 * g_{\sigma_{\text{agg}}}$  ▷ Gaussian kernel aggregation
5:   end for
6:    $j^*(\vec{p}) \leftarrow \text{argmax}_j \tilde{c}_j^2(\vec{p})$ 
7:    $d \leftarrow \text{INTERPDEPTH}(\tilde{c}^2, j^*)$ 
8:    $\xi \leftarrow \text{ZSCORE}(\tilde{c}^2, j^*)$  ▷ pixel-wise z-scores
9:   return  $d, \xi$ 
10: end procedure

11: procedure LOCALFM( $\tilde{\mathbf{I}}_j$ )
12:   return  $\frac{[\tilde{\mathbf{I}}_j * \mathbf{L}_{3 \times 3}]^2}{[\tilde{\mathbf{I}}_j * \mathbf{U}_{3 \times 3}]^2}$  ▷ normalized squared Laplacian
13: end procedure

14: procedure ZSCORE( $\tilde{c}^2, j^*$ )
15:    $\tilde{c}_{\text{med}}^2(\vec{p}) = \text{median}_j \tilde{c}_j^2(\vec{p})$ 
16:    $\text{MAD}(\vec{p}) = \text{median}_j |\tilde{c}_j^2(\vec{p}) - \tilde{c}_{\text{med}}^2(\vec{p})|$ 
17:   return  $\frac{\tilde{c}_{j^*}^2 - \tilde{c}_{\text{med}}^2}{\text{MAD}}$  ▷ pixel-wise robust z-score
18: end procedure

```

D.2 Depth recovery algorithms

Algorithms 1 and 2 detail the methods we use in § 6 for recovering depth from a focal stack. Algorithm 1 performs one aggregation over a patch of fixed size to compute a Laplacian-based focus measure. Algorithm 2 is a pixel-adaptive variant: At each pixel, it performs multiple consecutive aggregations—each progressively increasing the effective patch size—until the robust z-score at that pixel reaches a certain threshold. Thus this algorithm varies per-pixel patch size to adapt to areas of different texture contrast. In both algorithms, the routine INTERPDEPTH uses three-point Gaussian interpolation of the focus measure to determine a peak dithered between focus indices [9].

D.3 Monte Carlo simulation

Our Monte Carlo simulation is a direct implementation of the description of second-order texture in Appendix F and the full noise model in Appendix G. Below, we explain the sequence of steps alongside implementation details and references to corresponding equations in the supplement and main text. We use a pseudorandom number generator for all sampling steps. Algorithm 3 summarizes the full Monte Carlo pipeline.

1. We simulate a surface that is fronto-parallel to the camera. We model its texture using the blurred outgoing radiance $\mathcal{L}^{(\Delta_1)}(\vec{q}_m)$ (Equation (73)) as

Algorithm 2 Pixel-adaptive depth recovery using progressive aggregation.

```

1: procedure PIXELADAPTIVEDFF( $\{\tilde{I}_j\}_{j=1}^J, \xi_{\text{th}}$ )
2:    $\xi(\vec{p}) \leftarrow 1; d(\vec{p}) \leftarrow \text{NaN}; \chi(\vec{p}) \leftarrow \text{false}$   $\triangleright$  init. z-scores, depth, completion
3:    $l \leftarrow 0$ 
4:   for  $j \in \{1, \dots, J\}$  do
5:      $\tilde{c}_j^2 \leftarrow \text{LOCALFM}(\tilde{I}_j)$   $\triangleright$  init. focus measure stack
6:   end for
7:   while  $l < l_{\text{max}}$  and  $\neg \text{all}(\chi)$  do
8:     for  $j \in \{1, \dots, J\}$  do
9:        $\tilde{c}_j^2 \leftarrow \text{WEIGHTEDAGGSTEP}(\tilde{c}_j^2, \xi)$ 
10:    end for
11:     $j^*(\vec{p}) \leftarrow \text{argmax}_j \tilde{c}_j^2(\vec{p})$ 
12:     $\xi \leftarrow \text{ZSCORE}(\tilde{c}_j^2, j^*)$   $\triangleright$  update pixel-wise z-scores
13:     $\chi_{\text{new}} \leftarrow \xi > \xi_{\text{th}} \wedge \neg \chi$ 
14:     $d[\chi_{\text{new}}] \leftarrow \text{INTERPDEPTH}(\tilde{c}_j^2, j^*)[\chi_{\text{new}}]$ 
15:     $\chi \leftarrow \chi \vee \chi_{\text{new}}$ 
16:     $l \leftarrow l + 1$ 
17:  end while
18:  return  $d$ 
19: end procedure

20: procedure WEIGHTEDAGGSTEP( $\tilde{c}_j^2, \xi$ )
21:   return  $\frac{[(\tilde{c}_j^2 \cdot \xi) * g_{\sigma_1}]}{[\xi * g_{\sigma_1}]}$   $\triangleright$  z-score weighted aggregation
22: end procedure

```

a two-dimensional grid of spacing Δ_i , where $\{\vec{q}_m\}$ is the set of grid centers (based on Proposition 10 in Appendix F).

2. We consider first-order textureless appearance, *i.e.*, $\bar{\mathcal{L}}^{(\Delta_i)}$ is constant across grid locations, or $\kappa_{\bar{\mathcal{L}}^{(\Delta_i)}}^2 = \kappa_E^2 = 0$. Additionally, we assume that $\bar{\mathcal{L}}^{(\Delta_i)}$ at each location is proportional to spectral bandwidth Δ_k according to Equation (159), to model brightness attenuation due to narrowing bandwidth; thus: $\bar{\mathcal{L}}^{(\Delta_i)} = (\bar{\mathcal{L}}^{(\Delta_i)})_{\text{orig}} \Delta_k$, where $(\bar{\mathcal{L}}^{(\Delta_i)})_{\text{orig}}$ is the original value of $\bar{\mathcal{L}}^{(\Delta_i)}$ without spectral filtering.
3. We sample each cell of the coherence area grid independently from a Gamma distribution according to Proposition 10. We report related parameters in Table 3a.
4. We compute the irradiance reaching the sensor at the in-focus setting according to the discrete convolution of the PSF P_c with the outgoing radiance grid $\mathcal{L}^{(\Delta_i)}(\vec{q}_m)$ in Equation (77). We model the PSF as a discrete Gaussian kernel with standard deviation $\Delta_c/2$ as in Equation (113). We compute the in-focus PSF width in object space using Equation (13). We scale the irradiance at the sensor by a factor $\propto m^2/N_{\text{eff}}^2$ to model how the energy reaching the sensor

Algorithm 3 Monte Carlo focus measure sampler

default parameters $g, \sigma_{\text{pre}}, \sigma_{\text{post}}$ *etc.* from Tables 3a–3c
Pixel resolution $H \times W$.

- 1: **procedure** MCFMSAMPLER(Δ_k, t)
- 2: $\bar{\mathcal{L}}^{(\Delta_i)} \leftarrow g_p \Delta_k \cdot \bar{\mathcal{L}}^{(\Delta_i, \text{orig})}$ ▷ Equation (171)
- 3: $\mathcal{L}^{(\Delta_i)}(\vec{q}_m) \leftarrow \text{Gamma}\left(M, \frac{M}{\bar{\mathcal{L}}^{(\Delta_i)}}\right)$ ▷ Equation (82)
- 4: $E_{\text{if}} \leftarrow \text{PSFBLUR}(\mathcal{L}^{(\Delta_i)})$
- 5: $\tilde{I}_{\text{if}}(\vec{p}) \leftarrow \text{SENSORREADOUT}(E_{\text{if}}, t)$
- 6: $\tilde{c}_{\text{if}}^2(\vec{p}) \leftarrow \text{SAMPLECONTRASTFM}(\tilde{I}_{\text{if}})$ ▷ in-focus sample
- 7: $\tilde{c}_{\text{oof}}^2(\vec{p}) \leftarrow \text{SAMPLECONTRASTFM}(\mu_{\text{I}} \cdot \text{ones}(H, W))$ ▷ out of focus sample
- 8: **return** $\tilde{c}_{\text{if}}^2, \tilde{c}_{\text{oof}}^2$
- 9: **end procedure**

- 10: **procedure** PSFBLUR($\mathcal{L}^{(\Delta_i)}$)
- 11: $E(\vec{q}) \leftarrow \sum_m \Delta_{\text{p}}^2 P_{\text{c}}(\vec{q} - \vec{q}_m) \mathcal{L}^{(\Delta_i)}(\vec{q}_m)$ ▷ Equation (77)
- 12: **return** E
- 13: **end procedure**

- 14: **procedure** SENSORREADOUT(E, t)
- 15: $\Phi(\vec{p}) \leftarrow E(\vec{p}) \cdot \Delta_{\text{p}}^2$ ▷ sample irradiance coarsely at pixel resolution
- 16: $P \leftarrow \text{Poisson}((\zeta \Phi + D)t)$
- 17: $Z_{\text{pre}} \leftarrow \mathcal{N}(0, \sigma_{\text{pre}}^2); Z_{\text{post}} \leftarrow \mathcal{N}(0, \sigma_{\text{post}}^2)$
- 18: $\tilde{I}(\vec{p}) \leftarrow \min(\lfloor 1/g(\min(P, P_{\text{max}}) + Z_{\text{pre}}) + Z_{\text{post}} \rfloor, I_{\text{max}})$ ▷ Equation (120)
- 19: **return** $\tilde{I}$
- 20: **end procedure**

- 21: **procedure** SAMPLECONTRASTFM($\tilde{I}$)
- 22: **return** $\frac{\tilde{s}^2}{\tilde{m}^2}$ ▷ Equation (1)
- 23: **end procedure**

through the lens varies as a function of m and N . We report related parameters in Table 3b.

5. We sum the resulting irradiance grid $E(\vec{p})$ —now mapped to image space—over the pixel area $\Delta_{\text{p}} \times \Delta_{\text{p}}$ to get the flux $\Phi(\vec{p})$ at each pixel at the camera’s resolution.
6. Given the flux at each pixel, we use the full noise model of Equation (120) to sample the noisy intensity $\tilde{I}(\vec{p})$ at each pixel. We report related parameters in Table 3c.
7. Given the noisy intensity measurements, we compute the focus measure $\tilde{c}^2$ (§ 3) for a square-shaped patch of size $\sqrt{T} \times \sqrt{T}$ (5×5) centered at the pixel of interest.
8. We repeat this process multiple times to obtain Monte-Carlo estimates of $\Pr(\tilde{c}_{\text{if}}^2 > \tilde{c}_{\text{oof}}^2)$, for different values of Δ_k and t . We plot these estimates in Figure 7.

