Hi-SPAD: Video-Rate Hyperspectral Imaging and Inference with Single-Photon Cameras

HAEJOON LEE, Carnegie Mellon University, USA
MOHIT GUPTA, University of Wisconsin-Madison, USA
VIJAYAKUMAR BHAGAVATULA, Carnegie Mellon University, USA
ASWIN C. SANKARANARAYANAN, Carnegie Mellon University, USA

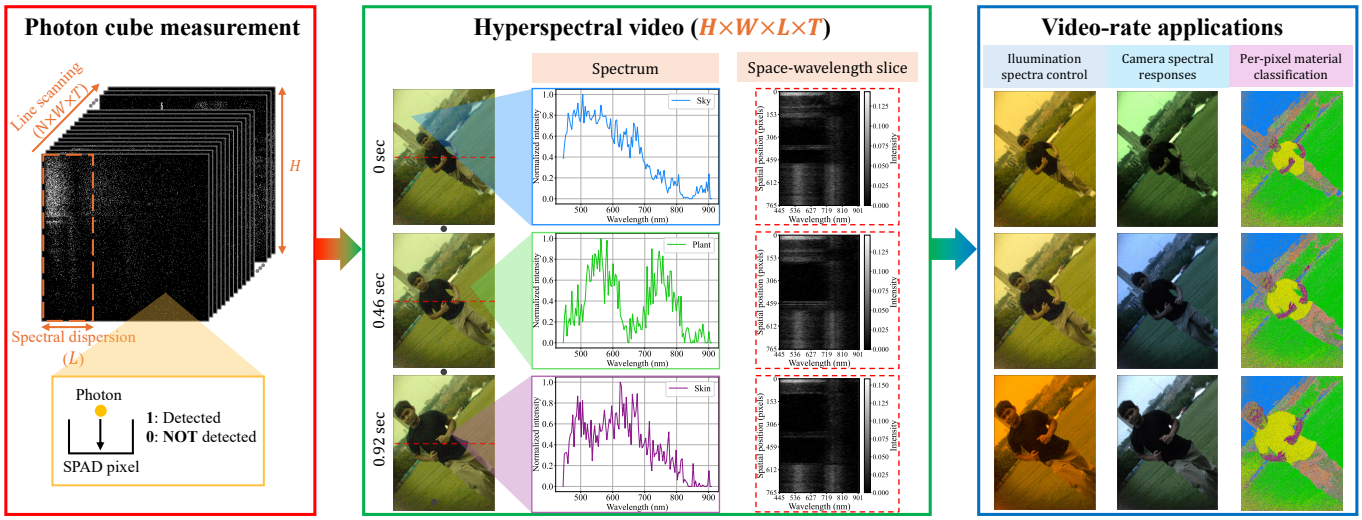


Fig. 1. We present a technique that uses a SPAD image sensor, recording a photon cube (as shown in the red box) to enable sensing of high-resolution hyperspectral videos (green box) and, subsequently, inference with them (blue box). We reconstruct a hyperspectral video with width $W = 768$ pixels, height $H = 1024$, and $L = 123$ spectral bands spanning 445 nm to 910 nm, with $T = 26$ frames captured at a rate of 26 frames per second. Each hyperspectral frame was reconstructed from $N = 5$ binary SPAD measurements. Green box: Reconstructed hyperspectral video from the photon cube using a deep neural network. Space–wavelength slices are cross-sections of the reconstructed 3D hyperspectral cube taken along the vertical spatial axis and the wavelength axis. The positions of these slices are indicated by red dashed lines in the color images. Blue box: Video-rate applications on two different scenes enabled by the captured hyperspectral video data. The first column shows illumination control performed by multiplying the illumination spectrum with the spectrally calibrated hyperspectral video. The second column shows video rendering by multiplying the spectral sensitivity functions of different camera models. The third column presents video-rate pixel-wise material classification results on the raw hyperspectral video.

We present Hi-SPAD, a hyperspectral imaging system that combines a pushbroom architecture with single-photon avalanche diode (SPAD) sensors to enable high spatial, spectral and temporal resolutions under passive ambient light. While SPAD-based spectral sensing and the pushbroom configuration have been explored separately, our key contribution lies in combining these technologies while addressing SPAD-specific challenges. We provide a theoretical analysis showing that, despite the heavily quantized and nonlinear nature of SPAD measurements, accurate material classification and full-spectrum reconstruction can be achieved with exponentially decreasing error bounds as the number of measurements increases. These results are validated through simulations and real data from our lab prototype, establishing the feasibility of photon-counting hyperspectral video in dynamic scenes. Additionally, we evaluate reconstruction methods for Hi-SPAD that support robust imaging under passive ambient illumination. Our lab-built Hi-SPAD camera enables hyperspectral video at 26 frames per second with 123 spectral channels, even under ambient lighting.

Authors' Contact Information: Haejoon Lee, Carnegie Mellon University, Pittsburgh, PA, USA, haejoonl@andrew.cmu.edu; Mohit Gupta, University of Wisconsin-Madison, Madison, WI, USA, mohitg@cs.wisc.edu; Vijayakumar Bhagavatula, Carnegie Mellon University, Pittsburgh, PA, USA, vk16@andrew.cmu.edu; Aswin C. Sankaranarayanan, Carnegie Mellon University, Pittsburgh, PA, USA, saswin@andrew.cmu.edu.

ACM XXXX-XXXX/2026/6-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS Concepts: • **Computing methodologies** → **Hyperspectral imaging**; **Computational photography**.

Additional Key Words and Phrases: Hyperspectral Imaging, SPAD Sensors

ACM Reference Format:
Haejoon Lee, Mohit Gupta, Vijayakumar Bhagavatula, and Aswin C. Sankaranarayanan. 2026. Hi-SPAD: Video-Rate Hyperspectral Imaging and Inference with Single-Photon Cameras. 1, 1 (June 2026), 29 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Imaging problems often have to contend with low light levels. An extreme instance of this occurs in plenoptic imaging, where we seek to resolve the light across the dimensions of space, time, angle, spectrum, and polarization. Since the amount of light in any scene is finite, resolving it across such a large number of dimensions makes even the most favorable of situations—say, an outdoor scene illuminated with direct sunlight—photon starved. This remains one of the enduring challenges toward enabling tractable solutions for sensing the plenoptic function.

In this work, we study imaging hyperspectral video (HSV), a four-dimensional subset of the plenoptic function characterizing how a scene's intensity varies across space, time, and spectrum. Even this lower-dimensional slice presents significant challenges for sensing systems due to the signal's high dimensionality. Consider a megapixel image resolved across 100 spectral channels, and sensed at 30 samples per second; this signal has a bandwidth of three billion samples per second. When a finite amount of light incident on the camera is resolved across these spatial, spectral, and temporal "bins," we can expect the average signal in each of them to be minuscule. This is especially concerning as imaging systems have to contend with two dominant sources of noise: read noise, which arises from thermal effects in the electronics, and photon noise, which is a fundamental property of light describing the stochastic nature of photon arrivals. When the actual amount of signal is very low, both noise sources can overwhelm the signal component.

Identifying noise as a critical challenge for video-rate hyperspectral imaging raises an interesting question: would sensing with photon-counting devices such as single-photon avalanche diode (SPAD) sensors provide a viable operating point? These sensors can detect the arrival of individual photons, returning binary measurements—one if one or more photons are detected in a given exposure time and zero otherwise. A key benefit of such cameras for sensing HSVs is that their measurements are free of read noise owing to the absence of analog-to-digital conversion. SPADs also operate at very high time resolutions, reaching up to 100,000 binary frames per second, thereby providing the requisite bandwidth for sensing HSVs. Yet, despite these natural synergies, sensing such low-light signals—especially with SPADs—must contend with photon noise, the stochastic arrival of photons that is a fundamental property of light. Since signal-to-noise ratio under photon noise decreases with light level, SPAD measurements under low light levels are highly noisy despite the absence of read noise. In addition, SPAD measurements are extremely quantized and nonlinear—specifically, 1-bit values indicating whether one or more photons arrived. Any proposed approach must also contend with these aspects stemming from the unconventional nature of SPAD measurements, all of which require the design of novel methods.

However, photon noise has an interesting property that arises from modeling photon arrivals as a Poisson process. The sum of two Poisson random variables is also Poisson with a commensurate increase in the arrival rates; hence, if we "split" a measurement into multiple lower-exposure measurements, then we do not pay a penalty (as long as there is no read noise and as long as we can handle the additional bandwidth). However, this property of Poisson variables does not naturally extend to SPAD measurements, which are better modeled with a Binomial distribution indicating the presence or absence of photon arrivals. Under this Binomial SPAD measurement model, we conceptualize basic results for a broader range of classification and regression problems commonly studied with HSVs. In the context of classification problems, we show that the misclassification rate associated with a simple cosine-similarity-based classifier decreases exponentially with the number of SPAD measurements. Similarly, for regression problems, we show that the mean squared error exhibits similar behavior. Together, this suggests that even with a few photons per measurement, we can obtain reliable estimators of spectra and accurate classification of signals of interest. In turn, this suggests that SPAD-based sensing of HSVs can indeed be a viable approach.

Contributions. This work develops the first scanning-based video-rate hyperspectral imager without any spectral multiplexing or compressive sensing techniques, using SPAD devices as the sensor of choice, making the following contributions:

- We present a theoretical analysis for simple reconstruction and classification problems, showing that errors in both cases decrease exponentially with the number of binary hyperspectral SPAD measurements.
- We demonstrate effectiveness through simulation experiments with a public dataset.
- We present a high-speed hyperspectral imaging system, Hi-SPAD, capable of capturing high-speed dynamic scenes with high spectral resolution. Specifically, our lab prototype senses HSVs at a resolution of 1024×768 spatial pixels, with 123 spectral bands spanning visible and near-infrared wavelengths, at a maximum rate of 26 frames per second.
- We evaluate our lab-built prototype on a range of scenes, both indoors and outdoors, and across multiple applications of interest.

2 Prior Work

In this section, we review two categories of relevant literature: 1) traditional HSI systems based on CMOS sensors, and 2) emerging approaches using SPAD sensors for spectral measurement. This highlights the trade-offs in classical HSI and how SPAD technology addresses low-light and high-speed limitations.

2.1 HSI with CMOS Sensors

Conventional hyperspectral imaging relies on two broad approaches: scanning-based and snapshot.

Scanning-Based HSI systems capture spectral data sequentially over time (e.g., pushbroom line-scan [Arablouei et al. 2016; Lim and Murukeshan 2015; Turner et al. 2017] or multishot capture with tunable filters [Kurihara et al. 2020; Takamatsu et al. 2024; Wang and Zhang 2016]). This method can achieve high spectral resolution since many narrow bands are measured, but it requires long acquisition times and is prone to motion artifacts in dynamic scenes. High spectral fidelity, therefore, comes at the cost of slower imaging speed to acquire the full hyperspectral cube.

Snapshot HSI. Snapshot imagers use a focal-plane array to collect a full 3D hyperspectral cube in a single capture. Optical arrangements (e.g., spectral filter mosaics [Cheung and Christophersen

2023; Monakhova et al. 2020], coded aperture optics [Lin et al. 2014; Wagadarikar et al. 2008], diffractive optical elements (DOEs) [Jeon et al. 2019], or spatial light modulators [Saragadam et al. 2021]) encode spectral information across a 2D sensor, yielding an instantaneous spectral snapshot. This enables video-rate HSI of transient events. However, snapshot techniques often sacrifice spatial or spectral resolution for speed since finite sensor pixels must be divided among many spectral channels [Hagen and Kudenov 2013].

Recent works on HSI often combine scanning-based and snapshot-based approaches to achieve high-speed HSI with multiple encoding measurements. For example, Aydin et al. [2025] proposed an HSI system that employs a chromatic focal stack to achieve high-speed HSI with ~ 0.2 fps and 12 spectral channels. In the context of scanning with active illumination, Shin et al. [2025] uses multiplexed dispersed structured light to achieve high-speed 3D HSI (HSI + 3D depth), with 6.6 fps and 22 spectral channels. A drawback of the method is that it is inherently limited by the need for active illumination; the system must contend with ambient light and requires the objects to be close enough to achieve sufficient signal throughput.

A fundamental trade-off in HSI systems is between *resolution* in time, spatial, and spectral dimensions, and *signal throughput (exposure time)*. Capturing more and narrower bands for each dimension means each band receives fewer photons, lowering the signal-to-noise ratio (SNR) unless the exposure is increased. Under low-light conditions, conventional CMOS-based HSI faces severe limitations: when photon counts are low, the signal is overwhelmed by readout noise. Moreover, high-speed CMOS cameras—necessary for capturing video-rate HSI using scanning-based methods—tend to exhibit elevated readout noise due to the demands of fast sensor readout, further degrading performance under photon-limited conditions [Baudot et al. 2009; Oxford Instruments Andor 2023].

In this study, we show that SPAD sensors can overcome this limitation. Unlike CMOS-based systems, Hi-SPAD exploits readout-noise-free photon counting to maintain signal fidelity under low light *without increasing exposure time*, thereby achieving video-rate hyperspectral imaging with significantly higher spectral and temporal resolution than prior approaches. As a result, our lab-prototype Hi-SPAD camera, as shown in Figure 1, achieves full video-rate high-dimensional HSI (megapixel images at 26 fps with 123 spectral channels) under ambient lighting.

2.2 Computational Imaging with SPAD Sensors

SPADs are ultrasensitive detectors that trigger an avalanche current upon detecting a single photon, producing a digital pulse for each photon event. They effectively perform *direct photon counting*, eliminating readout noise and enabling detection of extremely weak optical signals [Niclass et al. 2008, 2005; Qian et al. 2023]. These characteristics have made SPAD arrays a popular choice in low-light imaging applications such as LiDAR [Callenberg et al. 2021; Gupta et al. 2019; Lee et al. 2023; Tontini et al. 2020] and fluorescence lifetime imaging [Mai et al. 2023; Ulku et al. 2018; Zickus et al. 2020]. Recent advances in SPAD array technology have enabled their use as full-frame imaging sensors, opening the door to SPAD-based computational imaging systems that leverage their high-speed, noise-free operation. For instance, Ma et al. [2020] demonstrated photon-level burst photography using a SPAD array in passive mode, achieving high-quality intensity imaging under extreme conditions like fast motion and ultra-low light. More recently, SoDaCam [Sundar et al. 2023] proposed a software-defined imaging framework that reinterprets the SPAD's spatio-temporal binary photon detections (photon cube) through post-processing to flexibly simulate diverse camera designs. Beyond monochrome single-photon imaging, Ma et al. [2023] introduced a computational photography approach that reconstructs high-quality color images from mosaicked color-filtered binary SPAD frames, demonstrating robustness even under high-dynamic-range scenes with rapid motion.

In this context, our Hi-SPAD camera extends this line of work by targeting the spectral domain. As illustrated in Figure 1, Hi-SPAD captures the photon cube via line scanning and disperses light spectrally along each scan line. The resulting measurement is then processed to reconstruct video-rate hyperspectral data, making Hi-SPAD a novel computational imaging system that enables spectral inference under severe photon constraints.

2.3 SPAD Sensors for Spectrum Measurement

SPAD sensors in 1D arrays have been widely utilized in spectrometers for point-wise spectral scanning [Erdogan et al. 2019; Krstajić et al. 2015; Kufcsák et al. 2017; Maruyama et al. 2013]. In particular, Erdogan et al. [2019] used a SPAD-based spectrometer for spectral fluorescence lifetime imaging, employing point-wise spectral scanning. This approach, however, is limited to slower scanning speeds (approximately 0.11 frames per second) due to its point-wise scanning nature. More recently, Griffiths et al. [2018] introduced a multispectral imager that sequentially illuminated the scene with nine LEDs of different wavelengths and captured the scene with a SPAD image sensor for each illumination band. Similarly, Zhang et al. [2023] proposed a snapshot hyperspectral imaging system using a SPAD sensor and coded aperture for space target detection. However, these approaches rely on active illumination with only a few spectral bands or use snapshot imaging, which compromises both the spectral and spatial resolution. While these works highlight the potential of SPAD imagers for spectral measurement, none of them have provided a theoretical study of how classification and reconstruction perform under the binary, quantized measurements endemic to SPAD sensors, nor have they shown a viable video-rate hyperspectral imager.

Advantages of SPAD-Based HSI: Summarizing, SPAD imagers offer several advantages over CMOS-based hyperspectral cameras.

- **Low light sensitivity:** SPAD sensors excel in photon-starved conditions. They are *free of readout noise* and can detect individual photons, enabling reliable spectral capture where a CMOS sensor's noise would overwhelm the signal.
- **High frame rates:** SPAD arrays can be read out rapidly (e.g., up to *100,000 binary frames per second* for our lab-built prototype), making them well-suited for video-rate hyperspectral imaging. Such high temporal resolution allows acquiring high-dimensional hyperspectral data without the long scan times required by typical traditional systems, which is critical for dynamic scene analysis.
- **High spectral resolution:** Because of the above advantages of SPAD, one can increase spectral resolution (narrower bands

or more bands) *without sacrificing data fidelity* as severely as in CMOS systems. Even if fewer photons are present per band, the SPAD will register them without added read noise, so fine spectral distinctions can still be made.

By enabling *photon-counting hyperspectral imaging*, SPAD-based systems effectively address the noise and speed limitations of CMOS sensors, paving the way for high-speed, high-resolution hyperspectral imaging.

3 Theoretical Performance Analysis of Hi-SPAD

In this section, we provide a theoretical analysis of the proposed Hi-SPAD camera by establishing error bounds for both spectrum reconstruction and classification.

3.1 Background: CMOS vs. SPAD Measurement Models

It is important to note that the SPAD sensor measurement is not the same as a CMOS sensor measurement without readout noise, because its statistical nature differs from that of a CMOS sensor.

In a CMOS sensor (assuming absence of dark current noise), the total number of photons arrived at a pixel during an exposure time Δt can be modeled as a Poisson random variable. Specifically, if Φ (photons/sec) denotes the mean photon flux at that pixel, then over time Δt the expected photon count is $\Phi\,\Delta t$, and the observed count can be modeled as:

$$L \;\sim\; \text{Poisson}\big(\Phi\,\Delta t\big). \tag{1}$$

However, due to the analog nature of CMOS, the measured signal contains additive readout noise stemming from voltage fluctuations of electronic components:

$$I_{\text{CMOS}} = \min\big(L \,+\, n_{\text{ro}},\; C_{\text{FWC}}\big),\;\text{where } n_{\text{ro}} \sim \mathcal{N}(0, \sigma_{\text{ro}}), \tag{2}$$

where C_{FWC} denotes the full-well capacity of the CMOS pixel. Thus, when the light level ($\Phi\,\Delta t$) is low, the desired signal (L) is often buried under the readout noise (n_{ro}), while at high flux levels the measurement saturates once the accumulated charge reaches C_{FWC} [Hasinoff et al. 2010]. In contrast, SPADs do not suffer from readout noise, their dynamic range is not limited by well-capacity of the sensor, and increases with overall exposure time [Ingle et al. 2019].

From Poisson Process to Binomial Model in SPAD. A SPAD sensor, however, operates differently: each detection event is triggered by the *first* photon arrival within a short gating time. Concretely, rather than measuring a *total* photon count over exposure time (Eq. (1)), the SPAD is run in a series of short "frames," each of duration $\Delta t'$; note that $\Delta t'$ is often extremely short, on the order of a few microseconds. In each frame, the SPAD records either *one arrival* (if at least one photon arrives) or *no arrival*.

Let p be the probability that at least one photon arrives in a single frame of duration $\Delta t'$. Since arrivals are Poisson with rate Φ, the probability of zero arrivals is $\exp(-\Phi\,\Delta t')$. Thus,

$$p \;=\; 1 \;-\; e^{-\Phi\,\Delta t'}. \tag{3}$$

If we collect N such frames, the SPAD measurement is the number of frames that detected at least one photon. Each frame is an independent Bernoulli trial with success probability p, so the overall measurement (neglecting dark current) follows a Binomial distribution:

$$I_{SPAD} \;\sim\; \text{Binomial}(N,\; p = 1 - e^{-\Phi\,\Delta t'}). \tag{4}$$

Equation (4) succinctly captures the difference between CMOS and SPAD measurements. In the CMOS model, each pixel integrates the total photon arrivals over Δt, yielding a *Poisson* random variable. By contrast, a SPAD in "photon-counting" mode counts how many short frames had at least one photon arrival, resulting in a *Binomial* random variable whose parameter is set by the probability of detecting an arrival in each short frame. Owing to the direct digital conversion of photon arrivals, SPAD measurements are free from additive readout noise.

3.2 The Hi-SPAD Measurement Model

Building on the distinction between CMOS and SPAD measurements, we now formulate the Hi-SPAD measurement model, which extends photon-counting statistics to hyperspectral video capture with a pushbroom architecture. Let $h(x, y, \lambda, t)$ denote the hyperspectral intensity of a scene at spatial location (x, y), wavelength λ, and time t. The Hi-SPAD camera captures this 4D signal using a classical pushbroom scanning mechanism, where each frame of the SPAD sensor measures the spectrally dispersed light from a single scan line. Over time, sequential line scans across the scene reconstruct the full hyperspectral video cube.

Formally, the hyperspectral cube at a fixed time t can be written as:

$$\mathbf{I} \in \mathbb{R}^{H\times W\times L},$$

where H and W denote the spatial dimensions, and L is the number of spectral channels.

Each pixel (x, y) in the spatial domain is associated with a raw spectral vector:

$$\mathbf{s}_{x,y} = \begin{bmatrix} \mathbf{I}(x, y, \lambda_1) \\ \mathbf{I}(x, y, \lambda_2) \\ \vdots \\ \mathbf{I}(x, y, \lambda_L) \end{bmatrix}.$$

We define the normalized spectral profile as:

$$\mathbf{d} = \frac{\mathbf{s}_{x,y}}{\max_n \mathbf{s}_{x,y}[n]},$$

where $\mathbf{d} \in \mathbb{R}^L$ and $\max_n \mathbf{d}[n] = 1$. If the maximum photon rate across all spectral channels at the pixel is Φ, the photon rate at each spectral channel n is given by $\Phi \cdot \mathbf{d}[n]$. As we derived in the previous section, the probability of at least one photon arrival at channel n over $\Delta t'$ is

$$\mathbf{p}[n] \;=\; 1 - \exp\big(-\Phi\,\mathbf{d}[n]\,\Delta t'\big). \tag{5}$$

Defining

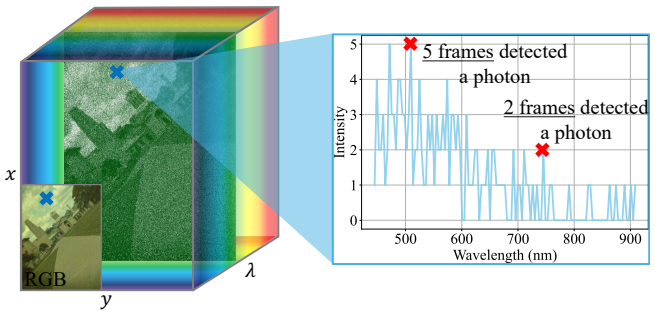
$$\Phi_0 \;=\; \Phi\,\Delta t', \tag{6}$$

we obtain

$$\mathbf{p}[n] \;=\; 1 - \exp\big(-\Phi_0\,\mathbf{d}[n]\big). \tag{7}$$

If N binary frames are captured, the total SPAD measurement at channel n is the number of frames in which at least one photon was detected at that channel. Thus,

$$\mathbf{y}[n] \;\sim\; \text{Binomial}\Big(N,\; \mathbf{p}[n] = 1 - e^{-\Phi_0\,\mathbf{d}[n]}\Big), \quad n \;=\; 1, \ldots, L. \tag{8}$$



(a) Raw Hi-SPAD HSI ($N = 5$)

(b) Measured spectrum ($N = 5$) at a pixel $(231, 121)$

Fig. 2. Example of Hi-SPAD measurement captured by the lab-built prototype. (a) Single hyperspectral cube, captured with $N = 5$ binary frames. The color image at the bottom left of the cube is rendered from the HSI cube. In the middle of the cube, we show a single channel image at wavelength of 400 nm. (b) Spectrum measurement from a single pixel of the hyperspectral cube, denoted as $\mathbf{y}$ in Eq. (8). Intensity values at each spectral channel indicate the number of frames in which at least one photon was detected at that channel.

Figure 2 shows an example of Hi-SPAD measurement captured by our lab-built prototype.

Remark. From Eq. (8), we can notice the SPAD measurement itself is a non-linear function of the true spectrum $\Phi_0\,\mathbf{d}$. Specifically, the expected value of the Hi-SPAD measurement is

$$\mathbb{E}[\mathbf{y}[n]] = N\left(1 - e^{-\Phi_0\,\mathbf{d}[n]}\right) \neq N\,\Phi_0\,\mathbf{d}[n]. \tag{9}$$

The bias arises because the SPAD measurement counts whether each short exposure contains at least one photon arrival, not a direct photon count. When the incident rate of photons is very low,

$$1 - e^{-\Phi_0\,\mathbf{d}[n]} \approx \Phi_0\,\mathbf{d}[n], \quad \text{and} \quad \mathbb{E}[\mathbf{y}[n]] \approx N\,\Phi_0\,\mathbf{d}[n],$$

and the bias goes away. This statistical property has been similarly observed in passive free-running SPAD imaging [Ingle et al. 2019], where inter-photon timing estimators become linear at low flux. Moreover, analogous phenomena have been reported in quanta image sensors [Gnanasambandam et al. 2019]. This statistical property is important in the design of the following classifiers and reconstruction techniques.

3.3 Analysis of Classification Performance

Material classification has long been one of the primary applications for spectrum measurement [Erickson et al. 2018; Hege et al. 2004; Ilehag et al. 2019; Lee and Sankaranarayanan 2024]. We now analyze the performance of cosine classifiers under a SPAD measurement model. Cosine similarity–based classifiers have been widely used in hyperspectral image analysis [Ahmad and Mazzara 2024; Galal et al. 2012; Zou et al. 2019]. The idea is simple: spectra are treated as high-dimensional vectors, and classification is performed by comparing the angle between a measured spectrum and reference spectra from a dictionary. Because cosine similarity depends only on the relative orientation of spectral vectors rather than their absolute magnitude, it is robust to variations in illumination intensity and sensor gain. Our choice of cosine classifiers is motivated by their simplicity, which enables tractable analysis; although they do not model spatial relationships in material composition, they provide a useful baseline as well as important insights into classification using SPAD measurements.

Suppose that we have a dictionary $\mathbf{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_M\}$ consisting of reference spectra for the materials of interest; we assume that each $\mathbf{d}_i$ is unit norm, i.e.,

$$\|\mathbf{d}_i\|_2 = 1.$$

A simple approach to classifying materials in a scene is to compare the spectrum at each spatial location (or pixel) to that of a dictionary of reference spectra for each material. Since we are classifying each pixel in isolation, for the rest of the discussion, we ignore all references to spatial location, and only consider the spectrum profile. Let $\mathbf{d}_k$ be the true (but unknown) spectrum at the pixel under consideration. The Hi-SPAD camera provides us with a measurement of the form:

$$\mathbf{y}[n] \sim \text{Binomial}\Big(N,\ 1 - e^{-\Phi_0\,\mathbf{d}_k[n]}\Big), \quad n = 1, \ldots, L, \tag{10}$$

where Φ_0 indicates the light level during the exposure time for each SPAD frame. Our goal is to design a classifier based on $y[n]$ and analyze its performance.

Exponentially Transformed Spectrum Dictionary. Given that the Hi-SPAD measurement is a biased measurement of the true spectrum, we propose an *exponential transformation* of the reference spectra to enhance classification performance.

We define the exponentially transformed dictionary as

$$\mathbf{p}_i = \frac{1 - e^{-\Phi_0\,\mathbf{d}_i}}{\|1 - e^{-\Phi_0\,\mathbf{d}_i}\|}, \quad i = 1, \ldots, M. \tag{11}$$

Now our measurement is modeled as

$$\mathbf{y} \sim \text{Binomial}\Big(N,\ \alpha_k\,\mathbf{p}_k\Big), \tag{12}$$

with

$$\alpha_k = \|1 - e^{-\Phi_0\,\mathbf{d}_k}\|. \tag{13}$$

Note that now the measurement is an unbiased measurement of the transformed true spectrum $\mathbf{p}_k$ with the scale factor $N\,\alpha_k$:

$$E[\mathbf{y}] = N\,\alpha_k\,\mathbf{p}_k. \tag{14}$$

We can assign the measurement $\mathbf{y}$ to the material whose reference spectrum yields the highest cosine similarity as follows:

$$\ell = \arg\max_j \ \langle \mathbf{y},\ \mathbf{p}_j \rangle. \tag{15}$$

This classifier fails in identifying the k-th material as the correct one, if

$$\exists j \neq k, \langle \mathbf{y},\ \mathbf{p}_k \rangle < \langle \mathbf{y},\ \mathbf{p}_j \rangle \implies \exists j \neq k, \langle \mathbf{y},\ \mathbf{p}_k - \mathbf{p}_j \rangle < 0.$$

We now analyze the probability of misclassification under the Hi-SPAD measurement model.

3.3.1 Bound for Misclassification Probability. We define the similarity score as

$$S_j = \langle \mathbf{y},\, \mathbf{p}_j \rangle, \tag{16}$$

and characterize the term

$$e_j = S_k - S_j = \langle \mathbf{y},\, \mathbf{p}_k - \mathbf{p}_j \rangle. \tag{17}$$

Since misclassification occurs when $e_j \le 0$, our goal is to derive bounds on the probability $P(e_j \le 0)$.

Theorem 3.1 (Exponential Decay via Exponential Transformation). *Let $\mathbf{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_M\}$ be a dictionary of normalized spectra and define the exponentially transformed dictionary as*

$$\mathbf{p}_i = \frac{1 - e^{-\Phi_0 \mathbf{d}_i}}{\|1 - e^{-\Phi_0 \mathbf{d}_i}\|}, \quad i = 1, \ldots, M. \tag{18}$$

Then, the misclassification probability associated with the cosine-similarity classifier (using the transformed dictionary) satisfies

$$\Pr\{e_j < 0\} \le \exp\left\{ -\frac{N \|1 - e^{-\Phi_0 \mathbf{d}_k}\| \,(1 - \cos\theta)}{L \,\max_n \left|\mathbf{p}_k[n] - \mathbf{p}_j[n]\right|} \right\}, \tag{19}$$

where $\cos\theta = \langle \mathbf{p}_k,\, \mathbf{p}_j \rangle$ and L is the number of spectral channels. In particular, the misclassification probability decays exponentially with the number of measurements N.

Sketch of Proof. The proof follows a similar strategy to the low flux case by applying a Chernoff bound to the difference in cosine similarity scores. However, after the exponential transformation, the Hi-SPAD measurement acts as an unbiased estimator for the transformed spectrum. Analyzing the corresponding MGF and choosing an appropriate Chernoff parameter $s > 0$ yields the stated bound. (Detailed steps are provided in the Supplementary Material.) □

Remark. The exponential transform of the spectrum dictionary ensures an exponentially decreasing misclassification probability for all light levels, including the high-flux regime where the nonlinearity of the SPAD measurement is significant. However, the transformed dictionary requires additional information about the effective light level Φ_0 for constructing the exponential transform of dictionary.

From the bound in Eq. (19), we note that the misclassification probability decays faster when (1) the cosine similarity between the true spectrum of the measurement and the reference spectrum, $\cos\theta$, is smaller, and (2) the number of spectral channels, L, is smaller. The dependence on the effective light level Φ_0 is more subtle: increasing Φ_0 increases $\|1 - e^{-\Phi_0 \mathbf{d}_k}\|$, but it also changes the transformed spectra $\mathbf{p}_k$ and $\mathbf{p}_j$ (and hence $\cos\theta$ and $\max_n |\mathbf{p}_k[n] - \mathbf{p}_j[n]|$), so the overall decay rate can either improve or degrade depending on how the transformed dictionary separation varies with Φ_0.

Corollary 3.2 (Low Flux Approximation). *In the low-flux regime, the exponential transformation of the dictionary is not necessary. As noted earlier, when the incident rate of photons is extremely small,*

$$1 - e^{-\Phi_0 \mathbf{d}_k} \approx \Phi_0 \,\mathbf{d}_k,$$

and hence $\mathbf{p}_i$, as defined in Eq. (11) is identical to $\mathbf{d}_i$. Under this setting, the cosine classifier in Eq. (15) can be analyzed by looking at the score

$$S_j = \langle \mathbf{y},\, \mathbf{d}_j \rangle, \tag{20}$$

which is simply the direct cosine similarity between the measurement and the reference spectrum.

3.4 Analysis of Reconstruction Performance

Beyond classification, reconstructing the full spectrum from HSI measurements is another important application. Based on the statistical properties of SPAD measurements, we analyze the reconstruction performance.

Log-Transformed Estimator. Given that the Hi-SPAD measurement is a nonlinearly biased measurement of the true spectrum, we propose a *log-transformed estimator*, which is the maximum likelihood estimator (MLE) for the true spectrum. The log-transformed estimator is as follows:

$$\widehat{\mathbf{d}}[n] \;=\; -\frac{1}{\Phi_0}\,\ln\!\Big(1 - \tfrac{\mathbf{y}[n]}{N}\Big). \tag{21}$$

It is worthwhile to note that this log-transform estimator is the MLE for $\mathbf{d}[n]$. Let's consider the Binomial likelihood

$$L\big(\mathbf{d}[n]\big) \;=\; \binom{N}{\mathbf{y}[n]} \Big[1 - e^{-\Phi_0\,\mathbf{d}[n]}\Big]^{\mathbf{y}[n]} \Big[e^{-\Phi_0\,\mathbf{d}[n]}\Big]^{N - \mathbf{y}[n]}. \tag{22}$$

Taking the natural logarithm and dropping the constant term $\ln\binom{N}{\mathbf{y}[n]}$ gives the log–likelihood

$$\ell\big(\mathbf{d}[n]\big) \;=\; \mathbf{y}[n]\,\ln\!\Big(1 - e^{-\Phi_0\,\mathbf{d}[n]}\Big) - (N - \mathbf{y}[n])\,\Phi_0\,\mathbf{d}[n]. \tag{23}$$

Differentiating $\ell\big(\mathbf{d}[n]\big)$ with respect to $\mathbf{d}[n]$ and setting the derivative to zero yields

$$\frac{\mathbf{y}[n]\,\Phi_0\, e^{-\Phi_0\,\mathbf{d}[n]}}{1 - e^{-\Phi_0\,\mathbf{d}[n]}} - (N - \mathbf{y}[n])\,\Phi_0 \;=\; 0. \tag{24}$$

After rearrangement, this condition is equivalent to

$$e^{-\Phi_0\,\mathbf{d}[n]} \;=\; 1 - \frac{\mathbf{y}[n]}{N}, \tag{25}$$

or

$$\mathbf{d}[n] \;=\; -\frac{1}{\Phi_0}\,\ln\!\Big(1 - \frac{\mathbf{y}[n]}{N}\Big), \tag{26}$$

proving that the log-transformed estimator is the MLE for $\mathbf{d}[n]$.

3.4.1 Bound for Mean Squared Error.

Theorem 3.3 (Exponential Error Decay via Log-Transform). *In the general setting (no assumption on the light level), our measurement is modeled as*

$$\mathbf{y}[n] \;\sim\; \text{Binomial}\Big(N,\; 1 - e^{-\Phi_0\,\mathbf{d}[n]}\Big). \tag{27}$$

Let's define the log-transform estimator

$$\hat{\mathbf{d}}[n] \;=\; -\frac{1}{\Phi_0}\,\ln\!\Big(1 - \frac{\mathbf{y}[n]}{N}\Big). \tag{28}$$

Then there exists $\kappa_n > 0$ such that

$$P\Big(\big|\hat{\mathbf{d}}[n] - \mathbf{d}[n]\big| \;\ge\; \epsilon\Big) \;\le\; 2\,\exp\!\big(-\kappa_n\, N\big). \tag{29}$$

Moreover, the same union bound argument implies exponential decay of the overall ℓ_2 error across all L channels.

Sketch of Proof. The event $\{\hat{\mathbf{d}}[n] - \mathbf{d}[n] \geq \epsilon\}$ can be rewritten in terms of $\mathbf{y}[n] \geq N(1+\delta)(1-e^{-\Phi_0 \mathbf{d}[n]})$ for some $\delta > 0$. A Chernoff bound on $\mathbf{y}[n]$ then yields an exponential tail. The detailed proof appears in the Supplementary Material. □

Corollary 3.4 (Extension to All Channels). *If $\hat{\mathbf{d}}[n]$ is defined as above for each $n = 1, \ldots, L$, then a union bound implies*

$$P\Big(\max_{1\leq n\leq L}\big|\hat{\mathbf{d}}[n] - \mathbf{d}[n]\big| \geq \epsilon\Big) \leq \sum_{n=1}^{L} 2\,\exp(-\kappa_n N) \leq 2L\,\exp(-\kappa N), \tag{30}$$

where $\kappa = \min_n \kappa_n$. In particular, the ℓ_2 error also decays exponentially:

$$\|\hat{\mathbf{d}} - \mathbf{d}\|^2 \;\leq\; L \max_n \big|\hat{\mathbf{d}}[n] - \mathbf{d}[n]\big|^2. \tag{31}$$

Corollary 3.5 (Low Flux Approximation). *Under the low flux regime, the log-transformed estimator becomes approximately scaled estimator: Given that y/N is very small,*

$$\ln\Big(1 - \frac{\mathbf{y}[n]}{N}\Big) \;\approx\; -\frac{\mathbf{y}[n]}{N}, \tag{32}$$

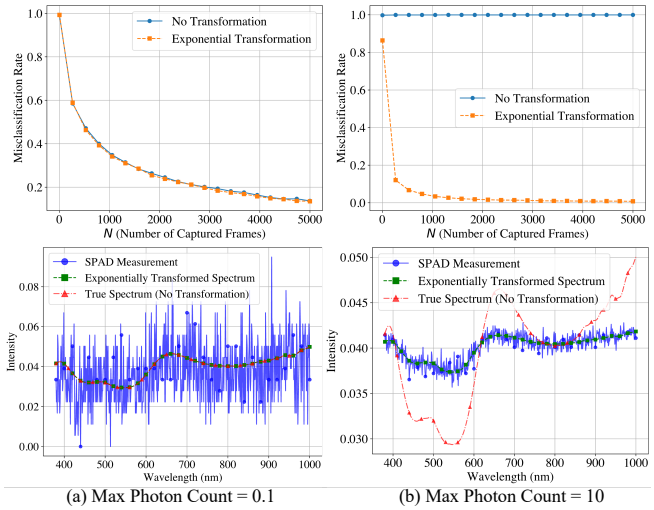
and the log-transformed estimator becomes approximately a scaled estimator:

$$\widehat{\mathbf{d}}[n] \;\approx\; \frac{\mathbf{y}[n]}{\Phi_0\, N}. \tag{33}$$

3.5 Validation of the Theoretical Results

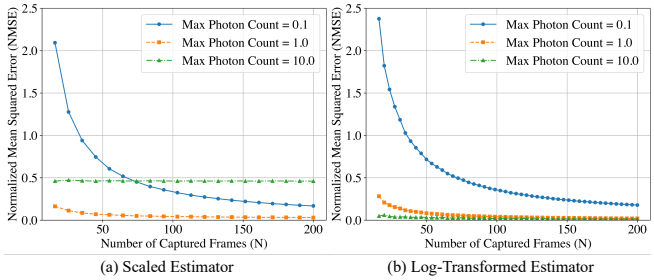
To validate the presented theoretical results as well as provide a comparison against a conventional CMOS sensor, we conducted simulation-based experiments using the USGS spectrum dataset [Kokaly et al. 2017].

3.5.1 Validation of Classification Performance. We simulated the Hi-SPAD measurements for each spectrum in the USGS dataset following Eq. (10) and applied the cosine classifier with both the unaltered and exponentially transformed dictionaries. The misclassification accuracy for both dictionaries at different light levels is shown in the top row of Figure 3. We define 'Max Photon Count' as the maximum number of photons that can be detected by the SPAD sensor in a single frame. The results demonstrate that the misclassification probability decays exponentially with the number of measurements only for low light levels (max photon count = 0.1) when we compute the similarities with the unaltered dictionary. Meanwhile, the exponentially transformed dictionary achieves exponential decay for both low and high light levels (max photon count = 10). The result could be intuitively understood in the bottom row of Figure 3 where we can see that the Hi-SPAD measurement is an unbiased estimator of the exponentially transformed spectrum rather than the true spectrum when the max photon count is 10. Due to the increased bias in Hi-SPAD measurements at high light levels, the misclassification probability remains at 1 (100%) even with an increased number of measurements. The results validate Theorem 3.1.



(a) Max Photon Count = 0.1 (b) Max Photon Count = 10

Fig. 3. Classification comparison between direct cosine classification on original spectra ("No Transformation") and exponentially transformed spectra for low light levels (max photon count = 0.1) and high light levels (max photon count = 10).



(a) Scaled Estimator (b) Log-Transformed Estimator

Fig. 4. Reconstruction error (MSE) with simulated Hi-SPAD measurements. The scaled estimator produces exponentially decaying error only for low light levels, while the log-transformed estimator shows exponential decay for all light levels.

3.5.2 Validation of Reconstruction Performance. To validate our theory on the spectrum estimation, we conducted experiments for true spectrum reconstruction with the USGS dataset. Each Hi-SPAD measurement was generated and applied to the scaled estimator and log-transformed estimator. Figure 4 presents the computed normalized MSE (NMSE) for each estimator with 3 different light levels. The NMSE was computed for every 10 measurements for each spectrum in the dataset. As shown in the figure, the log-transformed estimator shows exponential decay of the error with the number of measurements for all light levels, while non-decaying error is observed for the scaled estimator when max photon count is 10. This result validates Theorem 3.3.

3.5.3 Comparison with Simulated Conventional CMOS Sensor. Along with Hi-SPAD measurements, we also simulated a CMOS baseline and evaluated cosine classification performance. We modeled CMOS measurements as Gaussian, with mean equal to the true spectrum

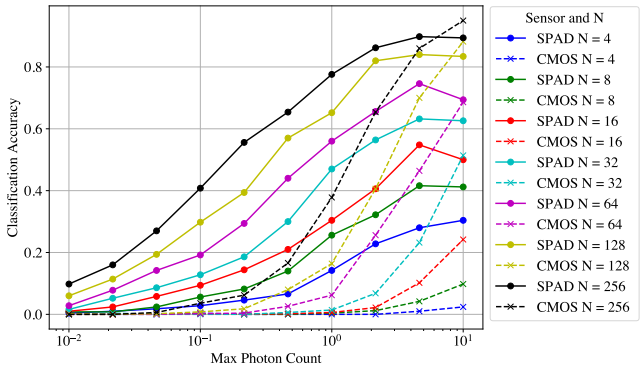


Fig. 5. Classification accuracy comparison between simulated SPAD and CMOS sensor measurement for different number of 1-bit frames (N). For CMOS sensor, we computed cosine angle directly on the measurement while for SPAD sensor we computed cosine angle with exponentially transformed dictionary. Exposure time for CMOS sensor was set as N times of max photon count to assure the equal exposure time for both sensors.

and variance determined by the sensor readout noise. To match the video-rate operation of our Hi-SPAD prototype with pushbroom acquisition, a CMOS-based line-scanning system would need to capture one frame per scanned column. With $W = 768$ columns per hyperspectral cube and 26 cubes per second, the required sensor frame rate is $W \times 26 \approx 2.0 \times 10^4$ frames per second. Such acquisition rates are beyond the operating regime of typical CMOS cameras for general purpose imaging. Therefore, for the CMOS comparison we use a readout-noise standard deviation of $24\,e^-$, consistent with a state-of-the-art high-speed CMOS camera capable of operation at the required frame rate of $\sim$ 20,000 fps [Pha 2025].

The classification accuracy comparison with different number of binary frames (N) for SPAD sensor and corresponding CMOS sensor is shown in Figure 5. For SPAD sensor measurement, we computed cosine angle with exponentially transformed dictionary while for CMOS sensor measurement it was computed with unaltered dictionary; because the CMOS measurement is an unbiased estimator for the true spectrum. Exposure time for CMOS sensor was set as N times of max photon count to assure the equal exposure time for both sensors. As shown in the figure, classification with Hi-SPAD measurements outperforms CMOS measurements under low-light conditions, where readout noise dominates the signal in CMOS sensors. As the light level increases (N times the maximum photon count for CMOS), the classification accuracy of CMOS measurements improves due to the higher SNR, while Hi-SPAD accuracy begins to decline because of the saturation effect in SPAD-based spectral measurements (where values are capped at N). Consequently, CMOS eventually surpasses Hi-SPAD in bright-light settings. As discussed in Section 2.1, capturing video-rate HSI with high spatial and spectral resolution is fundamentally a low-light imaging problem. These results highlight the advantages of SPAD sensors over high-speed CMOS sensors for enabling practical HSV cameras without compromising the spatial and spectral resolution.

4 Hardware Prototyping and Evaluation

In this section, we present the design of the Hi-SPAD camera, a high-speed hyperspectral imaging system that leverages the single-photon sensitivity of SPAD sensors in combination with a line scanning mechanism.

4.1 Hardware Design

Our hardware prototype is built on a pushbroom camera architecture with a Vialux DLP7000 Digital Micromirror Device (DMD) for programmable line scanning and integrates a Pi Imaging SPAD 512 sensor to enable video-rate hyperspectral imaging under photon-limited conditions. Figure 6 showcases our lab-built prototype of the Hi-SPAD camera, with a list of specific components detailed in the supplemental material.

The line-scanning HSI acquisition process with a DMD is as follows. The light coming from the scene is focused on the DMD through an objective lens and a 4f relay system. Then the DMD selectively reflects a single column of the focused image by flipping a corresponding single column of micromirrors toward the imaging arm. The DMD that we used flips its micromirrors along their diagonal, deflecting the optical axis off the horizontal plane; to resolve this, we tilted the DMD and the SPAD sensor by 45°.

We place a spectral disperser between the DMD and the SPAD sensor. We designed a grism, a combination of a transmission grating and a prism [Bowen and Vaughan 1973], to simplify the overall alignment and avoid additional skew in the optical axis; specifically, we used a transmissive grating with 70 grooves/mm and a prism with a wedge angle of 5.22°, so that light at 668nm pass through undeflected. With this, the reflected light from the DMD first reaches the grating, where the dispersion is performed, and the prism redirects the first-order dispersion to the desired location on the sensor. A closer look at the captured column-wise dispersed images is shown in the first row of Figure 7. By sequentially displaying columns on the DMD and capturing the spectrally dispersed images of each column, we can acquire the data for reconstructing the hyperspectral cube.

The spatial and spectral dimensions of the hyperspectral cube are determined by the DMD, SPAD sensor resolution, and optical magnification. The DMD (1024×768 micromirrors) serves as the spatial scanning interface, while the SPAD sensor (512×512 pixels) captures the spectrally dispersed light from each selected column. The magnification was chosen so that all 1024 DMD rows are mapped onto the 512-pixel height of the SPAD sensor. The horizontal axis of the SPAD sensor is reserved for spectral dispersion, with 123 pixels allocated for dispersion, yielding 123 spectral channels. To avoid overlap between first- and second-order diffraction, we computed the angular separation of the disperser and restricted the usable spectrum with a bandpass filter, resulting in an operational range of 445 nm to 910 nm. While the 768 DMD columns are preserved through sequential scanning, the vertical dimension is limited by the SPAD resolution. The 1024 DMD rows are optically compressed onto 512 SPAD pixels, leading to an effective spatial resolution of 512×768 in the reconstructed hyperspectral cube.

As a final step, hyperspectral video is captured by repeating the above process for each hyperspectral frame, exploiting the high

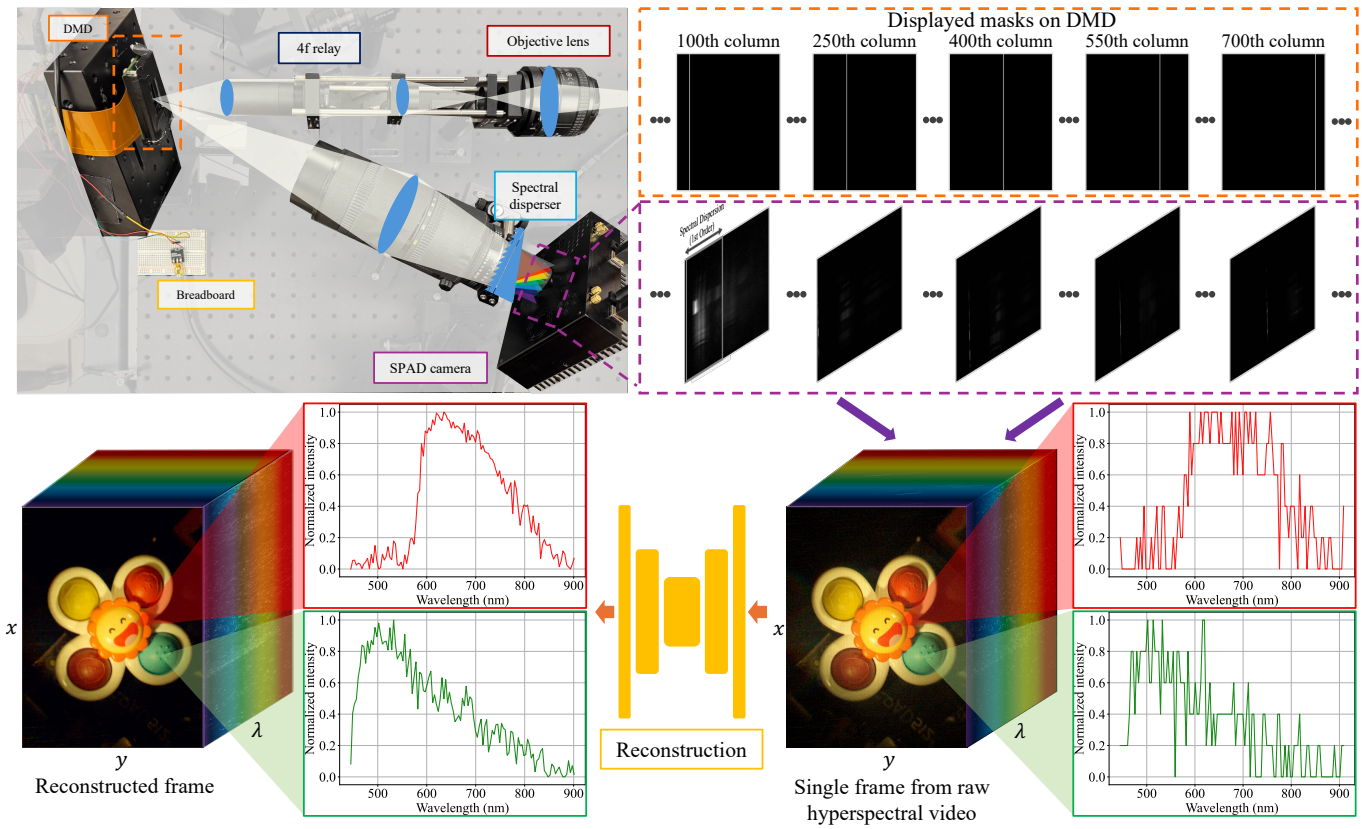


Fig. 6. The overall process of Hi-SPAD. Top left figure shows the lab-built prototype of the Hi-SPAD camera. The top right figure shows displayed mask patterns on the DMD for line scanning and capture of dispersed image with SPAD camera. Bottom figure shows the reconstruction process of the full hyperspectral cube from the SPAD measurement.

frame rates of the DMD and SPAD sensor. The maximum speed of the DMD is 20,000 patterns per second and the SPAD sensor is 100,000 binary frames per second; the fastest (and noisiest) capture per single column is five 1-bit frames at 20,000 fps. With this operation, the system can acquire 26 fps hyperspectral video with 123 spectral channels in the wavelength range of 445 nm to 910 nm. More details of the hardware operation are described in the supplementary material.

4.2 Spectral Calibration and HSI Reconstruction

Figure 7 shows the captured spectrally dispersed images of the sequence of DMD patterns and captured images when a monochromatic light source is illuminated onto the DMD with a dot-column pattern used for calibration. We define a mapping function F that relates the opened DMD pixel coordinates $(x_{\text{dmd}}, y_{\text{dmd}})$ and wavelength (λ) to the corresponding SPAD sensor coordinates:

$$(x_{\text{img}},\ y_{\text{img}}) = F(x_{\text{dmd}},\ y_{\text{dmd}}, \lambda). \tag{34}$$

From the images in the second row of Figure 7, we detected pixel locations of dots corresponding to first-order dispersion and built a dataset of the DMD pixel coordinates and the corresponding SPAD sensor coordinates under different wavelengths: An approximation $\widehat{F}$ of F is then obtained by least-squares fitting on the calibration dataset. Using the obtained mapping function $\widehat{F}$, the hyperspectral cube is reconstructed by remapping the pixel-wise SPAD measurements to pixel coordinates of the DMD at the corresponding wavelength channels. More details of the calibration process are provided in the supplementary material.

Since the SPAD sensor is highly sensitive to photon-level signals, stray light can easily contaminate measurements. Despite enclosing the setup with black cloth and cardboard to block stray light, scene light passing through the objective lens can still reflect off the DMD and reach the SPAD sensor, even when an all-zero pattern is displayed (i.e., all micromirrors are directed away from the imaging path). As shown in the top left corner of Figure 8, this back-reflected light triggers consistent sensor activations. During line scanning, this appears as spatially translating bright spectral peaks, resulting in rainbow-like streaks in the reconstructed RGB image or persistent bright pixels in individual monochromatic bands. To mitigate stray-light artifacts, we apply background subtraction. Using the calibrated mapping, we first mask out the first-order dispersion region and estimate a background image by averaging the remaining pixels, which are dominated by stray light. This background is then subtracted from each column-wise raw measurement to obtain

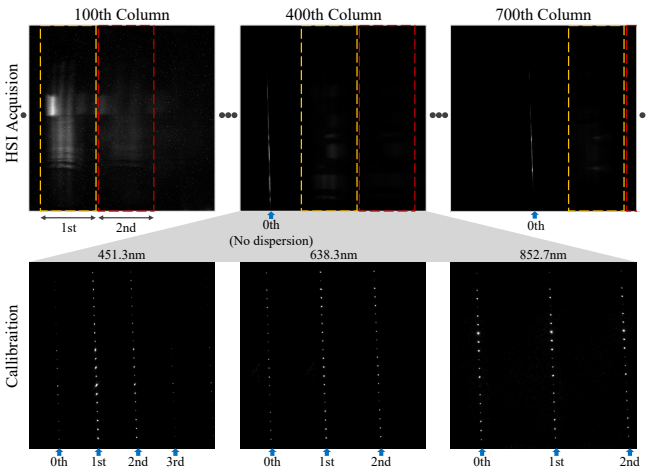


Fig. 7. Images captured by SPAD camera with column-wise DMD patterns. First row shows the column-wise dispersed images of the scene illuminated by broadband light source. Each diffraction order that lies on the sensor is labeled. The 1st order dispersion is only employed for HSI reconstruction. Second row shows the captured images with 400^{th} dot-column pattern on DMD directly illuminated by monochromatic light source with different wavelengths for spectral calibration of the Hi-SPAD camera. Calibration process is described in Section B.2.

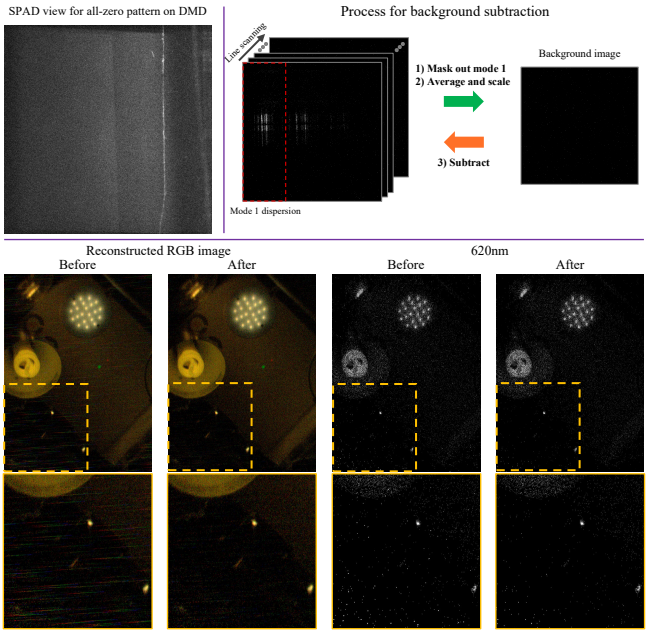


Fig. 8. Background subtraction for removing stray light artifacts. Top right corner shows the SPAD camera view with all-zero pattern displayed on the DMD. Top left corner shows the process of background subtraction. Bottom shows the comparison of the raw HSI cube and the background subtracted HSI cube for the scene with low-light condition captured with smallest aperture of the objective lens. After the background subtraction, the rainbow streaks in the reconstructed RGB image and the bright pixels in the monochromatic channel are significantly reduced.

the corrected image. Through this method, we avoid capturing a separate background image for each acquisition, which would add additional time to the capture process and slow down the frame rate. The bottom row of Figure 8 illustrates the effect: after background subtraction, rainbow streaks and spurious bright pixels are substantially reduced, even under extreme low-light conditions captured using the smallest aperture setting. More details are provided in the supplement section.

4.3 Spectrum Reconstruction Methods

We have shown that the log-transformed estimator is the MLE for the true spectrum and provides exponential error decay with an increasing number of measurement frames. However, under the low-flux regime such as hyperspectral video capture, the log-transformed estimator may not be the best choice. In the scenario of video-rate capture (small number of frames N) with a low probability of photon arrival such as an indoor scene, the number of detected photons is very sparse and highly quantized, as shown in the raw measurement of our lab-built prototype in Figure 6. Thus, we propose multiple methods to reconstruct the full spectrum from the noisy SPAD measurements and compare the reconstructed spectra with the ground truth spectrum.

Low-Rank Denoising. The second method applies low-rank denoising to the Hi-SPAD measurements. Hyperspectral data are often well-approximated by a low-dimensional subspace because material spectra can be represented using a small number of underlying basis components. This low-rank prior has been widely adopted in hyperspectral image denoising [Chen et al. 2024; Fan et al. 2019; Zhuang et al. 2021]. In our implementation, we reshape the hyperspectral cube into a matrix by stacking all pixel spectra as column vectors. We then perform Singular Value Decomposition (SVD) on this matrix and retain only the top r singular components. The denoised spectral matrix is reconstructed using these leading singular values and corresponding singular vectors, effectively suppressing noise while preserving the dominant spectral structure.

Super-Pixel Rank-1 Denoising. We further exploit local spectral redundancy by assuming that neighboring pixels within a homogeneous region share the same spectral *shape* up to a scale factor. Following the super-pixelated HSI prior in [Saragadam et al. 2021], we first form a guidance image by averaging the hyperspectral cube across wavelengths, lightly smooth it to suppress noise while preserving edges, and then segment it into super-pixels (We compute super-pixels using the `quickshift` algorithm from the `scikit-image` library (`skimage.segmentation.quickshift`).) For each super-pixel, we stack the spectra of its pixels into a matrix and compute a rank-1 approximation (via SVD), which enforces a single dominant spectral signature per region while allowing per-pixel scaling.

Pixel-Wise Reconstruction Network. To exploit the learnable prior from public dataset, we trained a pixel (spectrum)-wise reconstruction network on the USGS spectrum dataset. After preprocessing, we obtained a dataset with 1797 spectra with unique material labels.

We cropped each spectrum to the range of 380 nm to 1000 nm (total 621 channels) where our SPAD sensor in the prototype Hi-SPAD camera operates, and normalized to have unit ℓ_2 norm. Then we downsampled the spectrum to 123 channels to match the number of channels of Hi-SPAD. The network has autoencoder architecture with 1D convolutional layers and takes the Hi-SPAD measurement as input and outputs the reconstructed spectrum. Raw Hi-SPAD measurements were simulated with randomly selected values of N and light levels every epoch in training phase, and the network was trained to minimize the mean squared error between the reconstructed spectrum and the ground truth spectrum.

Patch-Wise Reconstruction Network. The last method is a patch-wise reconstruction network trained on simulated data derived from the ICVL hyperspectral image dataset [Arad and Ben-Shahar 2016]. We selected 50 hyperspectral images from the dataset, downsampled the spectral dimension to 123 bands to match the Hi-SPAD configuration, and used them for training and validation. As in the pixel-wise reconstruction network, the raw Hi-SPAD measurements were simulated during training by sampling from the corresponding Binomial measurement model.

The network operates on hyperspectral patches of size $64 \times 64 \times 123$ (width×height×bands) and outputs the reconstructed denoised patch. We adopt the Hybrid Spectral Denoising Transformer (HSDT) architecture [Lai et al. 2023], a state-of-the-art HSI denoising model that combines convolutional inductive bias with spectral self-attention. Specifically, HSDT follows a hierarchical U-shaped encoder–decoder structure with skip connections and integrates (i) spectral–spatial separable convolution (S3Conv) for efficient local feature extraction, (ii) guided spectral self-attention (GSSA) for modeling long-range inter-band dependencies, and (iii) self-modulated feed-forward networks (SM-FFN) for adaptive feature refinement. These components allow the model to jointly capture spatial correlation and global spectral structure. Compared to the pixel-wise approach, the patch-wise reconstruction network exploits spatial correlation and cross-pixel spectral consistency. More architectural and implementation details can be found in the original HSDT paper [Lai et al. 2023].

4.4 Prototype Evaluation

To evaluate the spectral resolution of the lab-built Hi-SPAD camera, we illuminated a close object with monochromatic lasers of different wavelengths, then captured the hyperspectral image of the scene. For each capture, we illuminated the same spatial region to ensure the imager's ability to capture spectral peaks consistently. We then computed the Full Width at Half Maximum (FWHM) of spectral peaks in the spectra averaged over the pixels of the illumination spots for each wavelength. The result is shown in Figure 9. Because of differing laser powers and the wavelength-dependent spectral sensitivity of our lab prototype, the spectral peaks are not identical. Resolution notably decreases in the NIR wavelengths. We attribute this to the incomplete spectral response of our optics, such as compound lenses.

As a proof of concept, we captured the hyperspectral image of the scene of different light sources with distinct spectrum and reconstructed the spectrum with one of the proposed methods, patch-wise

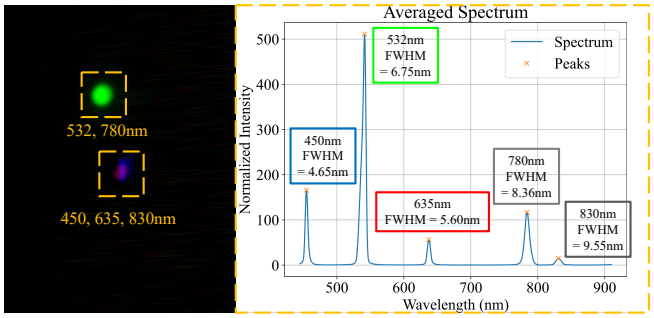


Fig. 9. Spectral resolution analysis. Left figure shows the scene with compounded laser illumination, and right figure plots averaged spectra over the illumination spots. Full Width at Half Maximum (FWHM) was computed for each spectral peak. Because of the distinct power of the laser and wavelength-dependent spectral sensitivity of our lab-prototype, peaks of the spectrum are not identical.

reconstruction network. Each hyperspectral image with different numbers of 1-bit frames was computed by averaging different numbers of frames from the hyperspectral video for the static scene with a 10 µs exposure time per frame. The result is shown in Figure 10 (a). In the spectrum plots, we show the spectrum of captured HSI with different numbers of 1-bit frames. Notably, the reconstructed spectra from the sparsest measurement $N = 5$ are very close to the spectra from the dense measurement $N = 502$. In the Figure 10 (b), we compare the captured spectra of two Incandescent light sources, one LED and one Compact Fluorescent light (CFL) lamp with corresponding spectrometer measurements (lasers are excluded because spectrometer measurements of reflected laser light are difficult to obtain) after calibration by dividing by the spectral sensitivities of the measurement devices. We observe that both the dense measurements and the reconstructed spectra from the sparsest measurements align well with the spectrometer measurements. Moreover, they accurately capture the characteristic spectral peaks of the CFL lamp. However, the Hi-SPAD measurements exhibit notably higher noise at longer wavelengths for the two incandescent light sources after spectral calibration. This is caused by extremely sparse photon detections in that range, due to the reduced sensitivity of the optics and SPAD sensor in the near-infrared wavelengths.

5 Experiments and Applications

We now showcase the performance of the Hi-SPAD system with both simulated measurements as well as real-world data captured with our lab-built prototype camera.

5.1 Comparison against State-of-the-Art

To validate the video-rate performance of the Hi-SPAD camera, we compared it against a CMOS-based line-scanning HSI system operating with the same pushbroom mechanism as Hi-SPAD and a CASSI-based HSI system, a compressive-sensing approach for video-rate HSI, using the ICVL dataset. For the CMOS-based line-scanning HSI (denoted as "HSI-CMOS" in the following), we simulated the HSI data with the assumption that the measurement follows a Gaussian distribution with the mean equal to the true spectrum and the

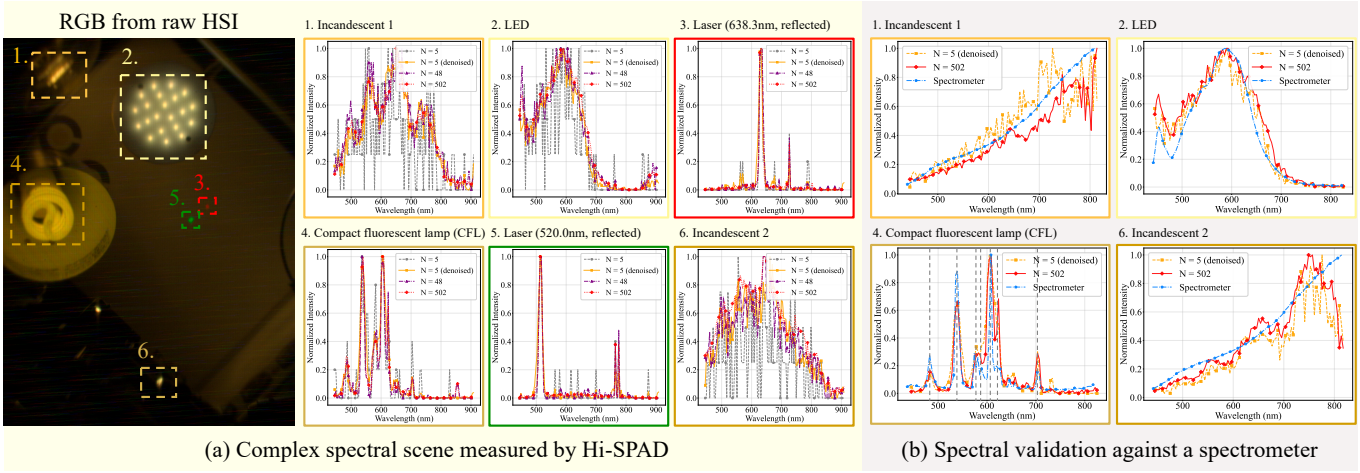


Fig. 10. Comparison of Hi-SPAD spectral measurements from hyperspectral cubes captured under multiple light sources. (a) Raw and reconstructed Hi-SPAD spectra. The "denoised" curve shows the spectrum reconstructed by the patch-wise reconstruction network from the sparsest measurement ($N = 5$), corresponding to video-rate capture. Laser spectra were obtained from reflected illumination, and therefore include peaks from other light sources present in the scene. (b) Spectral validation against a spectrometer. Spectra from two incandescent sources, one LED, and one CFL lamp are compared with spectrometer measurements (lasers are excluded because spectrometer measurements of reflected laser light are difficult to obtain). All spectra are calibrated by dividing by the spectral sensitivity of the respective measurement devices. The results show that the Hi-SPAD imager accurately captures characteristic spectral peaks, such as the distinctive emission lines of the CFL lamp. Peak labels for CFL light source follow [Deglr6328 2005]. Noisy measurements at longer wavelengths for the two incandescent light sources are due to the relatively low sensitivity of the optics and SPAD sensor on non-visible wavelengths and even amplified after spectral calibration.

variance equal to the readout noise. As in Section 3.5.3, we set the standard deviation of the noise to $24\,e^-$, which matches a latest high-speed CMOS camera model [Pha 2025] capable of enabling a video-rate pushbroom camera (~ 20,000 fps). For the CASSI-based HSI, we employed RndHRNet [Zhang et al. 2024], one of the latest deep-learning-based CASSI techniques. For the simulated single-shot CASSI measurements, which serve as input to RndHRNet, we added readout noise with a standard deviation of $2\,e^-$, reflecting a realistic value for conventional CMOS sensors capable of operating at sufficient speed for video-rate single-shot HSI. For the Hi-SPAD camera, we simulated the HSI data following Eq. 8. For HSI-CMOS and Hi-SPAD measurements, we employed simple super-pixel rank-1 denoising to obtain the reconstructed spectrum. The simulation light level was set to a max photon count of 2.5, which is a realistic level for the lab-built prototype operating at maximum speed outdoors on a sunny day.

Quantitative comparison of the reconstruction quality is shown in Table 1. We computed the PSNR, SSIM, and Spectral Angle Mapper (SAM) for the 20 reconstructed hyperspectral cubes randomly selected from the ICVL dataset. The SAM value indicates the spectral similarity between the ground truth hyperspectral cube and the reconstructed hyperspectral cube. It is computed for each pixel in the hyperspectral cube as follows:

$$\mathrm{SAM}(\mathbf{d}, \widehat{\mathbf{d}}) = \arccos\left(\frac{\mathbf{d}\cdot\widehat{\mathbf{d}}}{\|\mathbf{d}\|\|\widehat{\mathbf{d}}\|}\right), \tag{35}$$

where $\mathbf{d}$ and $\widehat{\mathbf{d}}$ are the spectra of the pixel in the ground truth hyperspectral cube and the reconstructed hyperspectral cube, respectively.

Table 1. Quantitative comparison of hyperspectral image reconstruction from three different methods, including a pushbroom camera using a high-speed CMOS sensor, a CASSI imager with the reconstruction technique of Zhang et al. [2024], and the proposed SPAD-based imager with a super-pixel rank-1 denoiser. Metrics are computed for 20 randomly selected hyperspectral images from the ICVL dataset and averaged over the test set. Higher values are better for PSNR and SSIM, and lower values are better for SAM. SAM values highlight cross-band spectral shape errors that are not well captured by PSNR/SSIM; Hi-SPAD produces consistently lower angles, indicating better spectral fidelity despite similar PSNR/SSIM.

Method	Rec. algo.	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓
HSI-CMOS	Super-Pixel	13.38	0.059	11.76
CASSI	RndHRNet	28.48	**0.811**	5.07
Hi-SPAD	Super-Pixel	**28.88**	0.800	**3.23**

We report SAM alongside PSNR and SSIM because SAM directly measures per-pixel spectral fidelity, which is the quantity of interest in HSI. Unlike PSNR and SSIM, which assess band-wise intensity errors and spatial structure, SAM compares the direction of the spectral vectors and is invariant to global rescaling (illumination or albedo). Consequently, it better reflects material discriminability and spectral correctness. In our results (Table 1), Hi-SPAD attains the lowest SAM (3.23°), a 35% reduction relative to CASSI-RndHRNet (5.07°), indicating more accurate spectral shapes. The advantage arises from Hi-SPAD's read-noise-free photon counting and the rank-1 super-pixel denoiser, which together stabilize spectral shape under photon-limited conditions. The PSNR and SSIM differences

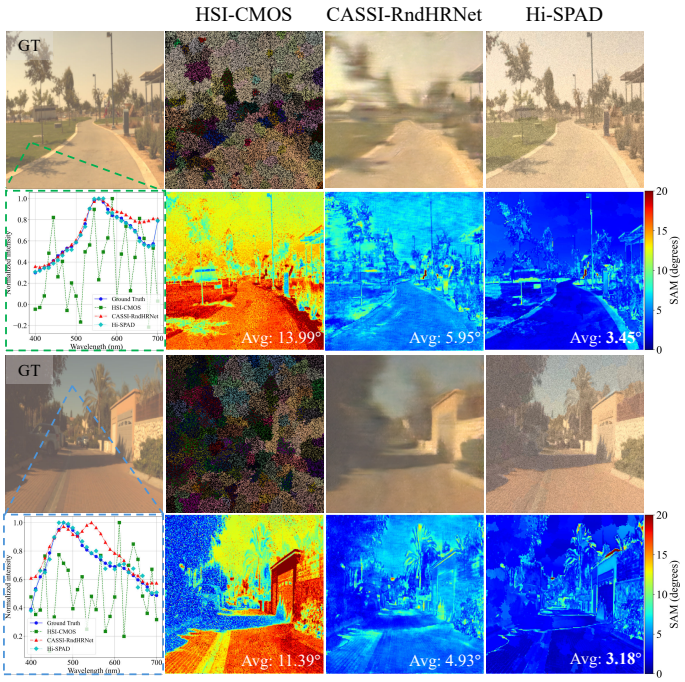


Fig. 11. Reconstruction comparison of three simulated methods. Each color image is rendered from the reconstructed hyperspectral cube. The leftmost images in the first and third rows show the RGB renderings from the ground truth hyperspectral cube, while those in the second and fourth rows display the ground-truth spectra and reconstructed spectrum of a selected pixel. Spectral Angle Mapper (SAM) maps are shown to the right of the spectral plots. The Hi-SPAD measurements achieve comparable or superior reconstruction quality relative to the deep-learning-based CASSI approach, even when using simple super-pixel rank-1 denoising.

are modest (Hi-SPAD: 28.88 dB / 0.800; CASSI: 28.48 dB / 0.811) because these metrics are dominated by amplitude errors and spatial texture priors.

Figure 11 showcases the two test scenes with the reconstruction results of the 3 different methods. Color images were rendered from the reconstructed hyperspectral cube. As shown in the figure, CMOS-HSI shows the worst performance because, under low light, the signal is dominated by readout noise. As a result, the reconstructed spectra with simple super-pixel rank-1 denoising deviate greatly from the ground truth spectrum. In the rendered color image, patchwise color errors are prominent due to incorrect spectrum estimation. CASSI-RndHRNet provides better reconstruction quality than CMOS-HSI, but still shows a loss of spatial details in the reconstructed hyperspectral cube due to the compressive-sensing nature. Meanwhile, the Hi-SPAD measurement shows comparable or better reconstruction with well-preserved spatial details, even with simple super-pixel rank-1 denoising.

5.2 Material Classification with Hi-SPAD

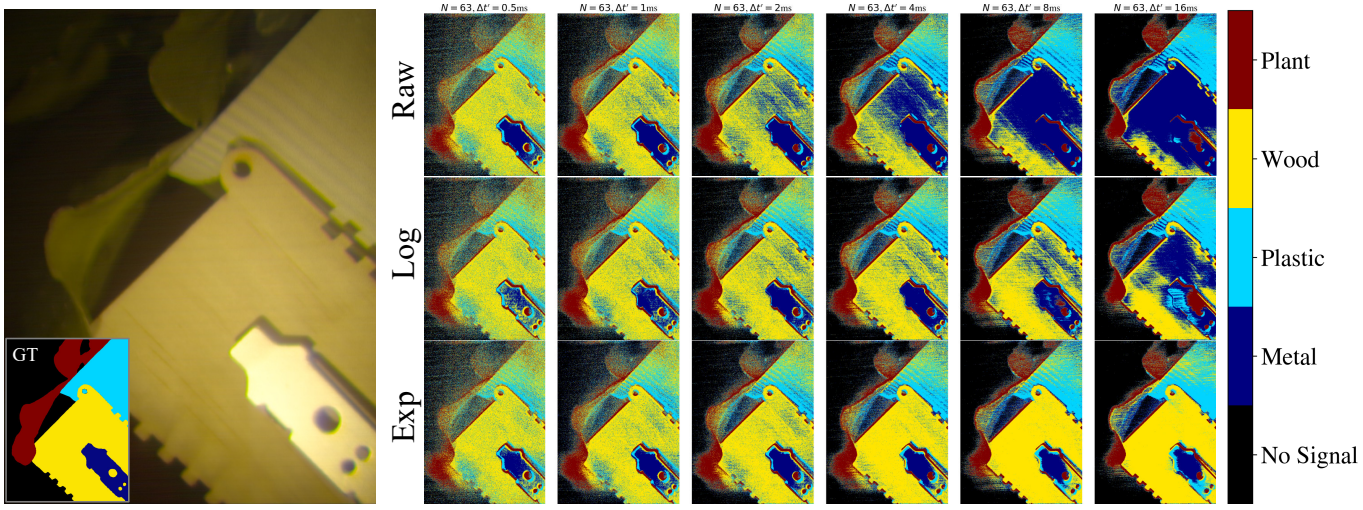
As a first real-world result, we performed classification with real-world data to demonstrate our theory on material classification with SPAD measurements (Section 3.3) and the classification capabilities of the Hi-SPAD prototype. We first collected a spectral dictionary for four common material classes: plant, wood, plastic, and metal. Then, a test scene containing different objects from these categories was captured in 6-bit mode ($N = 63$, since the 6-bit SPAD counter records values from 0 to 63) with increasing exposure time (scene light level). We compare classification results using a cosine classifier under three pairings: (1) raw dictionary and raw measurement, (2) raw dictionary and log-transformed measurement, and (3) exponentially transformed dictionary and raw measurement. Here, "raw measurement" indicates the raw hyperspectral cube measurement without any denoising. The classification maps are presented in Figure 12.

For the raw dictionary and raw measurement, we can notice that the classification accuracy decays as we increase the exposure time (prominently wood and metal in the scene are misclassified). This is due to the increasing non-linearity of the SPAD sensor as the light level increases. The log-transformed measurement shows the better classification accuracy than the raw measurement, but still decays as we increase the exposure time. This is because the log transformation is not able to fully linearize the nonlinearity, especially when the sensor is saturated, which occurs when a SPAD pixel fires for every capture, clipping the value to $N = 63$. Meanwhile, the exponentially transformed dictionary with raw measurements shows the best classification accuracy, especially when the light level is high, with improved performance as exposure time increases. Since the dictionary is nonlinearly transformed according to the scene's light level (Theorem 3.1), which matches the transformation of the true spectrum in the mean of the SPAD measurement (Eq. 8), classification accuracy does not drop as light level increases but instead improves as the variance of SPAD measurements decreases. The results validate the Theorem 3.1 and the classification capabilities of the Hi-SPAD camera.

5.3 Comparison of Spectrum Reconstruction Techniques

We show a comparison of the methods presented in Section 3.4 for true spectrum estimation by applying them to raw HSI data captured with our Hi-SPAD camera prototype. The scene consists of a color checkerboard and a continuously varying bandpass filter (CVBF). CVBF is a specialized bandpass filter with center wavelength that varies continuously along the filter. In Figure 13, the color image reconstructed from the hyperspectral cube is shown on the left, and spectra for selected pixels are shown on the right. The ground truth hyperspectral cube is captured using a large number of frames ($N = 409600$) with a Hi-SPAD camera, with a short exposure time for each binary frame. Under this configuration, the Binomial SPAD measurements converge to a Poisson distribution with rate $N\Phi$, where Φ denotes the photon flux, making the raw Hi-SPAD measurements effectively unbiased estimates of the true spectrum.

The raw Hi-SPAD measurements were captured in 6-bit mode ($N = 63$) with a 1200 fps frame rate of the SPAD sensor, which is equivalent to 0.64 seconds of acquisition time for the whole hyperspectral cube. From the noisy raw Hi-SPAD measurement, we reconstruct the spectrum with the proposed methods in Section 3.4.



(a) Test Scene (b) Classification Maps

Fig. 12. (a) Test scene for material classification with prototype Hi-SPAD camera. RGB image of test scene was reconstructed from captured hyperspectral cube. The scene contains plant, wood, plastic, and metal objects. On the corner of the image, we added the ground truth material label map. (b) Classification maps from cosine classifier with different pairs of dictionary and measurement with increasing exposure time in 6-bit mode ($N = 63$). Each row indicates different pairs of dictionary and measurement: 1) Raw: raw dictionary and raw measurement, 2) Log: raw dictionary and log-transformed measurement, 3) Exp: exponentially transformed dictionary and raw measurement. Due to the non-linear transformation of the dictionary accounting for the light level of the scene, only the Exp pair shows the improved classification accuracy with increasing exposure time.

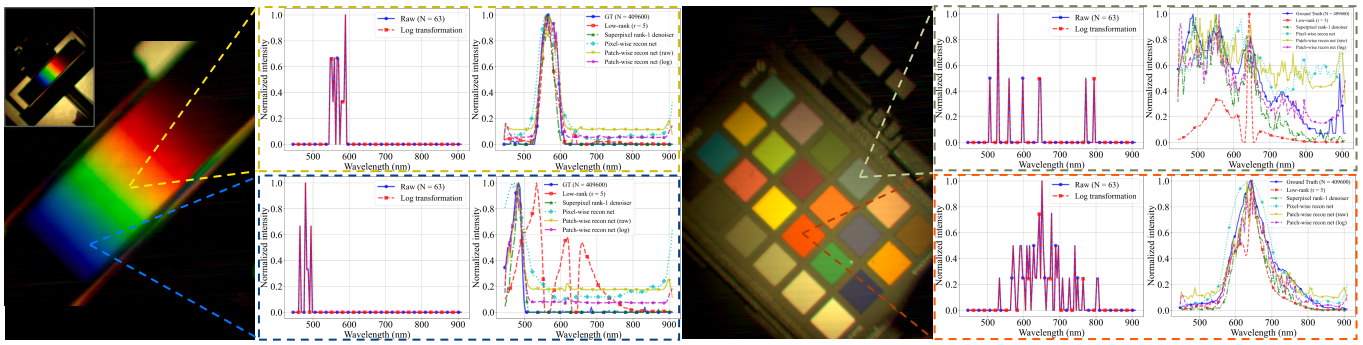


Fig. 13. Lab prototype Hi-SPAD camera measurements for CVBF (left) and color checkerboard (right) scenes and reconstructed spectra from proposed methods. For each scene, color image reconstructed from the captured hyperspectral cube is shown on the left, and the spectra plot of the selected pixel location are shown on the right plots. For CVBF scene, we show magnified view for the better visualization.

The measurement is plotted together with the log-transformed spectrum separately from the reconstructed spectra from other methods. Notably, the two spectra look almost identical. This is because only a few photons are detected in the raw Hi-SPAD measurement and the log transformation is approximately a scaling operation. For patch-wise reconstruction network trained on the ICVL dataset, we exhibit two separate results: "Patch-Wise Reconstruction Network (Raw)" was trained on simulated raw SPAD measurement, while "Patch-Wise Reconstruction Network (Log)" was trained on the log estimated spectra following Equation 28. For both scenes, the network reconstructing the spectrum from the log-estimated spectrum shows better reconstruction quality.

For quantitative evaluation, we computed the SAM images computed with the ground truth hyperspectral cube are depicted in Figure 14, visualizing the reconstruction quality map. Please note that the averaged SAM values are relatively higher than the results in Section 5.1 due to the pixels with extremely low signal under the indoor light condition. From the computed map and averaged SAM values, we conclude that the patch-wise reconstruction network with log-transformed spectrum works the best for the both scenes. Therefore, we employ the method for the following experiments requiring the full HSI reconstruction.

Additionally, to further assess the advantage of Hi-SPAD measurements over CMOS-based line-scanning HSI, we trained the

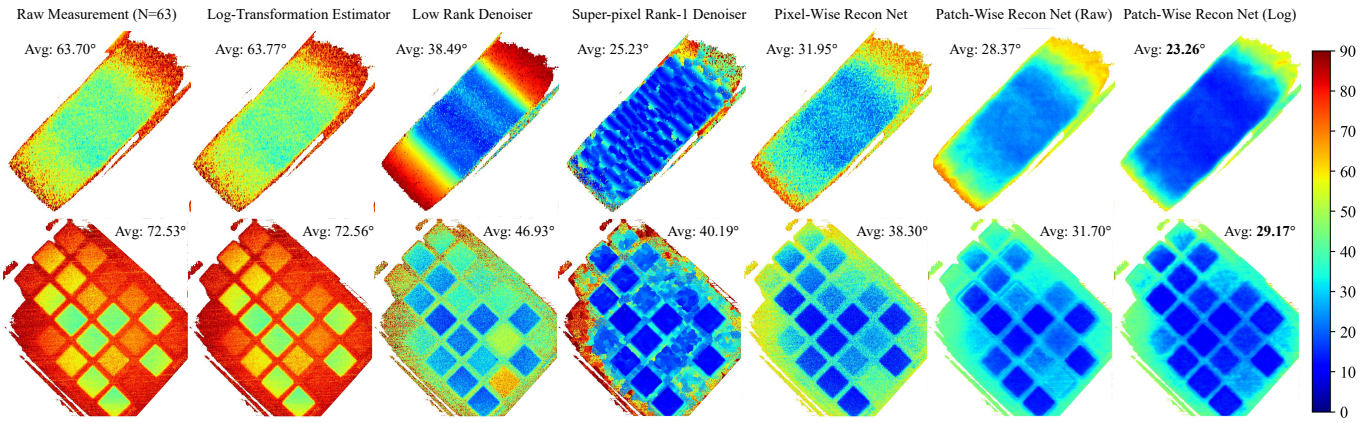


Fig. 14. SAM images computed with ground truth hyperspectral cube and reconstructed hyperspectral cube for the CVBF and color checkerboard scenes. The SAM image is computed for each pixel in the hyperspectral cube and the result is shown as a color image. The pixels where even ground truth cube has extremely low signal are masked out. On the corner of each map, we describe the averaged SAM values for each reconstruction result.

Table 2. Quantitative comparison of reconstruction quality from patch-wise reconstruction network trained on simulated CMOS HSI data and raw Hi-SPAD measurement with log-transformed spectrum. Metrics are computed for 20 randomly selected hyperspectral images from the ICVL dataset (same scenes in Section 5.1) and averaged over the test set. Hi-SPAD produces significantly higher PSNR and SSIM, and lower SAM compared to CMOS-HSI, indicating better reconstruction quality.

Method	Rec. algo.	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓
HSI-CMOS	Patch-Wise	31.73	0.84	7.76
Hi-SPAD	Patch-Wise	**38.84**	**0.92**	**3.98**

patch-wise reconstruction network on simulated CMOS data generated from the ICVL dataset and compared its performance against the same network trained on raw Hi-SPAD measurements with log-transformed spectra. The CMOS-HSI data were simulated as described in Section 3.5.3. Both networks were trained with the same architecture and hyperparameters until convergence. Testing was performed on the same 20 randomly selected hyperspectral images from the ICVL dataset used in Section 5.1.

As shown in Table 2, the network trained with Hi-SPAD measurements achieves significantly better reconstruction quality than its CMOS-based counterpart. While the averaged SAM values are comparable to those obtained with super-pixel rank-1 denoising (Section 5.1), the PSNR and SSIM values are substantially higher. This difference arises because super-pixel rank-1 denoising explicitly enforces a shared spectral direction within local regions, effectively removing orthogonal noise components that dominate the SAM metric. In contrast, the patch-wise reconstruction network is trained with a per-pixel objective and does not explicitly impose region-wise spectral colinearity.

5.4 Hyperspectral Video Results

In this section, we showcase the hyperspectral video results captured with our Hi-SPAD camera. The operational speed of the camera is 26 frames per second, with $N \approx 5$ captured binary frames for each hyperspectral cube. For thorough justification we captured various dynamic scenes under different light conditions, including indoor scenes, outdoor scenes with sunlight, and outdoor scenes with cloudy sky. Figure 15 shows raw and reconstructed hyperspectral video frames from super-pixel rank-1 denoising and patch-wise reconstruction network (with log-transformed measurement). Super-pixel rank-1 denoising is producing piece-wise smooth reconstruction with artifacts in the edges of super-pixels, while the patch-wise reconstruction network is producing still grainy reconstruction. Since the patch-wise reconstruction network produced higher reconstruction accuracy in spectrum-wise in Section 5.3, we employed the patch-wise reconstruction network for the following results showcasing the hyperspectral video results.

With the enabled video-rate hyperspectral imager, we provide results for 3 different applications: 1) Illumination control of captured video, 2) Rendering video from different camera models with different color sensitivity, and 3) Video-rate material classification.

Illumination Control of Captured Video. By capturing the hyperspectral video of the same scene with known light source (sunny day outdoor spectrum), we can compute the spectral reflectance of the scene by dividing the captured hyperspectral video with the known light source spectrum. With the estimated spectral reflectance, then we can post-process the captured video to control the illumination of the scene with different light source coming with different spectrum by re-multiplying the estimated spectral reflectance with the light source spectrum. Figure 16 shows the relighted hyperspectral video frames with numerous light sources with different spectrum; Daylight, sunset, LED, