Lensless Aerial Navigation in Dark

Deepak Singh*[1], Hudson Kortus*[1], Jahnavi Prudhivi*[1], Vivek Reddy Kasireddy*[1], Nitin J Sanket[1]
[1]*Perception and Autonomous Robotics (PeAR) Group*
Worcester Polytechnic Institute
Worcester, USA
{dsingh1, hkortus, jprudhivi, vkasireddy1, nitin}@wpi.edu

***Abstract*—Palm-sized aerial robots are promising for search and rescue operations, but limited sensing and onboard compute make autonomous flight challenging, especially in dark scenes. We present a lensless navigation system that replaces the conventional lens with an ultralight piece of tape, which passively suppresses background structure while preserving nearby obstacles, thereby embedding depth cues directly in the measurements. Combined with low-power infrared (IR) illumination, this enables autonomous navigation in complete darkness within the robot's sensing and compute constraints. To further reduce onboard computation, scene reconstruction is performed only when an obstacle is detected from raw measurements, and the recovered structure is then used for dodging. Our reconstruction pipeline runs at 14 Hz on a Jetson Orin™ Nano, and fewer than 30% of frames are reconstructed during flight, improving both computational efficiency and power usage. Real world autonomous flight experiments demonstrate robust traversal in cluttered environments in complete darkness.**

***Index Terms*—Lensless Imaging, Aerial Robots, Passive Computation.**

I. Introduction

Search and rescue often requires navigating tight, cluttered spaces that are unsafe for humans and often completely dark. Palm-sized aerial robots are well-suited to these settings, but reliable autonomy remains difficult under strict size, weight, and power constraints.

Vision is a primary sensing modality for such platforms due to its compact form factor, low cost, and rich scene information. However, safe navigation still requires recovering scene structure, particularly depth, for obstacle avoidance. Conventional robotic cameras do not encode depth cues directly in their measurements, and therefore rely on multiple views and computationally intensive processing to estimate depth [1]–[3]. Although effective, these approaches introduce significant hardware and computational overhead, which limits their suitability for small, resource-constrained aerial robots. In dark environments, the challenge becomes even greater, as active illumination and additional computation are often required, further increasing the burden on onboard resources. Optical approaches such as depth from defocus [4] and coded aperture imaging [5], [6] can embed depth cues within the captured signal, but they still depend on conventional lens-based designs and thus retain their associated size and weight constraints.

This motivates a fundamental shift in perspective: rather than recovering depth through increasingly complex post-processing or using multiple sensors, we can instead design the sensing process itself to encode scene structure. By adopting a *physics-informed input* approach, the image formation process can be shaped so that captured measurements inherently contain depth relevant information, thereby reducing the need for heavy downstream computation.

*Equal Contribution. Author order decided randomly.

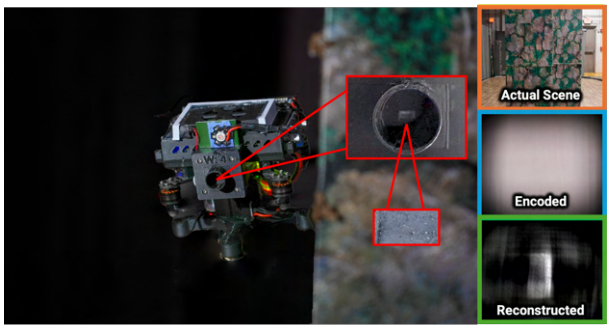

Fig. 1. **Our palm-sized aerial robot uses a lensless camera to navigate cluttered real-world scenes in complete darkness**. Replacing a conventional lens with a simple tape diffuser passively attenuates background structure, causing nearby obstacles to dominate the reconstruction and thereby providing a physics-informed signal for navigation. We illustrate the imaging process for a scene containing a foreground box: actual scene, raw image from the lensless camera, and reconstructed frames.

Recent work in computational imaging has explored lensless systems as lightweight alternatives to conventional cameras, using coded apertures, diffusers, and phase masks to directly encode scene structure into measurements [7]–[11]. These approaches reduce hardware complexity and enable compact imaging designs, but have largely been demonstrated in controlled indoor settings, often on objects displayed on screens and placed only a few centimeters from the camera.

In this work, we take lensless imaging beyond reconstruction in controlled settings and for the first time ever, demonstrate its potential as a viable sensing modality for real time aerial robot navigation, that too in complete darkness. We show that a lensless system can be deployed on a small aerial robot to enable autonomous flight in complete darkness (Fig. 1). At the core of our approach is an extremely simple optical element: a piece of translucent tape acting as a diffuser, inspired by [8], that reshapes the incoming light. Rather than reconstructing the entire scene, the diffuser's inability to preserve fine detail becomes an advantage, suppressing distant and irrelevant background information and effectively restricting perception to a navigation relevant depth range while still encoding ordinal depth cues through passive computation [3], [12].

To operate in complete darkness, we pair this with low-power infrared illumination (0.5W), enabling active sensing with minimal impact on system size, weight, and power. Because the captured measurements already encode coarse scene structure, we avoid expensive high-fidelity reconstruction and instead use low-resolution grayscale outputs from a *single* iteration of Alternating Direction Method of Multipliers (ADMM) optimizer for reconstruction. This enables real-time onboard perception on a resource-constrained platform,

Fig. 2. **Weight comparison between the proposed tape diffuser and conventional imaging optics.** The tape diffuser (right) weighs only 5 g, compared to several hundred grams for DSLR lenses (left) and tens of grams for compact robotic camera lenses. This corresponds to up to 73× lower weight than DSLR lenses and about 8× lower weight than small robotic lenses, highlighting its advantage for resource-constrained aerial robots.

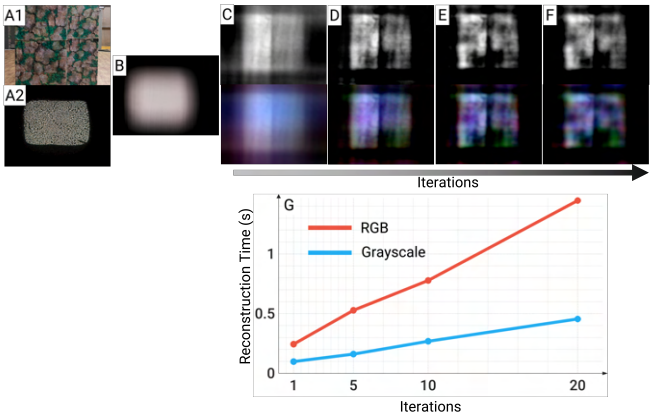

Fig. 3. **Grayscale vs. RGB lensless reconstruction.** (A1) Ground-truth RGB scene, (A2) Tape Point Spread Function (PSF) and (B) corresponding raw lensless measurement. (C)-(F) Reconstructions with increasing ADMM iterations, with grayscale on top and RGB on bottom. A single iteration already recovers the dominant scene structure, while additional iterations provide only minor gains. RGB reconstruction, however, incurs significantly higher runtime, as shown in (G), with limited added structural benefit.

allowing navigation in cluttered environments in complete darkness while remaining significantly lighter than traditional camera systems (Fig. 2). To the best of our knowledge, this is the first demonstration of lensless imaging for embodied robot navigation in unstructured real-world environments, highlighting a shift toward simple optics and task aware sensing over complex hardware and computation.

Our key contributions are summarized as follows:

- We present, to the best of our knowledge, the first lensless imaging-based navigation framework for robots in complete darkness ($< 1\text{milliLux}$).
- We demonstrate autonomous navigation in unstructured environments.
- We propose a reconstruction on demand strategy to reduce computation and improve efficiency.

II. Lensless Navigation Using Passive Computation

In our setup, we use a piece of translucent tape and a 0.5W IR source for illuminating the scene. Let $\mathbf{x} \in \mathbb{R}^N$ denote the scene irradiance and $\mathbf{y} \in \mathbb{R}^M$ the captured sensor measurement. The lensless image formation process is modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ is the diffuser-induced Point Spread Function (PSF) and $\mathbf{n}$ denotes measurement noise. We recover the scene by solving the TV-regularized inverse problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \tau\|\nabla\mathbf{x}\|_1, \tag{2}$$

using an ADMM-based solver [8]. Since our goal is obstacle avoidance rather than high-fidelity imaging, we reconstruct only coarse scene structure in grayscale. Unlike conventional lensless pipelines that solve separately for RGB channels, we recover a single intensity image, reducing the problem dimensionality by a factor of three and correspondingly lowering memory and computation cost (Fig. 3). Further, because the system must operate onboard a resource-constrained robot, we perform only a single reconstruction iteration. As shown in Fig. 3, additional iterations increase runtime, while the navigation task does not require fine-grained image detail.

A. Foreground Emphasis

The central idea of our approach is to emphasize nearby obstacles while suppressing less relevant background structure. This foreground-biased sensing behavior can be understood from the forward model. Let $\mathbf{x}_d$ denote the scene contribution at depth d. The captured measurement can then be written as

$$\mathbf{y} = \sum_{d\in\mathcal{D}} \alpha(d, P)\, \mathbf{H}_d\mathbf{x}_d + \mathbf{n}, \tag{3}$$

where $\mathbf{H}_d$ denotes the depth-dependent forward operator (PSF), and $\alpha(d, P)$ models attenuation as a function of scene depth d and illumination power P.

Under active illumination, the received signal strength follows

$$\alpha(d, P) = \kappa\, \frac{P}{d^2}, \tag{4}$$

where κ captures reflectance and diffuser-dependent effects.

The depth-dependent signal-to-noise ratio can be expressed as

$$\mathrm{SNR}(d) = \frac{\left\|\kappa\, \frac{P}{d^2}\, \mathbf{H}_d\mathbf{x}_d\right\|_2^2}{\|\mathbf{n}\|_2^2}, \tag{5}$$

which decreases quadratically with depth. Therefore, nearby obstacles produce measurements for which

$$\left\|\kappa\, \frac{P}{d^2}\, \mathbf{H}_d\mathbf{x}_d\right\|_2 \gg \|\mathbf{n}\|_2, \tag{6}$$

and are reliably recovered, whereas distant structures satisfy

$$\left\|\kappa\, \frac{P}{d^2}\, \mathbf{H}_d\mathbf{x}_d\right\|_2 \approx \|\mathbf{n}\|_2 \quad \text{or} \quad \left\|\kappa\, \frac{P}{d^2}\, \mathbf{H}_d\mathbf{x}_d\right\|_2 < \|\mathbf{n}\|_2, \tag{7}$$

causing them to blend into the noise floor and be poorly reconstructed. This depth-dependent attenuation naturally suppresses background regions while preserving foreground obstacles, yielding an implicit foreground-background separation in the recovered structure.

For navigation, we only require this foreground-background separation (or depth ordinality). To quantify this, we define a foreground emphasis metric $\mathcal{F}$ over the reconstructed image $\hat{\mathbf{x}}$. Let F and B denote foreground and background regions, corresponding to obstacle depths Z_f and Z_b, respectively. We define

$$\mathcal{F} = \frac{\mu_f}{\mu_b}, \tag{8}$$

where

$$\mu_f = \mathbb{E}[\hat{\mathbf{x}}(u) \mid u \in F], \qquad \mu_b = \mathbb{E}[\hat{\mathbf{x}}(u) \mid u \in B]. \tag{9}$$

Since the measurement contribution scales as $1/d^2$, the reconstructed foreground and background responses diverge as their depths separate, and hence

$$\mathcal{F} \propto |Z_b - Z_f|. \tag{10}$$

When the foreground and background are well separated in depth, $\mu_f > \mu_b$, yielding a large $\mathcal{F}$ and clear obstacle delineation. Conversely, when Z_f and Z_b are similar, their attenuated responses become comparable, so that $\mu_f \approx \mu_b$ and $\mathcal{F} \approx 1$, indicating weak foreground-background separation.

B. Obstacle-Aware Reconstruction Triggering

To reduce onboard computation, we avoid continuous reconstruction and trigger it only when the raw measurement deviates from an obstacle-free reference $I_0(x, y)$. Given an incoming frame $I_t(x, y)$, we first compute a smoothed image

$$I_t^s(x, y) = (I_t * G)(x, y), \tag{11}$$

where G is a Gaussian kernel. From I_t^s, we extract a compact set of statistics that capture both global intensity variation and localized bright structure. In particular, we compute the bright-pixel ratio

$$r_t = \frac{1}{N} \sum_{x,y} \mathbf{1}\big(I_t^s(x, y) > T_b\big), \tag{12}$$

the mean intensity μ_t, the total energy

$$E_t = \sum_{x,y} I_t^s(x, y), \tag{13}$$

the maximum intensity $I_{\max,t}$, and the 99th percentile intensity $I_{99,t}$.

These quantities are compared against their corresponding reference values computed from I_0. We then form binary indicators $C_1, \ldots, C_4$ that test whether the current frame exhibits a sufficient increase in bright-pixel ratio, mean intensity, total energy, maximum response, or 99th percentile response. The indicators are aggregated using

$$V_t = C_1 + C_2 + C_3 + C_4, \tag{14}$$

and reconstruction is triggered when $V_t \geq 2$. Since I_0 captures the nominal free-space response of the system, deviations in these statistics provide an efficient cue for obstacle presence directly from the raw lensless measurement, without requiring reconstruction at every frame.

C. Navigation Policy

Our navigation policy follows the formulation of [13]. In our setup, distant structures appear darker in the reconstructed image $\hat{\mathbf{x}}$ due to attenuation, making image intensity a useful cue for obstacle proximity. Thresholding $\hat{\mathbf{x}}$ separates nearby obstacles from background structure, and the resulting foreground map is used for obstacle avoidance.

III. Experiments and Results

We validate the proposed system through real-world autonomous flight experiments in indoor zero-light settings. All experiments are conducted in a netted flight arena of size $11 \times 4.5 \times 3.65$ m under near-zero ambient illumination (< 1 mLux). The robot is a custom-built quadrotor (Corgi140, 140 mm wheelbase) equipped with a Jetson Orin Nano for onboard perception and planning, a MatekSys H743 Mini running ArduPilot for low-level control, a forward-facing lensless camera with a tape diffuser placed 3 mm from the sensor, a 0.5 W IR illumination source, and a downward optical flow sensor for hover stabilization.

We consider two obstacle configurations (Table I): `Play tunnels` (Fig. 4(A)), consisting of upright fabric tunnels in a moderately cluttered layout, and `Poles` (Fig. 4(B)), consisting of thin PVC poles arranged vertically and horizontally, creating a sparser but geometrically challenging scene. For both scenes, we additionally compute traversability using the method of [14].

We also perform a benchtop study to quantify foreground-background separation. The scene consists of two box shaped obstacles with moss textures, where the depth gap between the foreground and background is varied and the resulting foreground emphasis metric $\mathcal{F}$ is measured. We compare ADMM reconstruction and Wiener deconvolution on the same Jetson Orin Nano platform. As shown in Fig. 5, ADMM consistently produces higher $\mathcal{F}$ values, indicating stronger separation between near and far structures, although it runs at 14 Hz compared to 80 Hz for Wiener deconvolution. We therefore adopt ADMM for autonomous traversal, as stronger obstacle separability is more critical for safe navigation, and reconstruction is performed only when needed rather than on every frame, keeping the overall computational load manageable.

For autonomous evaluation, we report *Success Rate* (SR), the number of collision-free traversals out of 15 runs, and *Trigger Rate* (TR), the percentage of frames requiring reconstruction. As shown in Table I, the robot achieves 13/15 successful runs in both scenes while reconstructing only 30.5% of frames in `Play tunnels` and 16.6% in `Poles`. These results show that robust autonomous navigation is possible while reconstructing only a small fraction of frames, substantially reducing onboard computation and improving power efficiency.

TABLE I
AUTONOMOUS NAVIGATION PERFORMANCE IN ZERO-LIGHT INDOOR ENVIRONMENTS. TR DENOTES TRIGGER RATE AND SR DENOTES SUCCESS RATE.

Scene	Traversability	Luminosity (lux)	TR (%)	SR (%)
`Play tunnels`	9.42	0.001	30.5	86.67 (13/15)
`Poles`	9.75	0.001	16.6	86.67 (13/15)

IV. Conclusion and Future Work

We presented the first demonstration of lensless sensing for autonomous aerial robot navigation, showing that a simple tape-based diffuser can enable real-world flight in complete darkness on a resource-constrained platform. By combining passive optical foreground emphasis with selective reconstruction, the proposed system achieves robust obstacle avoidance while reconstructing only a small fraction of frames, thereby reducing onboard computational load and improving power efficiency. Future work will investigate this paradigm in more

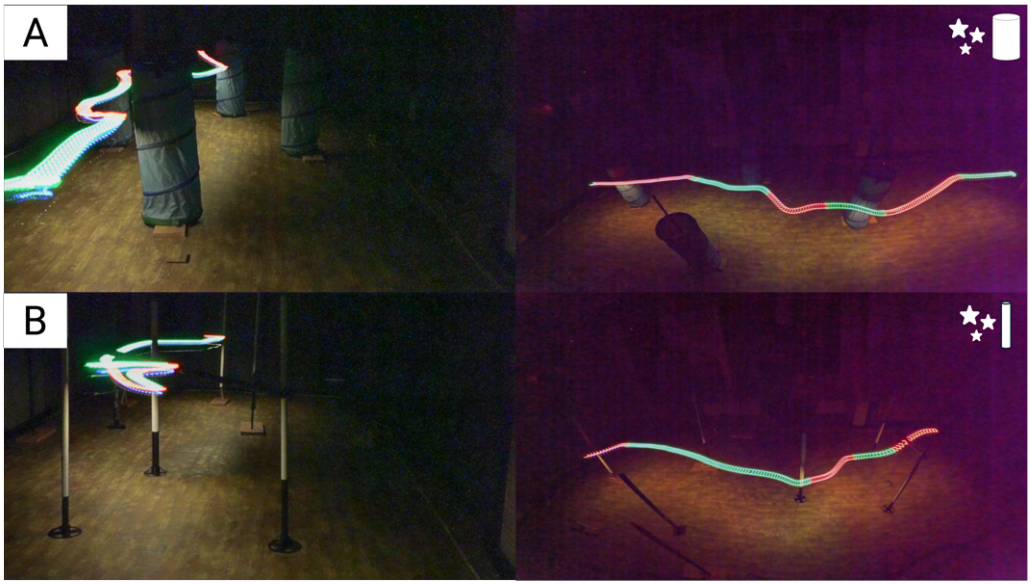

Fig. 4. **Real-world autonomous navigation in indoor zero-light scenes.** We evaluate the proposed system in two scenes: `Play tunnels` (A) and `Poles` (B). For each scene, the image on the left is taken from a SonyA7S3 Camera with a shutter speed of $1/100$, $f/5.6$ aperture, ISO 4096000. The right ones are top views recorded from a GoPro camera. In the images, red shows when obstacles are detected (thus reconstruction triggered) and green shows no obstacle detection.

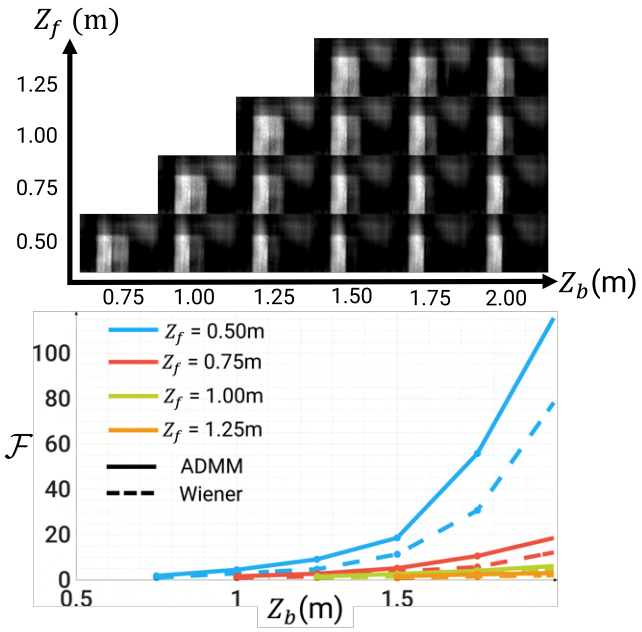

Fig. 5. **Foreground-background separation under varying depth gaps.** Two moss-textured box obstacles are placed at different depths, and the resulting foreground emphasis metric $\mathcal{F}$ is measured for ADMM and Wiener deconvolution. The top figure shows how the reconstruction varies with depth and the bottom plot shows how $\mathcal{F}$ varies for ADMM and Wiener.

demanding regimes, including high-speed navigation and dynamic obstacle avoidance, to better understand the limits and broader applicability of lensless sensing for agile autonomous flight.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002. [Online]. Available: https://doi.org/10.1023/A:1014573219977

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge: Cambridge University Press, 2004.

[3] D. Singh, S. Khobragade, and N. J. Sanket, "Asternav: Autonomous aerial robot navigation in darkness using passive computation," *IEEE Robotics and Automation Letters*, vol. 11, no. 3, pp. 3907–3914, 2026.

[4] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 523–531, 1987.

[5] A. Levin *et al.*, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70–es, Jul. 2007.

[6] H. Ikoma *et al.*, "Depth from defocus with learned optics for imaging and occlusion-aware depth estimation," in *2021 IEEE International Conference on Computational Photography (ICCP)*, 2021, pp. 1–12.

[7] L. A. Kabuli, H. Pinkard, E. Markley, C. S. Hung, and L. Waller, "Designing lensless imaging systems to maximize information capture," *Optica*, vol. 13, no. 2, pp. 227–235, Feb 2026. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-13-2-227

[8] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, Jan 2018. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-5-1-1

[9] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, Feb 1978. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-17-3-337

[10] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2017.

[11] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express*, vol. 27, no. 20, pp. 28 075–28 090, Sep 2019.

[12] H. Pawar, D. Singh, and N. J. Sanket, "Blurring for clarity: Passive computation for defocus-driven parsimonious navigation using a monocular event camera," in *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops*, February 2025, pp. 912–916.

[13] N. J. Sanket, C. D. Singh, C. Fermüller, and Y. Aloimonos, "Ajna: Generalized deep uncertainty for minimal perception on parsimonious robots," *Science Robotics*, vol. 8, no. 81, p. eadd5139, 2023. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.add5139

[14] C. Nous, R. Meertens, C. De Wagter, and G. de Croon, "Performance evaluation in obstacle avoidance," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 3614–3619.