Table 3: Parameters for Monte Carlo simulation in Figure 7.

(a) Second-order texture parameters.			(b) Camera lens settings.		
parameter	value	description	parameter	value	description
σ_r	3 μm	RMS height of surface	m	0.01	reproduction ratio
Δ_i	12 μm	width of coherence cell grid	N	7	f -number
$\bar{\lambda}$	532 nm	central wavelength			

(c) Camera sensor parameters.		
parameter	value	description
g	10 e^-/DN	sensor gain
σ_{pre}	10 e^-	pre-amp noise variance
σ_{post}	15 DN	post-amp noise variance
$\log_2(I_{\text{max}})$	12	ADC bit depth
v	0.4	quantum efficiency
D	1.0 e^-/s	dark current
P_{max}	35 000 e^-	full-well capacity
Δ_p	3.45 μm	pixel pitch

E Appearance model for second-order texture

We present a derivation of Proposition 4 consistent with physical-optics principles for an explicit micro-geometry of the surface, mapping wave quantities to conventional radiometric quantities used in computer vision. We then explain how traditional BRDF-based image formation models used in computer vision are a first-order abstraction of appearance based on the statistics of the surface heightfield, and then quantify the strength of second-order texture in Appendix F. Our formulations largely follow Goodman [46, Sections 5.8 & 6.3] and Steinberg and Yan [62].

E.1 Incident illumination model

Consider the 2D plane macroscopically tangent to the surface of the object at the point of interest with macroscopic outward normal $\hat{n}$ (Figure 3). At any point $\vec{\rho} \in \mathbb{R}^2$ in this plane, we model the incoming illumination as consisting of a superposition of monochromatic plane waves of varying direction and wavelength:

$$U_i(\vec{\rho}) = \int_{k \in \mathbb{R}^+} \int_{\mathcal{H}^2(\hat{n})} \mathcal{G}(\hat{\omega}, k) e^{ik\hat{\omega}_{xy} \cdot \vec{\rho}} d\sigma(\hat{\omega}) dk \quad (23)$$

Here, $\hat{\omega}$ is the outward unit vector antiparallel to the direction of the incoming plane wave, $k = 2\pi/\lambda$ represents its (angular) wavenumber, and

$$\mathcal{H}^2(\hat{n}) = \{\hat{\omega} \in \mathbb{R}^3, \|\hat{\omega}\| = 1, \hat{\omega} \cdot \hat{n} > 0\} \quad (24)$$

is the upper-half space of valid outward incident direction vectors. $\hat{\omega}_{xy}$ is the component of $\hat{\omega}$ parallel to the plane, *i.e.*, $\hat{\omega} = [\hat{\omega}_{xy}, \omega_z \hat{n}]$.

In the context of computer vision with direct and indirect illumination, one may consider these plane waves to have been generated by a variety of light sources in the scene. The amplitude of these plane waves $(\hat{\omega}, k)$ are modeled to be a result of the random process $\{\mathcal{G} : \mathcal{H}^2(\hat{n}) \times \mathbb{R}^+ \rightarrow \mathbb{C}\}$ (Units: $[\text{W}^{0.5}/\text{sr}]$). Any observable radiometric quantity such as irradiance or intensity is to be computed by averaging the result over this ensemble. [71]

Mutual incoherence assumption. We assume these plane waves to be *mutually incoherent*, *i.e.* $\mathcal{G}(\hat{\omega}, k)$ are uncorrelated for any non-identical pair $(\hat{\omega}, k)$, that is,

$$\langle \mathcal{G}(\hat{\omega}_1, k_1) \mathcal{G}^*(\hat{\omega}_2, k_2) \rangle = R_i(\hat{\omega}_1, k_1) \delta_k(k_1 - k_2) \delta_{\hat{\omega}}(\hat{\omega}_1 - \hat{\omega}_2) \quad (25)$$

Here, $\langle \cdot \rangle$ represents an average⁷ over the ensemble $\mathcal{G}$, and $R_i(\hat{\omega}_1, k_1)$ equals the *foreshortened radiance*

$$R_i(\hat{\omega}, k) = L_i(\hat{\omega}, k)(\hat{\omega} \cdot \hat{n}). \quad (26)$$

$L_i(\hat{\omega}, k)$ (Units: $[\text{W m}^{-1} \text{sr}^{-1}]$) is the incoming spectral radiance environment map introduced in § 4. The assumption of mutual incoherence allows us make

⁷ We use the symbol $\langle \cdot \rangle$ to distinguish it from expectations over texture and heightfields, for which we reserve $\mathbb{E}[\cdot]$ as used in the main text.

a correspondence between plane wave amplitudes with incoming radiance as defined in radiometry. In addition, we assume the environment map is separable in k and $\hat{\omega}$, *i.e.* $L_i(\hat{\omega}, k) = \ell_i(k) \cdot L_i(\hat{\omega})$, where the normalized spectral profile is $\ell_i(k)$, which integrates to unity.

The above is a suitable model for real-world ambient illumination, where plane waves from different incident directions due to a variety of far-field light sources (*e.g.* sun, lamps, ceiling lamps, indirect reflections) and wavenumbers are statistically uncorrelated due to randomness of the emission process at each point on the light source(s). By considering the plane-wave amplitudes $\mathcal{G}$ to be position-independent, we have thus described an environment lighting model. This is sufficient for our purposes since we only seek to describe the scattering interaction with the surface in the local neighborhood at the point of interest.

We re-express the complex incident field created by the superposition of these plane waves (Equation (23)) as

$$U_i(\vec{\rho}) = \int_k U_i(\vec{\rho}, k) dk \quad (27)$$

$$U_i(\vec{\rho}, k) = \int_{\mathcal{H}^2(\hat{n})} \mathcal{G}(\hat{\omega}, k) e^{jk\hat{\omega}_{xy} \cdot \vec{\rho}} d\sigma(\hat{\omega}) \quad (28)$$

Example: incident irradiance. We now provide a simple example of computing an observable radiometric quantity using the above illumination model. The incident *irradiance* (Units:[W m⁻²]) at a point on the local 2D plane is given by the average magnitude-squared of the incident field:[72]

$$E_i(\vec{\rho}) = \langle |U_i(\vec{\rho})|^2 \rangle \quad (29)$$

$$= \int_{\mathbb{R}^+} \langle |U_i(\vec{\rho}, k)|^2 \rangle dk \quad (30)$$

$$= \int_{\mathbb{R}^+} E_i(\vec{\rho}, k) dk. \quad (31)$$

The cross-terms vanish since $\mathcal{G}(\cdot, \cdot)$ is uncorrelated for non-identical wavenumbers. We now similarly expand the *spectral density* of incident irradiance $E_i(\vec{\rho}, k)$ for wavenumber k , as (Units:[W m⁻¹])

$$E_i(\vec{\rho}, k) = \langle |U_i(\vec{\rho}, k)|^2 \rangle \quad (32)$$

$$= \int_{\mathcal{H}^2(\hat{n})} \int_{\mathcal{H}^2(\hat{n})} \langle \mathcal{G}(\hat{\omega}_1, k) \mathcal{G}^*(\hat{\omega}_2, k) \rangle e^{jk\vec{\rho} \cdot (\hat{\omega}_{xy,1} - \hat{\omega}_{xy,2})} d\sigma(\hat{\omega}_1) d\sigma(\hat{\omega}_2) \quad (33)$$

$$= \ell_i(k) \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) d\sigma(\hat{\omega}) \quad (34)$$

which is independent of position $\vec{\rho}$, as expected, since we use an environment map. Hence, we can drop the position dependence and use $E_i(k)$ and E_i instead. In summary, thanks to the assumption of mutual incoherence of the incident plane waves, the above example illustrates that for the computation of observable

radiometric quantities, we can consider the effect of each incident plane wave $(\hat{\omega}, k)$ independently of others (using $L_i(\hat{\omega}, k)$) and then sum up their contributions linearly.

E.2 Coherence function of the incident illumination

Before proceeding to compute the scattered field and irradiance on the sensor of the camera, we introduce the following definition (Equation (9) in main text):

Definition 8: Coherence Function

We define the *coherence function* $A_i(\cdot)$ of the illumination environment map $L_i(\cdot)$ as its projected Fourier transform:

$$A_i(\vec{\rho}; k) := \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) e^{jk\hat{\omega}_{xy} \cdot \vec{\rho}} d\sigma(\hat{\omega}) \quad (35)$$

$$L_i(\hat{\omega}) = \frac{k^2}{4\pi^2} \int_{\mathbb{R}^2} A_i(\vec{\rho}; k) e^{-jk\hat{\omega}_{xy} \cdot \vec{\rho}} d\vec{\rho}, \quad \hat{\omega} \in \mathcal{H}^2(\hat{n}) \quad (36)$$

It is easily shown using the results of the previous section that the coherence function represents the two-point correlation of the incident field $U_i(\cdot, k)$, where the Fourier variable $\vec{\rho}$ represents the *separation* between these two points⁸

$$A_i(\vec{\rho}_1 - \vec{\rho}_2, k) = \langle U_i(\vec{\rho}_1, k) U_i^*(\vec{\rho}_2, k) \rangle \quad (37)$$

Observe that by definition, $A_i(\vec{0}, k) = E_i(k)/\ell_i(k)$.

Example: Gaussian environment lighting. As a concrete example, consider the following lighting configuration⁹

$$L_i(\hat{\omega}) = L_p \exp\left(-\frac{\|\hat{\omega}_{xy} - \hat{\Omega}_{xy}\|^2}{2\Delta_\omega^2}\right) \quad (38)$$

The angular distribution of radiance is a Gaussian lobe (when projected onto the x-y plane) centered at $\hat{\Omega}_{xy}$ with (projected) angular spread $2\Delta_\omega$ and peak value L_p

The corresponding coherence function (Fourier dual of lighting) according to Definition 8 is given by

$$A_i(\vec{\rho}; k) = 2\pi\Delta_\omega^2 L_p \exp\left(-\frac{\Delta_\omega^2 k^2}{2} \|\vec{\rho}\|^2\right) \exp(jk\hat{\Omega}_{xy} \cdot \vec{\rho}) \quad (39)$$

$$= \frac{E_i(k)}{\ell_i(k)} \exp\left(-\frac{\|\vec{\rho}\|^2}{2\Delta_i^2}\right) \exp(jk\hat{\Omega}_{xy} \cdot \vec{\rho}) \quad (40)$$

⁸ Readers with a background in optics may recognize this as a form of the van-Cittert Zernike theorem.

⁹ The projected Gaussian here is technically truncated at $\|\hat{\omega}_{xy}\| < 1$. We consider the width Δ_ω to be sufficiently small such that the truncation effect is negligible.

where $\Delta_i = 1/k\Delta_\omega$ is the *spatial coherence length*, a measure of how quickly the magnitude of field correlation (coherence) decreases with separation, and $E_i(k) = 2\pi\Delta_\omega^2 L_p \ell_i(k)$ is the spectral density of incident irradiance at the surface. We will use the above example in Appendix F to prove Proposition 6 from the main text.

E.3 Scattering and Measurement

Our goal now is to express the irradiance measured at a point on the camera sensor given the incident environment map $L_i(\hat{\omega})\ell_i(k)$ and the surface heightfield $h(\cdot)$, which together (along with the camera lens) determine the scattered field reaching the camera sensor. For a given point on the camera sensor, consider $\vec{p}$ as the point on the tangent plane it is mapped to via pinhole projection (center of the lens aperture). The viewing angle of the camera for point $\vec{p}$ is $\hat{v}$ (refer Figure 3).

Roughness model. We characterize the roughness of the random heightfield h using the *standard deviation* $\sigma_r := \sqrt{\text{var}(h(\vec{x}))}$ and *correlation length* Δ_r ($\text{corr}(h(\vec{x}), h(\vec{y})) \approx 0$ if $\|\vec{x} - \vec{y}\| > \Delta_r$). We consider 1. σ_r comparable to wavelength, 2. Δ_r smaller than the spatial coherence length Δ_i . The same conditions underlie standard BRDF models for optically rough surfaces [64–66]. Thus in practice, they are satisfied by any surface not optically smooth (i.e., near-specular) at visible wavelengths. The surface roughness conditions are also spelt out in Goodman [46, §5.10]

Scattering integral. For a single incident plane wave $\mathcal{G}(\hat{\omega}, k)e^{jk\hat{\omega}\cdot\mathbf{r}}$, the scattered wave reaching the camera sensor is given by

$$U_s(\vec{p}; k, \hat{\omega}) = \mathcal{G}(\hat{\omega}, k) \cdot \tilde{F}(\vec{p}; k(\hat{v} + \hat{\omega})) \quad (41)$$

where we have defined the *scattering integral* [46, 73]

$$\tilde{F}(\vec{p}; k, \vec{u}) = \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{\rho}) s(\vec{\rho}; k, u_z) e^{jk\vec{u}_{xy}\cdot\vec{\rho}} d\vec{\rho} \quad (42)$$

where $K_c(\cdot)$ is the *coherent spread function* (CSF) of the camera describing the complex field impulse response of the camera lens at the depth of the surface (assumed fronto-parallel to the camera), and $s(\vec{\rho}; k, u_z) = r(\vec{\rho}, k)e^{jk u_z h(\vec{\rho})}$ as introduced in Definition 3. The reflection coefficient $r(\vec{\rho}, k)$ is a slowly varying (compared to h) non-negative factor that models the spatially-varying spectral albedo of the surface. Note that $s(\cdot; k, u_z)$ is a random signal through its dependence on the heightfield $h(\cdot)$. This therefore results in the total field

$$U_s(\vec{p}) = \int_{\mathbb{R}^+} \int_{\mathcal{H}^2(\hat{n})} \mathcal{G}(\hat{\omega}, k) \cdot \tilde{F}(\vec{p}; k(\hat{v} + \hat{\omega})) d\sigma(\hat{\omega}) dk \quad (43)$$

Measured irradiance. The irradiance reaching the sensor is the average mean-squared of the field reaching the sensor, *i.e.*

$$E_s(\vec{p}) = \left\langle |U_s(\vec{p})|^2 \right\rangle \quad (44)$$

$$= \int_{k \in \mathbb{R}^+} \int_{\hat{\omega} \in \mathcal{H}^2(\hat{n})} \ell_i(k) L_i(\hat{\omega}) (\hat{\omega} \cdot \hat{n}) \left| \tilde{F}(\vec{p}; k, (\hat{v} + \hat{\omega})) \right|^2 d\sigma(\hat{\omega}) dk \quad (45)$$

following a simplification similar to the example of incident irradiance in Appendix E.1 based on the mutual incoherence assumption.

To obtain the irradiance *measured* by the camera sensor, we must weight the spectral integration by the sensor spectral sensitivity function $w_c(k)$:

$$E(\vec{p}) = \int_{\mathbb{R}^+} w_c(k) \ell_i(k) E(\vec{p}, k) dk \quad (46)$$

defining

$$E(\vec{p}, k) = \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) \left| \tilde{F}(\vec{p}; k, \hat{v} + \hat{\omega}) \right|^2 d\sigma(\hat{\omega}) \quad (47)$$

We therefore have the following proposition:

Proposition 9: Measured Sensor Irradiance (Exact)

$$E(\vec{p}) = \int_{\mathbb{R}^+} w_c(k) \ell_i(k) E(\vec{p}, k) dk \quad (48)$$

$$E(\vec{p}, k) = \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) \left| \tilde{F}(\vec{p}; k, \hat{v} + \hat{\omega}) \right|^2 d\sigma(\hat{\omega}) \quad (49)$$

E.4 Approximating measured irradiance

To arrive at Proposition 4 in the form of a simple linear system, we now make two simplifying approximations.

Approximation 1: Mean z-component approximation. We now make a standard implicit approximation used in similar derivations [65, 66]. Consider the scattering integral Equation (42). The phase term $s(\vec{p}; k u_z) = e^{j k u_z h(\vec{p})}$ varies with incident direction $\hat{\omega}$, since $\vec{u} = \hat{\omega} + \hat{v}$ in Equation (49), preventing an interpretation of $\tilde{F}$ as a Fourier transform.

If we consider an illumination environment map similar to Equation (38) where the incident illumination is relatively concentrated around a peak direction $\hat{\Omega}$ with z-component Ω_z , then the variation in the z-component of the incident direction ω_z across the environment map is small, *i.e.* $\omega_z \approx \Omega_z$ for all $\hat{\omega}$ with significant contribution. Formally, the approximation is stated as:

$$(h(\vec{p}_1) - h(\vec{p}_2)) \Delta_{\omega_z} \ll \lambda \quad \forall \vec{p}_1, \vec{p}_2 \quad (50)$$

This gives us the following approximate form for the scattering integral:

$$\tilde{F}(\vec{p}; k \vec{u}) \approx \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{\rho}) s(\vec{\rho}; k, \vec{u}_z) e^{j k \vec{u}_{xy} \cdot \vec{\rho}} d\vec{\rho} \quad (51)$$

using $s(\vec{\rho}; k, \vec{u}_z) = e^{j k \vec{u}_z h(\vec{\rho})}$ with the mean direction z-component $\vec{u}_z = \Omega_z + v_z$.

Using this approximate form, we expand the squared magnitude term in Equation (49) as

$$\left| \tilde{F}(\vec{p}; k \vec{u}) \right|^2 \approx \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{\rho}_1) K_c^*(\vec{p} - \vec{\rho}_2) s(\vec{\rho}_1; k, \bar{u}_z) s^*(\vec{\rho}_2; k, \bar{u}_z) e^{jk \vec{u}_{xy} \cdot (\vec{\rho}_1 - \vec{\rho}_2)} d(\vec{\rho}_1) d(\vec{\rho}_2) \quad (52)$$

Using a change of variables

$$\begin{aligned} \vec{q} &= \frac{\vec{\rho}_1 + \vec{\rho}_2}{2} \\ \vec{r} &= \frac{\vec{\rho}_1 - \vec{\rho}_2}{2} \end{aligned} \quad (53)$$

we have

$$\begin{aligned} \left| \tilde{F}(\vec{p}; k \vec{u}) \right|^2 &\approx 4 \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{q} - \vec{r}) K_c^*(\vec{p} - \vec{q} + \vec{r}) \\ &\quad s(\vec{q} + \vec{r}; k, \bar{u}_z) s^*(\vec{q} - \vec{r}; k, \bar{u}_z) \\ &\quad e^{jk \vec{u}_{xy} \cdot (2\vec{r})} d\vec{r} d\vec{q} \end{aligned} \quad (54)$$

Using $S(\vec{q}, \vec{r}; k, \bar{u}_z)$ from Definition 3,

$$S(\vec{q}, \vec{r}; k, \bar{u}_z) := s(\vec{q} + \vec{r}; k, \bar{u}_z) s^*(\vec{q} - \vec{r}; k, \bar{u}_z) \quad (55)$$

and plugging the above into Equation (49), we have an approximate expression

$$\begin{aligned} E(\vec{p}, k) &\approx 4 \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{q} - \vec{r}) K_c^*(\vec{p} - \vec{q} + \vec{r}) \\ &\quad S(\vec{q}, \vec{r}; k \bar{u}_z) \\ &\quad \left[\int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) e^{jk \hat{\omega}_{xy} \cdot (2\vec{r})} d\sigma(\hat{\omega}) \right] \\ &\quad e^{jk \hat{v}_{xy} \cdot (2\vec{r})} d\vec{r} d\vec{q} \end{aligned} \quad (56)$$

Identifying the term in square brackets as the coherence function $A_i(2\vec{r}; k)$ from Definition 8, we have

$$\begin{aligned} E(\vec{p}, k) &\approx 4 \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{q} - \vec{r}) K_c^*(\vec{p} - \vec{q} + \vec{r}) \\ &\quad S(\vec{q}, \vec{r}; k \bar{u}_z) A_i(2\vec{r}; k) e^{jk \hat{v}_{xy} \cdot (2\vec{r})} d\vec{r} d\vec{q} \end{aligned} \quad (57)$$

At this point, we bring to attention the definitions of L_o from Definition 3

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

and Definition 3

$$\mathbf{f}_r^h(\vec{q}, \vec{u}, k) := \frac{1}{A} \int_{\mathbb{R}^2} S(\vec{q}, \vec{r}, k, u_z) e^{jk \vec{u}_{xy} \cdot 2\vec{r}} d\vec{r}, \quad (\text{rep. 4})$$

where A , with dimensions of area, is a constant inversely proportional to the amount of light collection by the lens aperture, scaling as $A \propto N_{\text{eff}}^2/4m^2$ (the proportionality absorbs the area scale set by the CSF, which cancels in all downstream results). The effective (sensor-side) f-number $N_{\text{eff}} = N(1 + m)$, N being the f-number of the camera (refer Appendix A)

With the definition of the coherence function in Definition 8, and using the same mean-z-component approximation already used for the scattering integral $\tilde{\mathbf{F}}$, we can write $L_o(\vec{q}, \hat{v}, k)$ approximately as

$$L_o(\vec{q}, \hat{v}, k) \approx \frac{1}{A} \int_{\mathbb{R}^2} S(\vec{q}, \vec{r}; k, \bar{u}_z) \Lambda_i(2\vec{r}; k) e^{jk \hat{v}_{xy} \cdot (2\vec{r})} d\vec{r} \quad (58)$$

Approximation 2: Weak-coherence approximation. For typical ambient-lighting conditions, the coherence function (*e.g.* Equation (39)) will have small width $\Delta_i = 1/k\Delta_\omega$. As long as Δ_i is small relative to the width of the CSF $K_c(\cdot)$, we can approximate the CSF to be constant within the area of the support of Λ_i , resulting in the expression (from Equation (57) and Equation (58)):

$$E(\vec{p}, k) \approx \int_{\mathbb{R}^2} P_c(\vec{p} - \vec{q}) L_o(\vec{q}, \hat{v}, k) d\vec{q} \quad (59)$$

where the normalized PSF (integrates to unity) of the camera $P_c(\vec{\rho}) := N_{\text{eff}}^2/m^2 |K_c(\vec{\rho})|^2$.

Combining the above with Equation (48), we have

$$E(\vec{p}) \approx \int_{\mathbb{R}^+} w_c(k) \ell_i(k) \int_{\mathbb{R}^2} P_c(\vec{p} - \vec{q}) L_o(\vec{q}, \hat{v}, k) d\vec{q} dk \quad (60)$$

thus proving Proposition 4.

E.5 Ensemble averaging

The BRDF defined in Stam [65] can be equivalently stated using our notation as the (normalized) power spectral density of the scattering function s :

$$\bar{\mathbf{f}}_r(\vec{q}, \vec{u}, k) = \frac{1}{4A} \mathcal{P}_s(k \vec{u}_{xy}), \quad (61)$$

$$\mathcal{P}_s(\vec{\kappa}) := \lim_{\mathcal{A} \rightarrow \infty} \frac{1}{\mathcal{A}} \mathbb{E}_h \left[\left| \int_{\mathcal{A}} s(\vec{\rho}; k, u_z) e^{j\vec{\kappa} \cdot \vec{\rho}} d\vec{\rho} \right|^2 \right], \quad (62)$$

where $\mathcal{A}$ is the illuminated surface area and $\mathcal{P}_s$ is the power spectral density of s (the Fourier transform of its spatial autocorrelation, by the Wiener-Khinchin theorem). The $1/\mathcal{A}$ normalization makes the BRDF independent of the illuminated area (a per-unit-area surface property). For this section, we are considering a

surface where the BRDF is constant (as indicated by the expression). The BRDF is proportional to the expected value of the squared magnitude of the absolute-squared fourier transform of $s(\vec{\rho}; k, u_z)$ over an ensemble of surface heightfield realizations. It is *almost* the mean of the absolute-square of the scattering integral $\tilde{F}$ in Equation (42), except for the camera lens CSF term $K_c(\cdot)$.

The expression for irradiance in our model depends on the particular realization of the surface heightfield $h(\cdot)$ through the *sample BRDF* f_r^h . We first prove Proposition 5—that the ensemble mean of the sample BRDF is the traditional BRDF—and then use the exact model to show how the finite-aperture CSF relates the two.

Proof of Proposition 5. If we compute the mean of our sample BRDF $f_r^h(\vec{q}, \vec{u}, k)$ from Definition 3 over an ensemble of surface heightfield realizations, we have

$$\mathbb{E}_h[f_r^h(\vec{q}, \vec{u}, k)] = \frac{1}{A} \int_{\mathbb{R}^2} \mathbb{E}_h[S(\vec{q}, \vec{r}, k, u_z)] e^{jk\vec{u}_{xy} \cdot 2\vec{r}} d\vec{r} \quad (63)$$

$$= \frac{1}{A} \int_{\mathbb{R}^2} \mathbb{E}_h[s(\vec{q} + \vec{r}; k, u_z) s^*(\vec{q} - \vec{r}; k, u_z)] e^{jk\vec{u}_{xy} \cdot 2\vec{r}} d\vec{r} \quad (64)$$

This step requires only *wide-sense stationarity* of s (equivalently, of the heightfield h): the correlation $\mathbb{E}_h[s(\vec{q} + \vec{r}) s^*(\vec{q} - \vec{r})]$ is then the autocorrelation of s at separation $2\vec{r}$, independent of the center $\vec{q}$.¹⁰ Substituting $\vec{\rho} = 2\vec{r}$ (Jacobian $d\vec{r} = 1/4 d\vec{\rho}$),

$$\mathbb{E}_h[f_r^h(\vec{q}, \vec{u}, k)] = \frac{1}{4A} \int_{\mathbb{R}^2} \mathbb{E}_h[s(\vec{\rho}' + \vec{\rho}; k, u_z) s^*(\vec{\rho}'; k, u_z)] e^{jk\vec{u}_{xy} \cdot \vec{\rho}} d\vec{\rho} \quad (65)$$

$$= \frac{1}{4A} \mathcal{P}_s(k\vec{u}_{xy}), \quad (66)$$

identifying the Fourier transform of the autocorrelation with the power spectral density $\mathcal{P}_s$ (Wiener–Khinchin). Comparing with the definition of $\bar{f}_r$, we obtain (Proposition 5)

$$\bar{f}_r(\vec{q}, \hat{\omega} + \hat{v}, k) = \mathbb{E}_h[f_r^h(\vec{q}, \hat{\omega} + \hat{v}, k)]. \quad (67)$$

The traditional BRDF as the pinhole limit. If we consider the expected value of the irradiance in Equation (49) from the exact model (Proposition 9), we obtain

$$\mathbb{E}_h[E(\vec{p}, k)] = \int_{\mathcal{H}^2(\hat{n})} L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) \mathbb{E}_h \left[\left| \tilde{F}(\vec{p}; k, \hat{v} + \hat{\omega}) \right|^2 \right] d\sigma(\hat{\omega}) \quad (68)$$

$$\mathbb{E}_h \left[\left| \tilde{F}(\vec{p}; k, \hat{v} + \hat{\omega}) \right|^2 \right] = \mathbb{E}_h \left[\left| \int_{\mathbb{R}^2} K_c(\vec{p} - \vec{\rho}) s(\vec{\rho}; k, u_z) e^{jk\vec{u}_{xy} \cdot \vec{\rho}} d\vec{\rho} \right|^2 \right] \quad (69)$$

¹⁰ We adopt the stronger *ergodicity* assumption in Proposition 5, as it is the conventional condition invoked for optically rough surfaces [46]; it implies wide-sense stationarity, which is all the argument here uses.

If we consider the case of a pinhole camera, then $K_c(\cdot)$ has infinite extent and is constant across the surface, so it factors out of $\tilde{F}$, leaving the plain Fourier transform of s that defines the traditional BRDF:

$$\mathbb{E}_h \left[\left| \tilde{F}(\vec{p}; k, \hat{v} + \hat{\omega}) \right|^2 \right] \propto \mathbb{E}_h \left[\left| \int_{\mathcal{A}} s(\vec{p}; k, u_z) e^{jk \vec{u}_{xy} \cdot \vec{p}} d\vec{p} \right|^2 \right] \propto \mathcal{P}_s(k \vec{u}_{xy}) \propto \bar{f}_r(\vec{q}, \vec{u}, k), \quad (70)$$

recovering the traditional BRDF-based rendering equation. The two agree *only* in this pinhole limit: the traditional BRDF omits the camera’s coherent spread function K_c , whereas our exact model retains it. For any finite aperture the CSF remains inside $\bar{F}$ and our model departs from the traditional BRDF—precisely the regime in which defocus and second-order texture (subjective speckle) arise. Our model thus *generalizes* the traditional BRDF to finite-aperture imaging of unresolved microgeometry.

F Characterization of second-order texture

In this section, we use the appearance model derived in the previous section to prove Proposition 6 in the main text, characterizing the distribution of irradiance through κ_E^2 .

Equation (59) can be approximated as:

$$E(\vec{p}, k) \approx \sum_m E_m^{(\epsilon)}, \quad \text{where} \quad (71)$$

$$E_m^{(\epsilon)}(k) := \epsilon^2 P_c(\vec{p} - \vec{q}_m) \mathcal{L}^{(\epsilon)}(\vec{q}_m, k), \quad (72)$$

$$\mathcal{L}^{(\epsilon)}(\vec{q}, k) := \int_{\mathbb{R}^2} G_\epsilon(\vec{q}' - \vec{q}) L_o(\vec{q}', \hat{v}, k) d\vec{q}' \quad (73)$$

defining $\mathcal{L}^{(\epsilon)}(\cdot, k)$ as the (blurred) local average of outgoing radiance $L_o(\cdot, \hat{v}, k)$ using a normalized Gaussian kernel in $\mathbb{R}^2$ of width ϵ centered at $\vec{q}_m$

$$G_\epsilon(\vec{q}' - \vec{q}_m) := \frac{2}{\pi \epsilon^2} \exp \left(-2 \frac{\|\vec{q}' - \vec{q}_m\|^2}{\epsilon^2} \right), \quad (74)$$

with $\{\vec{q}_m\}$ forming a rectangular grid of spacing ϵ in $\mathbb{R}^2$. In the limit of $\epsilon \rightarrow 0$, Equation (71) becomes an exact equality with Equation (59). This approximation is suitable when ϵ is smaller than the width of the PSF Δ_c .

Our objective now is to model the distribution of $E(\vec{p}, k)$. To do this, we consider the approximation above using the specific grid spacing of $\epsilon = \Delta_i$. In Appendix F.1, we show that considering the Gaussian environment lighting model in Equations (38)–(39) with spatial coherence length Δ_i , $\mathcal{L}^{(\Delta_i)}(\vec{q}, k)$ can be modeled as an exponential random variable:

$$\mathcal{L}^{(\Delta_i)}(\vec{q}, k) \mid \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}, k) \sim \text{Expo}(1/\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}, k)), \quad (75)$$

with mean equal to its first-order (ensemble-averaged) counterpart, $\bar{\mathcal{L}}^{(\Delta_i)}(\cdot, k) = \mathbb{E}_h[\mathcal{L}^{(\Delta_i)}(\cdot, k)]$. More explicitly in terms of the (first-order) BRDF:

$$\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}, k) = \int_{\mathbb{R}^2} G_\epsilon(\vec{\rho} - \vec{q}) \int_{\mathcal{H}^2(\hat{n})} \mathbb{E}_h[f_r^h(\vec{\rho}, \hat{\omega}, \hat{v}, k)] L_i(\hat{\omega})(\hat{\omega} \cdot \hat{n}) d\sigma(\hat{\omega}) d\vec{\rho} \quad (76)$$

We can also consider $\mathcal{L}^{(\Delta_i)}(\vec{q}_i, k)$ and $\mathcal{L}^{(\Delta_i)}(\vec{q}_j, k)$ are independent for $i \neq j$ since they correspond to (mostly) non-overlapping microstructure regions on the surface (and the correlation length of the heightfield is smaller than Δ_i). We can now approximate Equation (6) as:

$$E(\vec{p}) \approx \sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \mathcal{L}^{(\Delta_i)}(\vec{q}_m), \quad \text{where} \quad (77)$$

$$\mathcal{L}^{(\Delta_i)}(\vec{q}) := \int_{\mathbb{R}^+} \mathcal{L}^{(\Delta_i)}(\vec{q}, k) w_c(k) \ell_i(k) dk \quad (78)$$

While $\mathcal{L}^{(\Delta_i)}(\vec{q}_m, k)$ are independent¹¹ across the grid locations $\vec{q}_m$, they are *not* independent across wavenumbers k for a fixed m , *i.e.* $\mathcal{L}^{(\Delta_i)}(\vec{q}_m, k_1)$ and $\mathcal{L}^{(\Delta_i)}(\vec{q}_m, k_2)$ are not independent as they are both directly dependent on the local surface structure in the neighborhood of $\vec{q}_m$. Considering two wavenumbers k_1, k_2 they may be correlated or uncorrelated depending on the distribution from which the surface heightfield h is sampled.

Consider that the normalized spectral envelope of the light reaching the camera

$$f_0(k) := \frac{w_c(k) \ell_i(k)}{\int_{\mathbb{R}^+} w_c(k') \ell_i(k') dk'} \quad (79)$$

has a gaussian form

$$f_0(k) = \frac{1}{\sqrt{2\pi} \Delta_k} \exp\left(-\frac{(k - \bar{k})^2}{2\Delta_k^2}\right) \quad (80)$$

with bandwidth Δ_k and mean wavenumber $\bar{k}$. Goodman [46] show that for a Gaussian-distributed heightfield, Gaussian correlation function, and a Gaussian spectral profile, the number of uncorrelated spectral buckets for a given wavenumber bandwidth Δ_k is given by

$$M = \sqrt{1 + 8\pi^2 \left(\frac{\Delta_k}{\bar{k}}\right)^2 \left(\frac{\sigma_r}{\lambda}\right)^2} \approx G_1 \Delta_k \quad (81)$$

where $G_1 = 2\sqrt{2}\pi(\sigma_r/\lambda\bar{k})$, thereby allowing us to approximately model the integral over wavenumber in Equation (78) as a sum of M uncorrelated exponential random variables.

¹¹ conditioned on their first order means, see Appendix F.2

The conditional distribution of $\mathcal{L}^{(\Delta_i)}(\vec{q})$ can therefore be approximated as a Gamma distribution with shape parameter M :

$$\mathcal{L}^{(\Delta_i)}(\vec{q}_m) \mid \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m) \sim \text{Gamma}\left(M, \frac{M}{\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)}\right). \quad (82)$$

where we similarly define

$$\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}) := \int_{\mathbb{R}^+} \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}, k) w_c(k) \ell_i(k) dk \quad (83)$$

Proposition 10: Model for second-order texture

The irradiance at point $\vec{p}$ can be approximately modeled as

$$\mathbb{E}(\vec{p}) \approx \sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \mathcal{L}^{(\Delta_i)}(\vec{q}_m) \quad (84)$$

where the (blurred) outgoing radiance at grid locations $\vec{q}_m$ with spacing Δ_i can be modeled as Gamma-distributed random variables

$$\mathcal{L}^{(\Delta_i)}(\vec{q}_m) \mid \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m) \sim \text{Gamma}\left(M, \frac{M}{\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)}\right). \quad (85)$$

We correspondingly define a first-order counterpart of the approximate irradiance

$$\bar{\mathbb{E}}(\vec{p}) \approx \sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m) \quad (86)$$

Using the above model, we show in Appendix F.2 that κ_E^2 can be written as

$$\kappa_E^2 := \frac{1}{MQ} \left(1 + \kappa_{\bar{\mathcal{L}}^{(\Delta_i)}}^2\right) + \kappa_{\bar{\mathbb{E}}}^2. \quad (87)$$

where the respective coefficients of variation for $\bar{\mathbb{E}}$ and $\bar{\mathcal{L}}^{(\Delta_i)}$

$$\kappa_{\bar{\mathbb{E}}}^2 := \frac{\mathbb{V}[\bar{\mathbb{E}}(\vec{p})]}{\mu_E^2}, \quad (88)$$

$$\kappa_{\bar{\mathcal{L}}^{(\Delta_i)}}^2 := \frac{\mathbb{V}[\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)]}{\mathbb{E}[\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)]^2}. \quad (89)$$

If we further assume that first-order texture is independent across the grid locations $\vec{q}_m$ (*i.e.* $\bar{\mathbb{E}}(\vec{q}_m)$ are independent across $\vec{q}_m$), then we can write $\kappa_{\bar{\mathbb{E}}}^2 =$

$\frac{1}{Q}\kappa_{\mathcal{L}(\Delta_i)}^2$, resulting in the simplified expression

$$\kappa_E^2 = \frac{1}{MQ} \left(1 + \kappa_{\mathcal{L}(\Delta_i)}^2 \right) + \frac{1}{Q} \kappa_{\mathcal{L}(\Delta_i)}^2. \quad (90)$$

$$= \frac{1}{MQ} (1 + Q\kappa_E^2) + \kappa_E^2 \quad (91)$$

thus proving Proposition 6.

F.1 Proof of exponential distribution

Plugging Equation (58) into Equation (73), we have

$$\begin{aligned} \mathcal{L}^{(\epsilon)}(\vec{q}', k) &= \frac{1}{A} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} G_\epsilon(\vec{q} - \vec{q}') S(\vec{q}, \vec{r}, k) \Lambda_i(2\vec{r}, k) e^{jk\hat{v}_{xy} \cdot (2\vec{r})} d\vec{r} d\vec{q}, \quad (92) \\ &= \frac{2}{\pi A \epsilon^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} e^{-2\frac{\|\vec{q} - \vec{q}'\|^2}{\epsilon^2}} S(\vec{q}, \vec{r}, k) \Lambda_i(2\vec{r}, k) e^{jk\hat{v}_{xy} \cdot (2\vec{r})} d\vec{r} d\vec{q} \end{aligned} \quad (93)$$

To proceed we consider the special case of the coherence function for Gaussian environment lighting (Equations (38)–(39)):

$$\Lambda_i(2\vec{r}, k) = E_i(k) \exp\left(-\frac{\|2\vec{r}\|^2}{2\Delta_i^2}\right) \exp\left(jk\hat{\Omega}_{xy} \cdot 2\vec{r}\right) \quad (94)$$

Using a change of variables $\vec{t}_1 = \vec{q} + \vec{r}$, $\vec{t}_2 = \vec{q} - \vec{r}$, we have

$$\begin{aligned} \mathcal{L}^{(\epsilon)}(\vec{q}', k) &= E_i(k) \frac{2e^{-\frac{2\|\vec{q}'\|^2}{\epsilon^2}}}{\pi A \epsilon^2} \cdot \\ &\quad \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} s(\vec{t}_1, k) e^{-\frac{\|\vec{t}_1\|^2}{\epsilon^2} \left(\frac{1}{2} + \frac{1}{\Delta_i^2}\right) + \frac{2\vec{q}' \cdot \vec{t}_1}{\epsilon^2}} e^{jk(\hat{v}_{xy} + \hat{\Omega}_{xy}) \cdot \vec{t}_1} \\ &\quad s^*(\vec{t}_2, k) e^{-\frac{\|\vec{t}_2\|^2}{\epsilon^2} \left(\frac{1}{2} + \frac{1}{\Delta_i^2}\right) + \frac{2\vec{q}' \cdot \vec{t}_2}{\epsilon^2}} e^{-jk(\hat{v}_{xy} + \hat{\Omega}_{xy}) \cdot \vec{t}_2} \\ &\quad \exp\left(-\vec{t}_1 \cdot \vec{t}_2 \left(\frac{1}{\epsilon^2} - \frac{1}{\Delta_i^2}\right)\right) d\vec{t}_1 d\vec{t}_2 \end{aligned} \quad (95)$$

The integrand above is separable in $\vec{t}_1$ and $\vec{t}_2$ except for the cross-term in the last exponential above. Setting $\epsilon = \Delta_i$, the cross-term in the exponential vanishes, and the above integral then simplifies to

$$\mathcal{L}^{(\Delta_i)}(\vec{q}', k) = \frac{2E_i(k)}{\pi A \Delta_i^2} \left| \int_{\mathbb{R}^2} s(\vec{t}, k) e^{-\frac{\|\vec{t} - \vec{q}'\|^2}{\Delta_i^2}} e^{jk(\hat{v}_{xy} + \hat{\Omega}_{xy}) \cdot \vec{t}} d\vec{t} \right|^2 \quad (96)$$

$$= E_i(k) \frac{2\pi \Delta_i^2}{A} \left| \int_{\mathbb{R}^2} G_{\sqrt{2}\Delta_i}(\vec{t} - \vec{q}') s(\vec{t}, k) e^{jk(\hat{v}_{xy} + \hat{\Omega}_{xy}) \cdot \vec{t}} d\vec{t} \right|^2 \quad (97)$$

Without ensemble averaging over the distribution of surface heightfields, the above corresponds to *speckle* with a Gaussian window of width $\sqrt{2}\Delta_i$ around $\vec{q}'$ on the surface. Since it is a magnitude-squared of a sum of a large number of independent complex random variables (considering correlation length of the height field is sufficiently small relative to Δ_i), we can argue (as in Goodman [46]) that the real and imaginary parts of the integral are therefore normally distributed random variables (by central limit theorem), thereby making the blurred spectral density of outgoing radiance $\mathcal{L}^{(\Delta_i)}(\vec{q}', k)$ exponentially distributed.

F.2 Spatial integration

From Equation (77) we have the conditional mean of irradiance given the first-order blurred outgoing radiance $\bar{\mathcal{L}}^{(\Delta_i)}(\vec{p})$ as:

$$\mathbb{E}[\mathbb{E}(\vec{p}) \mid \bar{\mathcal{L}}^{(\Delta_i)}] = \sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \mathbb{E}[\mathcal{L}^{(\Delta_i)}(\vec{q}_m) \mid \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)] \quad (98)$$

$$= \sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m) \quad (99)$$

$$\approx \int_{\mathbb{R}^2} P_c(\vec{p} - \vec{\rho}) \bar{\mathcal{L}}^{(\Delta_i)}(\vec{\rho}) d\vec{\rho} \quad (100)$$

And the marginal mean of irradiance as (using ergodicity):

$$\mathbb{E}[\mathbb{E}(\vec{p})] = \int_{\mathbb{R}^2} P_c(\vec{p} - \vec{\rho}) \mathbb{E}_h[\bar{\mathcal{L}}^{(\Delta_i)}(\vec{\rho})] d\vec{\rho}. \quad (101)$$

$$= \mu_{\mathcal{L}^{(\Delta_i)}} = \mu_E \quad (102)$$

since the PSF integrates to unity. The conditional variance

$$\mathbb{V}[\mathbb{E} \mid \bar{\mathcal{L}}^{(\Delta_i)}] = \mathbb{V}\left[\sum_m (\Delta_i^2 P_c(\vec{p} - \vec{q}_m)) \mathcal{L}^{(\Delta_i)}(\vec{q}_m)\right] \quad (103)$$

$$= \sum_m (\Delta_i^2 P_c(\vec{p} - \vec{q}_m))^2 \mathbb{V}[\mathcal{L}^{(\Delta_i)} \mid \bar{\mathcal{L}}^{(\Delta_i)}] \quad (104)$$

$$= \sum_m (\Delta_i^2 P_c(\vec{p} - \vec{q}_m))^2 \frac{(\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m))^2}{M} \quad (105)$$

using conditional independence of $\mathcal{L}^{(\Delta_i)}(\vec{q}_m) \mid \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)$ across grid locations. Therefore the marginal variance of irradiance is given by

$$\mathbb{V}[\mathbb{E}] = \mathbb{V}[\mathbb{E}[\mathbb{E} \mid \bar{\mathcal{L}}^{(\Delta_i)}]] + \mathbb{E}[\mathbb{V}[\mathbb{E} \mid \bar{\mathcal{L}}^{(\Delta_i)}]] \quad (106)$$

The first term

$$\mathbb{V}[\mathbb{E}[\mathbb{E} \mid \bar{\mathcal{L}}^{(\Delta_i)}]] = \mathbb{V}\left[\sum_m \Delta_i^2 P_c(\vec{p} - \vec{q}_m) \bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)\right] \quad (107)$$

$$= \mathbb{V}[\bar{\mathbb{E}}(\vec{p})] \quad (108)$$

Further simplification isn't possible without information about the correlation structure of first-order texture $\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m)$ across grid locations, which are not necessarily independent. We will therefore leave it as is. The second term

$$\mathbb{E}\left[\mathbb{V}\left[\mathbb{E} \mid \bar{\mathcal{L}}^{(\Delta_i)}\right]\right] = \mathbb{E}\left[\sum_m (\Delta_i^2 \text{P}_c(\vec{p} - \vec{q}_m))^2 \frac{(\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m))^2}{M}\right] \quad (109)$$

$$= \frac{1}{M} \sum_m (\Delta_i^2 \text{P}_c(\vec{p} - \vec{q}_m))^2 \mathbb{E}\left[(\bar{\mathcal{L}}^{(\Delta_i)}(\vec{q}_m))^2\right] \quad (110)$$

$$= \mathbb{E}\left[(\bar{\mathcal{L}}^{(\Delta_i)})^2\right] \frac{\Delta_i^2}{M} \sum_m \Delta_i^2 (\text{P}_c(\vec{p} - \vec{q}_m))^2 \quad (111)$$

$$\approx \mathbb{E}\left[(\bar{\mathcal{L}}^{(\Delta_i)})^2\right] \frac{\Delta_i^2}{M} \int_{\mathbb{R}^2} (\text{P}_c(\vec{p} - \vec{\rho}))^2 d\vec{\rho} \quad (112)$$

Taking the PSF to be a Gaussian of width Δ_c (with parameter $\Delta_c/2$) centered at $\vec{p}$,

$$\text{P}_c(\vec{p} - \vec{\rho}) = \frac{2}{\pi \Delta_c^2} \exp\left(-2 \frac{\|\vec{p} - \vec{\rho}\|^2}{\Delta_c^2}\right) \quad (113)$$

$$\int_{\mathbb{R}^2} (\text{P}_c(\vec{p} - \vec{\rho}))^2 d\vec{\rho} = \frac{1}{\pi \Delta_c^2} \quad (114)$$

and thus the pointwise marginal variance of irradiance

$$\mathbb{V}[\mathbb{E}(\vec{p})] = \frac{1}{M} \left(\frac{\Delta_i^2}{\pi \Delta_c^2}\right) \mathbb{E}\left[(\bar{\mathcal{L}}^{(\Delta_i)})^2\right] + \mathbb{V}[\mathbb{E}(\vec{p})] \quad (115)$$

$$= \frac{1}{M} \left(\frac{\Delta_i^2}{\pi \Delta_c^2}\right) (\mu_{\mathcal{L}^{(\Delta_i)}}^2 + \mathbb{V}[\bar{\mathcal{L}}^{(\Delta_i)}]) + \mathbb{V}[\mathbb{E}(\vec{p})] \quad (116)$$

Therefore the coefficient of variation of irradiance is given by

$$\kappa_E^2 = \frac{\mathbb{V}[\mathbb{E}(\vec{p})]}{\mathbb{E}[\mathbb{E}(\vec{p})]^2} \quad (117)$$

$$= \frac{1}{M} \left(\frac{\Delta_i^2}{\pi \Delta_c^2}\right) \left(1 + \frac{\mathbb{V}[\bar{\mathcal{L}}^{(\Delta_i)}]}{\mathbb{E}[\bar{\mathcal{L}}^{(\Delta_i)}]^2}\right) + \frac{\mathbb{V}[\mathbb{E}]}{\mathbb{E}[\mathbb{E}]^2} \quad (118)$$

$$= \frac{1}{MQ} (1 + \kappa_{\mathcal{L}^{(\Delta_i)}}^2) + \kappa_E^2 \quad (119)$$

since $\mu_E = \mathbb{E}[\mathbb{E}(\vec{p})] = \mathbb{E}[\bar{\mathcal{L}}^{(\Delta_i)}] = \mu_{\mathcal{L}^{(\Delta_i)}}$, and defining $Q = \frac{\pi \Delta_c^2}{\Delta_i^2}$ as the effective number of coherence areas contributing to the irradiance at a sensor point through the PSF.

G Noise model

The (noisy) intensity $\tilde{I}$ (in DN, Digital Numbers) recorded by the sensor given an incident flux Φ [W] is modeled as follows

$$\begin{aligned}
\tilde{I}(\vec{p}) &= \min(\lfloor 1/g(\min(P, P_{\max}) + Z_{\text{pre}}) + Z_{\text{post}} \rfloor, I_{\max}) \\
P \mid \Phi &\sim \text{Poisson}((\zeta\Phi + D)t) \\
Z_{\text{pre}} &\sim \mathcal{N}(0, \sigma_{\text{pre}}^2), \quad [\text{e}^-] \\
Z_{\text{post}}(\vec{p}) &\sim \mathcal{N}(0, \sigma_{\text{post}}^2), \quad [\text{DN}] \\
\zeta &= \frac{v}{K}, \quad [\text{photoe}^- \text{J}^{-1}] \\
K &= \frac{hc_0}{\lambda}, \quad [\text{J}]
\end{aligned} \tag{120}$$

where, $P_{\max}$ is the full-well capacity of the pixel (in electrons), $I_{\max}$ is the maximum ADC value (in DN), $1/g$ is the gain factor (in [DN/e⁻]) determined by the ISO setting, v is the quantum efficiency, D is the dark current (in [e⁻s⁻¹]), t is the exposure time (in [s]), Z_{pre} is the pre-amplifier read noise (in electrons), and Z_{post} is the post-amplifier read noise (in DN) separate from noise due to ADC quantization. Parameters of the above noise model used for our numerical studies for Figure 7 are specified in Table 3c.

Simplifying assumptions for analysis. For analytical tractability of (noisy) focus measure statistics, we make the following simplifying assumptions:

1. We operate in a regime where saturation effects (full-well and ADC) can be neglected.
2. We ignore quantization effects, *i.e.*, we consider $\tilde{I}$ to be a continuous random variable.
3. We consider dark current to be negligible, *i.e.*, $D \approx 0$.

Resulting in the simplified model:

$$\begin{aligned}
\tilde{I}(\vec{p}) &= 1/g(P + Z_{\text{pre}}) + Z_{\text{post}} \\
P \mid \Phi &\sim \text{Poisson}(\zeta\Phi t)
\end{aligned} \tag{121}$$

Under the above assumptions, we define the noise-free intensity I as the conditional mean of $\tilde{I}$ given the incident flux Φ :

$$I := \mathbb{E}[\tilde{I} \mid \Phi] = \frac{\zeta t}{g} \Phi \tag{122}$$

$$\mathbb{V}[\tilde{I} \mid \Phi] = \mathbb{V}[\tilde{I} \mid I] = \frac{1}{g^2} [\zeta t \Phi + \sigma_{\text{pre}}^2] + \sigma_{\text{post}}^2 \tag{123}$$

$$= \frac{I}{g} + \frac{\sigma_{\text{pre}}^2}{g^2} + \sigma_{\text{post}}^2 \tag{124}$$

We may rewrite Equation (121) in an additive form, defining $n := \tilde{I} - I$:

$$\begin{aligned}\tilde{I} &= I + n \\ \mathbb{V}[n \mid I] &= \mathbb{V}[\tilde{I} \mid I] = \frac{I}{g} + \frac{\sigma_{\text{pre}}^2}{g^2} + \sigma_{\text{post}}^2\end{aligned}\quad (125)$$

which is the model described in § 3 (excluding saturation).

4. As a final simplification, we assume the incident irradiance $E(\vec{p})$ [W m^{-2}] is uniform over the pixel area

$$\Phi = E(\vec{p})\Delta_p^2, \quad [\text{W}] \quad (126)$$

where Δ_p^2 is the pixel area (in $[\text{m}^2]$). This therefore implies the noise-free intensity is proportional to the incident irradiance as follows:

$$I = \frac{\beta t}{g} E(\vec{p}) \quad (127)$$

$$\beta := \frac{v\Delta_p^2}{K}, \quad [\text{photoe}^- \text{J}^{-1} \text{m}^2] \quad (128)$$

We may also write the conditional mean and variance of $\tilde{I}$ given E as:

$$\begin{aligned}\mathbb{E}[\tilde{I} \mid E] &= \frac{\beta t}{g} E \\ \mathbb{V}[\tilde{I} \mid E] &= \frac{1}{g^2} [\beta t E + \sigma_{\text{pre}}^2] + \sigma_{\text{post}}^2\end{aligned}\quad (129)$$

We do not make the above assumptions (other than proportionality of flux to irradiance, Item 4) in our numerical implementation Figure 7, which is implemented according to Equation (120) with practical values chosen for sensor parameters including full-well capacity, dark current *etc.*, that can be found in Table 3c. Further details of the numerical implementation can be found in Appendix D.3.

H Depth from focus analysis

As described in § 3, given pixel intensity measurements in a patch P_T of T pixels, the squared sample coefficient of variation focus measure

$$\tilde{c}_j^2(\vec{p}) := \frac{\tilde{s}_j^2(\vec{p})}{\tilde{m}_j(\vec{p})}, \quad (130)$$

is computed from the *sample variance* and *sample mean* of patch intensities,

$$\tilde{m}_j(\vec{p}) := \frac{1}{T} \sum_{\vec{p}' \in P_T(\vec{p})} \tilde{I}_j(\vec{p}'), \quad (131)$$

$$\tilde{s}_j^2(\vec{p}) := \frac{1}{T-1} \sum_{\vec{p}' \in P_T(\vec{p})} \left(\tilde{I}_j(\vec{p}') - \tilde{m}_j(\vec{p}) \right)^2 \quad (132)$$

The noise-free focus measure c^2 is defined similarly :

$$c_j^2(\vec{p}) := \frac{s_j^2(\vec{p})}{m_j(\vec{p})}, \quad (133)$$

$$s_j^2(\vec{p}) := \frac{1}{T-1} \sum_{\vec{p}' \in P_T(\vec{p})} (I_j(\vec{p}') - m_j(\vec{p}))^2 \quad (134)$$

$$m_j(\vec{p}) := \frac{1}{T} \sum_{\vec{p}' \in P_T(\vec{p})} I_j(\vec{p}'), \quad (135)$$

For the purposes of analysis, we also define a normalized version of the sample variance

$$\tilde{C}^2(\vec{p}) := \frac{\tilde{s}^2(\vec{p})}{\mu_I^2} \quad (136)$$

$$C^2(\vec{p}) := \frac{s^2(\vec{p})}{\mu_I^2} \quad (137)$$

where $\mu_I = \mathbb{E}[I] = \mathbb{E}[\tilde{I}]$ is the marginal mean of intensities over the patch. This normalized sample variance is distinct from the actual focus measure defined above (and not directly measurable exactly). For analytical tractability, we study the statistics of $\tilde{C}^2$ instead of $\tilde{c}^2$, to approximately characterize $\tilde{c}^2$.

H.1 Statistics of the focus measure

Given the simplified sensor noise model in Equation (125), we can denote the pointwise marginal mean and variance of measured (noisy) pixel intensity $\tilde{I}$ as

$$\mu_{\tilde{I}} := \mathbb{E}[\tilde{I}] = \mu_I \quad (138)$$

$$\sigma_{\tilde{I}}^2 := \mathbb{V}[\tilde{I}] = \sigma_I^2 + \frac{\mu_I}{g} + \frac{\sigma_{\text{pre}}^2}{g^2} + \sigma_{\text{post}}^2 \quad (139)$$

where the marginal mean and variance of true intensities (as a consequence of assuming uniform irradiance over the pixel area, Equation (127)) are given by

$$\mu_I := \mathbb{E}[I] = \frac{\beta t}{g} \mu_E \quad (140)$$

$$\sigma_I^2 := \mathbb{V}[I] = \frac{\beta^2 t^2}{g^2} \sigma_E^2 \quad (141)$$

and as a consequence of this assumption,

$$\kappa_I^2 := \frac{\sigma_I^2}{\mu_I^2} = \frac{\sigma_E^2}{\mu_E^2} = \kappa_E^2 \quad (142)$$

We can now write the coefficient of variation of noisy intensity as

$$\kappa_{\tilde{\mathbf{I}}}^2 := \frac{\sigma_{\tilde{\mathbf{I}}}^2}{\mu_{\tilde{\mathbf{I}}}^2} = \kappa_{\mathbf{I}}^2 + \kappa_n^2 \quad (143)$$

$$= \kappa_{\mathbf{E}}^2 + \kappa_n^2 \quad (144)$$

where we defined in Equation (2) the squared reciprocal of pixel intensity SNR as the coefficient of variation of noise κ_n^2 :

$$\kappa_n^2 := \frac{\sigma_n^2}{\mu_{\mathbf{I}}^2} = \frac{1}{t\zeta\mu_{\Phi}} \left(1 + \frac{\sigma_{\text{read}}^2}{t\zeta\mu_{\Phi}} \right) \quad (\text{rep. 2})$$

with the marginal variance of noise n ,

$$\sigma_n^2 := \mathbb{V}[n] = \mathbb{E}[\mathbb{V}[n \mid \mathbf{I}]] + \mathbb{V}[0] = \frac{\mu_{\mathbf{I}}}{g} + \frac{\sigma_{\text{pre}}^2}{g^2} + \sigma_{\text{post}}^2 \quad (145)$$

using the conditional variance from Equation (125). The pointwise marginal distribution of pixel intensity $\tilde{\mathbf{I}}$ at each pixel is identical across pixels. We make the additional assumption that these pixel intensities are *independent across pixels*.¹² Therefore the sample variance of noisy pixel intensities has marginal statistics:

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{s}}^2] &= \mathbb{V}[\tilde{\mathbf{I}}] = \sigma_{\tilde{\mathbf{I}}}^2 \\ \mathbb{V}[\tilde{\mathbf{s}}^2] &= \frac{(\sigma_{\tilde{\mathbf{I}}}^2)^2}{T} \left(\kappa_4 - \frac{T-3}{T-1} \right) \end{aligned}$$

where κ_4 is the kurtosis (ratio of fourth moment to squared-variance) of the marginal distribution of pixel intensities. If the distribution of pixel intensities is approximated to be to Gaussian, then $\kappa_4 \approx 3$.

Normalizing by $\mu_{\tilde{\mathbf{I}}}^2$ we then have the marginal statistics of $\tilde{\mathbf{C}}^2$:

¹² This assumption is technically strictly valid only when the f -number of the camera is sufficiently small, such that the width of the PSF is smaller than the pixel size (see Appendix A.1). If the PSF width is too large, then adjacent pixels will have correlated intensities since they see overlapping regions of the surface. We use this assumption to simplify the analysis of focus measure statistics.

Lemma 11: Marginal statistics of $\tilde{C}^2$

$$\mathbb{E}[\tilde{C}^2] = \frac{\sigma_{\tilde{I}}^2}{\mu_{\tilde{I}}^2} = \kappa_{\tilde{I}}^2 = \kappa_I^2 + \kappa_n^2 \quad (146)$$

$$\mathbb{V}[\tilde{C}^2] = \frac{(\kappa_{\tilde{I}}^2)^2}{T} \left(\kappa_4 - \frac{T-3}{T-1} \right) \quad (147)$$

$$= \frac{2(\kappa_{\tilde{I}}^2)^2}{T-1} \quad (\text{if } \kappa_4 \approx 3, \text{ effectively } \tilde{I} \sim \mathcal{N}(\mu_I, \sigma_n^2)) \quad (148)$$

Statistics of focus measure for first-order textureless surface. For a first-order textureless surface, we have $\kappa_I^2 = \kappa_E^2 = 1/QM$, and so we have

$$\begin{aligned} \mathbb{E}[\tilde{C}^2] &= \frac{1}{QM} + \kappa_n^2 \\ \mathbb{V}[\tilde{C}^2] &= \frac{2}{T-1} \left\{ \frac{1}{QM} + \kappa_n^2 \right\}^2 \end{aligned}$$

H.2 Proof of Proposition 2

Appendix H.1 establishes the mean and variance of $\tilde{C}^2$. We now present a proof for Proposition 2, characterizing the α -recoverability of a surface from depth from focus, as defined in Definition 1.

We indicate the in-focus and out-of-focus normalized sample variances as $\tilde{C}_{\text{if}}^2$ and $\tilde{C}_{\text{oof}}^2$ respectively, and the corresponding noise-free focus measures as C_{if}^2 and C_{oof}^2 .

1. $\tilde{C}_{\text{if}}^2$ is the value of $\tilde{C}^2$ evaluated at the sharpest focus setting of the camera.
2. $\tilde{C}_{\text{oof}}^2$ is the value of $\tilde{C}^2$ evaluated at a focus setting of the camera that is sufficiently far from the in-focus setting, such that the true noise-free intensities in the patch are effectively constant (*i.e.*, the texture is completely blurred out), or $I(\vec{p}) \approx \mu_I$. This therefore implies $C_{\text{oof}}^2 = 0$.

To proceed, we model $\tilde{C}_{\text{if}}^2$ and $\tilde{C}_{\text{oof}}^2$ in and out of focus to be normally-distributed random variables, with the mean and variance expressions derived in Appendix H.1. The conclusions of this analysis are supported by our Monte-Carlo simulations in Figure 7, thereby justifying the use of this assumption for the purpose of analytical tractability.

Modeling $\tilde{C}_{\text{if}}^2$ and $\tilde{C}_{\text{oof}}^2$ to be normally-distributed, with the mean and variance derived in Lemma 11.

$$\tilde{C}_{\text{if}}^2 \sim \mathcal{N}\left(\kappa_I^2 + \kappa_n^2, \frac{2}{T-1}(\kappa_I^2 + \kappa_n^2)^2\right) \quad (149)$$

$$\tilde{C}_{\text{oof}}^2 \sim \mathcal{N}\left(\kappa_n^2, \frac{2}{T-1}(\kappa_n^2)^2\right) \quad (150)$$

In addition, since the out-of-focus intensity variations in $\tilde{C}_{\text{oof}}^2$ are purely due to sensor noise, we can say that $\tilde{C}_{\text{if}}^2$ and $\tilde{C}_{\text{oof}}^2$ are uncorrelated random variables. Therefore,

$$\Pr(\tilde{C}_{\text{if}}^2 > \tilde{C}_{\text{oof}}^2) = \Phi\left(\frac{\mathbb{E}[\tilde{C}_{\text{if}}^2] - \mathbb{E}[\tilde{C}_{\text{oof}}^2]}{\sqrt{\mathbb{V}[\tilde{C}_{\text{if}}^2] + \mathbb{V}[\tilde{C}_{\text{oof}}^2]}}\right) \quad (151)$$

$$= \Phi\left(\frac{\kappa_1^2}{\sqrt{\frac{2}{T-1}\{(\kappa_1^2 + \kappa_n^2)^2 + (\kappa_n^2)^2\}}}\right) \quad (152)$$

$$= \Phi\left(\frac{\Theta^2}{\sqrt{\frac{2}{T-1}\{(\Theta^2 + 1)^2 + 1\}}}\right) \quad (153)$$

where Φ is the CDF of the standard normal distribution, and we've defined Θ^2 in § 5.

$$\Theta^2 := \frac{\kappa_1^2}{\kappa_n^2} \quad (154)$$

Lemma 12: Probability of Error

Modeling $\tilde{C}_{\text{if}}^2$ and $\tilde{C}_{\text{oof}}^2$ to be normally-distributed, with the mean and variance derived in Lemma 11, the *probability of error*

$$\Pr(\tilde{C}_{\text{if}}^2 < \tilde{C}_{\text{oof}}^2) = 1 - \Phi\left(\frac{\Theta^2}{\sqrt{\frac{2}{T-1}\{(\Theta^2 + 1)^2 + 1\}}}\right) \quad (155)$$

Therefore we have a necessary and sufficient condition for α -recoverability (Definition 1) of a surface using depth from focus:

$$\frac{\Theta^2}{\sqrt{1 + (\Theta^2 + 1)^2}} \cdot \sqrt{\frac{T-1}{2}} > \Phi^{-1}(1 - \alpha) \quad (156)$$

thus proving Proposition 2.

Notice that as Θ is increased, the LHS saturates to $\sqrt{\frac{T-1}{2}}$, which is the maximum possible value of the argument of the CDF. Therefore, for a fixed value of T , there is a maximum achievable probability of correct recovery, which is given by $\Phi\left(\sqrt{\frac{T-1}{2}}\right)$.

I Recoverability with second-order texture

In this section, we derive Proposition 7, which characterizes the contrast-SNR product Θ as a function of exposure time t and bandwidth Δ_k , for a first-order textureless surface. For a first-order textureless surface, we have $\kappa_{\text{E}}^2 = 0$, and so from Proposition 6

$$\kappa_{\text{I}}^2 = \frac{1}{QM} \quad (157)$$

with $M \approx G_1 \Delta_k$ (Equation (81)). We can similarly expand κ_n^2 using the noise model in Equation (2).

$$\kappa_n^2 = \frac{1}{t\zeta\mu_\Phi} \left(1 + \frac{\sigma_{\text{read}}^2}{t\zeta\mu_\Phi} \right) \quad (158)$$

Spectral filtering reduces the mean flux μ_Φ by a factor of γ_g (described in Appendix I.1), which is a function of the filter bandwidth Δ_k .

$$\gamma_g := \frac{\mu_\Phi^1}{\mu_\Phi^0} \approx g_p \frac{\Delta_k}{\Delta_{\text{src}}} \quad (159)$$

where μ_Φ^0 and μ_Φ^1 are the mean flux before and after spectral filtering respectively, and g_p is the peak attenuation of the spectral filter. For sufficiently narrow spectral filter bandwidth, the effective bandwidth after filtering is approximately the same as that of the filter, and so we abuse notation and use Δ_k for both the spectral filter bandwidth as well as the effective bandwidth after filtering. We denote the fixed bandwidth of the original (unfiltered) source spectrum by Δ_{src} (written Δ_k in Equation (80)). These are clarified in Appendix I.1.

We define

$$G_2 := \frac{\zeta\mu_\Phi^0 \cdot g_p}{\Delta_{\text{src}}} \quad (160)$$

and plugging in $\zeta\mu_\Phi^1 = G_2\Delta_k$ into the expression for κ_n^2 , we have the contrast-SNR product for a first-order textureless surface as a function of exposure time and (filter) bandwidth

$$\Theta^2(t, \Delta_k) \approx \frac{1}{Q} \cdot \frac{G_2}{G_1} \cdot \frac{t}{\left(1 + \frac{\sigma_{\text{read}}^2}{G_2} \frac{1}{\Delta_k t} \right)} \quad (161)$$

Since bandwidth affects μ_Φ^1 , the upper limit of exposure time until saturation is bandwidth-dependent. A simple notion to model this is to limit the exposure time until $\mu_{\text{I}} = \zeta t \frac{\mu_\Phi^1}{g} = \frac{G_2 \Delta_k}{g} t$ equals I_{max} .

$$t \in \left[0, \frac{1}{\Delta_k} \cdot \frac{g \text{I}_{\text{max}}}{G_2} \right) \quad (162)$$

thus resulting in Proposition 7.

I.1 Attenuation due to spectral filtering

Consider the original normalized spectral profile (before spectral filtering) is denoted as $f_0(k)$, as in Equation (79). Let the attenuation profile of the spectral filter introduced be $g(k)$. Thus the effective attenuation in irradiance due to the spectral filter is given by the factor

$$\gamma_g := \frac{\mu_\Phi^1}{\mu_\Phi^0} = \int_{\mathbb{R}^+} g(k) f_0(k) dk \quad (163)$$

I.2 Gaussian spectral profile

The spectral profile of various light sources and object reflectance are wide and varied. However, to build intuition for the effect of spectral mismatch between the spectral filter and the illumination and reflectance spectrum, we consider a Gaussian profile for both $f_0(k)$ and $g(k)$.

We have previously introduced a Gaussian form for f_0 in Equation (80)

$$f_0(k) = \frac{1}{\sqrt{2\pi}\Delta_{\text{src}}} \exp\left(-\frac{(k - \bar{k}_{f_0})^2}{2\Delta_{\text{src}}^2}\right) \quad (164)$$

The spectral filter has a peak attenuation g_p ,

$$g(k) = g_p \exp\left(-\frac{(k - \bar{k}_g)^2}{2\Delta_g^2}\right) \quad (165)$$

The effective spectral profile after introducing the spectral filter will also maintain a Gaussian form,

$$f_0(k)g(k) = \gamma_g \cdot \left[\frac{1}{\sqrt{2\pi}\Delta_{\text{eff}}} \exp\left(-\frac{(k - \bar{k}_{\text{eff}})^2}{2\Delta_{\text{eff}}^2}\right) \right] \quad (166)$$

with effective mean wavenumber and bandwidth given by

$$\bar{k}_{\text{eff}} = \bar{k}_{f_0} \frac{\Delta_g^2}{\Delta_g^2 + \Delta_{\text{src}}^2} + \bar{k}_g \frac{\Delta_{\text{src}}^2}{\Delta_g^2 + \Delta_{\text{src}}^2} \quad (167)$$

$$\frac{1}{\Delta_{\text{eff}}^2} = \frac{1}{\Delta_{\text{src}}^2} + \frac{1}{\Delta_g^2} \quad (168)$$

and effective attenuation factor

$$\gamma_g = g_p \frac{\Delta_g}{\sqrt{\Delta_{\text{src}}^2 + \Delta_g^2}} \exp\left(-\frac{(\bar{k}_{f_0} - \bar{k}_g)^2}{2(\Delta_{\text{src}}^2 + \Delta_g^2)}\right). \quad (169)$$

implying a quadratic-exponential attenuation as the filter center frequency $\bar{k}_g$ moves away from the original center frequency $\bar{k}_{f_0}$.

Narrowband limit. If we consider a spectral filter with a bandwidth $\Delta_g \ll \Delta_{\text{src}}$, then we can approximate the effective attenuation factor as

$$\gamma_g \approx g_p \frac{\Delta_g}{\Delta_{\text{src}}} \exp\left(-\frac{(\bar{k}_{f_0} - \bar{k}_g)^2}{2\Delta_{\text{src}}^2}\right) \quad (170)$$

which is linear in the filter bandwidth Δ_g . Further, the effective bandwidth after filtering Δ_{eff} is approximately the same as the filter bandwidth Δ_g in this case. If we additionally consider that the width of the original spectrum Δ_{src} is sufficiently wide compared to the spectral mismatch $(\bar{k}_{f_0} - \bar{k}_g)$, then we can further approximate the attenuation factor as

$$\gamma_g \approx g_p \frac{\Delta_g}{\Delta_{\text{src}}} \approx g_p \frac{\Delta_{\text{eff}}}{\Delta_{\text{src}}} \quad (171)$$

I.3 Wavenumber and wavelength bandwidths

In our results, we specify bandwidths in terms of *wavelength* bandwidths as Δ_λ (specified in nm) because this is the common convention for specification of spectral filters as is more familiar to readers. However, we need to convert this to wavenumber bandwidth Δ_k for our analysis and monte-carlo numerical study (Equation (81)). Considering a central wavelength $\bar{\lambda}$, we consider the mean (angular) wavenumber $\bar{k} = 2\pi/\bar{\lambda}$.

$$k_{\text{low}} = \frac{2\pi}{\bar{\lambda} + \frac{\Delta_\lambda}{2}} \quad (172)$$

$$k_{\text{high}} = \frac{2\pi}{\bar{\lambda} - \frac{\Delta_\lambda}{2}} \quad (173)$$

$$\Delta_k = k_{\text{high}} - k_{\text{low}} \quad (174)$$

$$= 2\pi \cdot \frac{\Delta_\lambda}{\bar{\lambda}^2 - \frac{\Delta_\lambda^2}{4}} \quad (175)$$

Under the condition that $\Delta_\lambda \ll \bar{\lambda}$,

$$\Delta_k \approx \frac{2\pi}{\bar{\lambda}} \cdot \frac{\Delta_\lambda}{\bar{\lambda}} = \bar{k} \cdot \frac{\Delta_\lambda}{\bar{\lambda}} \quad (176)$$