Reimplementing Flash/No-Flash Photography: Joint Bilateral Detail Transfer on Captured Low-Light Scenes

Asya Ulger
Dartmouth College

Abstract

Photographing a dimly lit scene forces a compromise: no-flash images preserve the scene's true color and atmosphere but are noisy, while flash images are sharp but ruin the ambient lighting. We reimplement the flash/no-flash framework of Petschnigg et al. from scratch and evaluate it on scenes we captured to stress-test individual stages. Beyond the core pipeline, we add the paper's continuous-flash interpolation and a guided-filter detail layer that removes the bilateral's gradient-reversal halos. We also find that the paper's literal shadow test fails under our radiometric setup, and adopt a luminance-based reformulation.

1. Introduction

A central goal of photography is to reproduce the lighting of an environment, which in dim settings carries the mood of the image but is hard to capture cleanly: long exposures blur, wide apertures lose depth of field, and high ISO trades brightness for noise. Flash avoids all three, yielding a clean, well-lit image. However, it flattens the lighting, cools the colors, and adds cast shadows and specular highlights. Petschnigg et al. [1] observe that the two exposures are instead complementary: the flash image F locates the scene's edges and high-frequency detail far better than the noisy ambient image A. Even though it carries the wrong illumination, one can use F to guide an edge-preserving denoise of A, transfer F's detail as a multiplicative layer, and fall back to a safer filter where the flash misleads.

This project reimplements that framework on scenes we captured ourselves. Rather than introduce a new algorithm, we reproduce the method faithfully, test it on cases chosen to stress its individual stages, and report honestly where it succeeds and where it does not. We make the following contributions:

- A complete, stage-by-stage implementation of Sections 3 and 4 of the paper, each stage validated on captured data.
- An implementation of the continuous-flash interpolation of Section 6 as an interactive tool that adjusts flash intensity after capture.
- A guided-filter detail layer that removes the gradient-reversal halos the bilateral filter produces near strong edges, addressing a limitation the paper itself notes.
- A practical finding: the paper's literal shadow test does not transfer to our radiometric setup, and a luminance-based reformulation that restores sensible mask coverage.

2. Related Work

Flash/no-flash fusion. We reimplement Petschnigg et al. [1], who recover an ambient-looking image by using the flash frame to guide a cross-bilateral denoise of the noisy ambient frame and then transferring flash detail as a multiplicative ratio. We adopt this formulation because its pipeline decomposes into stages that can be validated independently. The concurrent work of Eisemann and Durand [2], published in the same SIGGRAPH volume, instead decomposes both frames into large-scale and detail layers with a bilateral filter and recombines the ambient large scale with the flash detail. Although this produces a cleaner relighting abstraction, its single fused step is harder to dissect for a stage-by-stage evaluation.

Edge-preserving filtering. Both methods rest on the bilateral filter of Tomasi and Manduchi [3], which preserves edges by weighting neighbors with a joint spatial–range kernel but has no notion of a separate guidance signal. This is the gap the cross-bilateral variant fills. He et al. [4] later showed that the bilateral filter produces gradient-reversal halos in detail-enhancement settings and proposed the guided filter, whose local linear model avoids them; this is the failure we observe in our own detail layer and the basis for the extension in Section 3.8.

3. Method

The implementation is written in Python, using NumPy and SciPy for the core array operations, OpenCV for fast filtering, rawpy for RAW decoding, and scikit-image for color-space conversions. Following the structure of the original paper, the system is organized as a sequence of stages — linearization, basic and joint bilateral filtering, detail transfer, masking, and compositing. Each stage outputs a diagnostic figure that we use to confirm its behavior on real data before proceeding. We denote the ambient (no-flash) frame A and the flash frame F throughout.

3.1. Acquisition

Each scene is captured with a tripod-mounted Canon EOS Rebel T7 (2000D) and its 18–55 mm kit lens, recording RAW at a fixed focal length and aperture with manual focus, so the only intended difference between the two frames is the flash. The ambient frame A is deliberately underexposed by two to three stops: because the T7's sensor is newer than the camera used in the original paper, a well-exposed ambient frame is not noisy enough to test denoising, so we underexpose to elicit sensor noise and recover brightness later with a matching digital gain (Section 3.2). The same underexpose-then-gain operation arises in handheld low-light capture, where a short anti-shake shutter forces underexposure—though that setting also adds motion blur and misalignment our tripod protocol avoids and the pipeline does not address. The flash frame F is shot with the camera's flash at ISO 200 and 1/125 s, at the same aperture as the corresponding ambient frame, so the difference $(F-A)$ reads as the light the flash added rather than a change in framing or depth of field. Table 1 lists the three scenes, each chosen to stress a different stage of the pipeline.

	Hero	Textile	Shadow/Spec.
Subject	candlelit	woven rug	glossy cup
ISO (A)	1600	1600	1600
Aperture	f/5.6	f/5.6	f/8
Shutter (A)	1.0 s	1.30 s	0.50 s
Underexposure	2 stops	3 stops	3 stops
Digital push	4×	8×	8×

Table 1. Capture settings. Each flash frame uses ISO 200, 1/125 s, at the same aperture as its ambient frame.

3.2. Linearization and Exposure Matching

Both frames are decoded from RAW using a linear tone curve (`gamma=(1,1)`) and no auto-brightening, so the output is proportional to scene radiance. We read a single set of white-balance multipliers from the flash frame and apply them to both images, so each color channel is scaled identically; otherwise the per-channel difference $(F-A)$ would be meaningless. We then place the ambient frame in the flash frame's radiometric space using the paper's exposure-matching relation,

$$A_{\text{Lin}} = A'_{\text{Lin}} \cdot \frac{\text{ISO}_F \, \Delta t_F}{\text{ISO}_A \, \Delta t_A} \tag{1}$$

where Δt is exposure time and the ratio is on the order of 10^{-3}. In this space, $F_{\text{Lin}} - A_{\text{Lin}}$ isolates the light the flash adds. For the denoising and detail-transfer path we then apply the digital gain of Section 3.1 (2^k for a k-stop underexposure), restoring display brightness while amplifying the sensor noise the pipeline is meant to remove. The shadow mask (Section 3.6) instead uses the un-pushed matched ambient, since the gain would otherwise corrupt the comparison $F_{\text{Lin}} - A_{\text{Lin}}$.

3.3. Basic Bilateral Filter

The denoising stages build on the bilateral filter of Tomasi and Manduchi [3] (paper Eqs. 2–3), an edge-preserving smoother that replaces each pixel with a weighted average of its neighbors,

$$A_p^{\text{Base}} = \frac{1}{k(p)} \sum_{p' \in \Omega} g_{\sigma_d}\big(\|p' - p\|\big)\, g_{\sigma_r}\big(A_{p'} - A_p\big)\, A_{p'} \tag{2}$$

where Ω is a spatial window around p, g_{σ_d} and g_{σ_r} are zero-mean Gaussians with standard deviations σ_d and σ_r, and $k(p)$ normalizes by the sum of the same weights. The spatial term favors nearby pixels, while the range term g_{σ_r} down-weights neighbors whose intensity differs considerably from A_p, so the filter averages within a region but not across its edges. We implement Eq. 2 in NumPy and verify it against OpenCV's *cv2.bilateralFilter*, which we use for full-resolution runs ($\sigma_d = 32$ px, $\sigma_r = 0.08$). Applied to the ambient frame, it yields our safe fallback A^{Base}: noise is removed, but without external guidance fine texture is smoothed away. This is the failure the joint bilateral filter (Sec. 3.4) is built to avoid. The same filter applied to the flash frame gives the base layer F^{Base} used in Sec. 3.5.

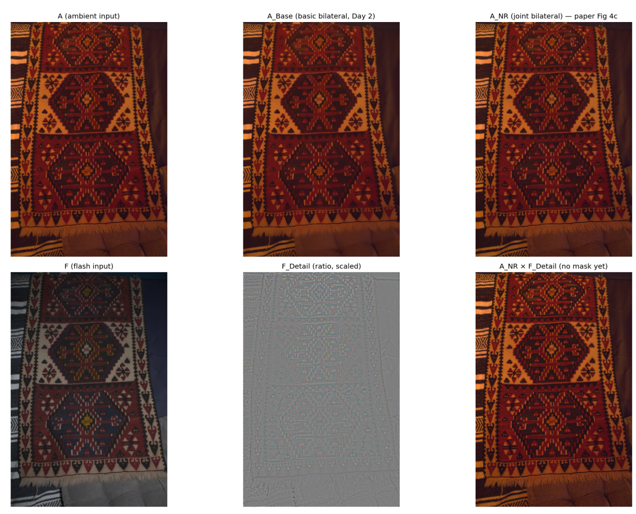

Figure 1. Pipeline stages on the textile scene. Top: ambient A; basic bilateral A^{Base} (weave smoothed away); joint bilateral A_{NR} (weave preserved). Bottom: flash F; detail layer F_{Detail}; detail-transferred result before masking.

3.4. Joint Bilateral Filter

The joint (cross) bilateral filter (paper Eq. 4) makes a single change to Eq. 2: the range Gaussian is evaluated on the flash frame F rather than on A, while the average is still taken over the ambient pixels $A_{p'}$,

$$A_p^{\mathrm{NR}} = \frac{1}{k(p)} \sum_{p' \in \Omega} g_{\sigma_d}\big(\|p' - p\|\big)\, g_{\sigma_r}\big(F_{p'} - F_p\big)\, A_{p'} \quad (3)$$

Because F is far less noisy than A, its edges are a reliable guide, so the range parameter can be shrunk to $\sigma_r = 0.001$ (0.1% of the intensity range, following the paper) without averaging across true edges. The result A^{NR} preserves the fine weave that the basic filter smoothed away while still suppressing ambient noise (Fig. 1, A^{NR} panel). This is the core denoising step: the flash supplies the edge information the noisy ambient frame cannot.

3.5. Detail Transfer

The flash frame carries high-frequency surface detail that the denoised ambient frame does not have. We extract it (paper Eq. 6) as the ratio of the flash frame to its own bilateral base layer,

$$F_{\mathrm{Detail}} = \frac{F + \varepsilon}{F^{\mathrm{Base}} + \varepsilon}, \qquad \varepsilon = 0.02 \quad (4)$$

where F^{Base} is the smoothed flash frame from Sec. 3.3. The ratio is near 1 on flat regions and departs from 1 where F has detail its base layer removed, so F_{Detail} is an image of fine texture alone (Fig. 1), and ε keeps it stable in dark regions. The product $A^{\mathrm{NR}} \cdot F_{\mathrm{Detail}}$ reintroduces the flash's sharpness while keeping the ambient color and tone (Fig. 1, final panel); this is the result before the shadow/specular mask of Sec. 3.6.

3.6. Shadow/Specular Mask and Composite

Detail transfer assumes the flash frame is a trustworthy source of detail, but this fails in two places. First is flash shadows, where the flash added no light and F carries no real detail. The second is specular highlights, where F is saturated. In both, the detail layer injects artifacts instead of texture. Following the paper (Eq. 7), we composite the detail-transfer result with the safe fallback A^{Base} through a mask M that is 1 in these unreliable regions,

$$A_{\mathrm{Final}} = (1 - M)\left(A^{\mathrm{NR}} \cdot F_{\mathrm{Detail}}\right) + M\, A^{\mathrm{Base}} \quad (5)$$

The paper detects shadows with an absolute test on the linearized flash contribution, $F^{\mathrm{lin}} - A^{\mathrm{lin}} < \tau$ (Eq. 8). We found this does not transfer to our captures: because exposure matching (Sec. 3.2) scales A^{lin} down by a factor on the order of 10^{-3}, the difference $F^{\mathrm{lin}} - A^{\mathrm{lin}}$ reduces to roughly F^{lin} everywhere, so a fixed threshold simply flags every dim pixel. On our metallic-cup scene this absolute test marked 87% of the frame as shadow. We instead use a *relative* test. A true flash shadow is a region where the flash contributed little compared to where it landed normally, so we normalize the contribution by a per-scene scale τ (a high percentile of $F^{\mathrm{lin}} - A^{\mathrm{lin}}$) and threshold the normalized value, additionally requiring that the ambient frame is not near-black so genuinely dark regions are not masked,

$$M_{\mathrm{shadow}} = \left[\tfrac{F^{\mathrm{lin}} - A^{\mathrm{lin}}}{\tau} < \beta\right] \,\wedge\, \left[A > \alpha_{\mathrm{floor}}\right], \quad (6)$$

$$M_{\mathrm{spec}} = \left[F^{\mathrm{lin}} > \tau_{\mathrm{spec}}\right] \quad (7)$$

with τ_{spec} a high percentile of the linear flash luminance. The two masks are combined, cleaned with a morphological opening, dilated slightly, and Gaussian-blurred so the composite transitions smoothly. The relative formulation generalizes across scenes where the absolute test does not, though its thresholds still require light per-scene adjustment (Sec. 5).

3.7. Continuous Flash

The pipeline so far produces a single fused image, but the same pair of frames also supports a continuous adjustment between them (paper Sec. 6). We convert both frames to YCbCr and interpolate with a single parameter α,

$$Y = (1-\alpha)\, Y_A + \alpha\, Y_F, \qquad C = \mathrm{clip}\big((1-\alpha)\, C_A + \alpha\, C_F\big) \quad (8)$$

where Y is luminance and C the two chrominance channels. At $\alpha = 0$ the result is the ambient frame and at $\alpha = 1$ the flash frame, with values outside $[0, 1]$ extrapolating to less-than-ambient or beyond-flash exposures. We let luminance extrapolate freely but clip the chrominance to the range spanned by the two sources, preventing out-of-gamut color shifts while still letting brightness extend past either input. The result is an interactive slider over flash intensity, set after capture rather than fixed at acquisition.

3.8. Guided Filter Extension

The detail layer of Sec. 3.5 inherits a known weakness of the bilateral filter: near strong edges it can produce *gradient-reversal halos*, where the recombined detail overshoots and reverses the gradient, leaving a bright or dark ring along the edge [4]. We observe these halos in F_{Detail} on our scenes. As an extension beyond the original paper, we replace the bilateral filter in the denoising and base-layer steps with the guided filter of He et al. [4], which assumes the output q is a local linear transform of a guidance image I within a window ω_k,

$$q_i = a_k I_i + b_k, \quad \forall i \in \omega_k \quad (9)$$

with (a_k, b_k) fit by regularized least squares against the input. Because $\nabla q = a_k \nabla I$ within each window, the output has an edge only where the guide does, avoiding the gradient reversals the bilateral introduces. Using the flash frame as the guide I ($r = 8$, $\varepsilon = 0.04$) yields a halo-free detail layer at comparable cost.

4. Evaluation

We evaluate on three scenes, each stressing a different stage (Table 1): the **textile** rug for denoising and detail transfer, the candlelit **hero** scene for color preservation under a strong warm/cool difference, and the **shadow/specular** cup for the mask.

Qualitative results. Across all three scenes, A_{Final} keeps the warm ambient color of the no-flash frame with the sharpness of the flash frame. The effect is clearest on the textile (Fig. 1), where the weave smeared by A^{Base} is preserved by A^{NR} and sharpened by detail transfer.

Quantitative noise. On a flat patch of the candlelit scene, the joint bilateral filter lowers the high-frequency noise standard deviation from 0.0216 to 0.0153, while the basic filter, with no detail to protect, smooths further to 0.0055 (Fig. 2). The joint filter deliberately trades flat-region smoothing to retain genuine texture.

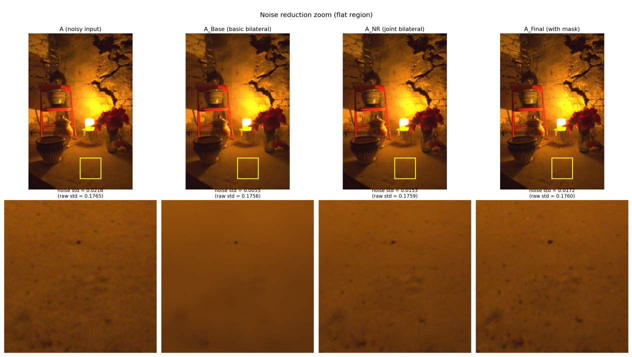

Figure 2. Noise reduction on a flat patch of the candlelit scene. The basic bilateral smooths most aggressively, while the joint bilateral preserves more detail.

Shadow/specular mask. The reformulated mask (Sec. 3.6) localizes to cast shadows and specular glints, leaving flat surfaces untouched; the amplified difference in Fig. 3 confirms it acts only at edges, flowers, and highlights. It generalizes where the paper's absolute test does not (Sec. 5).

Guided-filter detail layer. The guided-filter detail layer (Sec. 3.8) removes the bilateral's gradient-reversal halos; their difference concentrates along strong edges.

Continuous flash. YCbCr interpolation (Sec. 3.7) yields a smooth, post-capture sweep from below-ambient through full-flash to a synthetic beyond-flash exposure.

Figure 3. Shadow/specular mask on the candlelit scene. Top: A, F, and mask M (red) on F. Bottom: result without mask, final composite A_{Final} (Eq. 5), and the amplified difference.

Figure 4. Shadow/specular scene. Left: result without the mask. Center: final composite with the mask. Right: amplified difference, showing the mask localizes to the cup's specular streaks and the wall shadow, though reverting the specular object to the bilateral fallback dulls its highlights.

5. Limitations and Conclusion

We reimplemented the flash/no-flash framework of Petschnigg et al. and reproduced its central result (an image with the no-flash frame's lighting and the flash frame's clarity) while adding the paper's continuous-flash interpolation and a guided-filter detail layer that removes the bilateral's gradient-reversal halos. Several limitations remain. Our reformulated mask generalizes better than the paper's absolute test but still needs light per-scene tuning, and on a strongly specular scene reverting the object to the bilateral fallback dulls its highlights. Our underexposure-and-push protocol elicits sensor noise on a tripod but applies a large digital gain, and we rely on tripod capture for registration, so small inter-frame motion is misread as a flash artifact.

References

[1] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph. (SIGGRAPH)*, 23(3):664–672, 2004.

[2] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM Trans. Graph. (SIGGRAPH)*, 23(3):673–678, 2004.

[3] C. Tomasi and R. Manduchi. Bilateral filtering for gray and

color images. In *ICCV*, pages 839–846, 1998.

[4] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV*, pages 1–14, 2010.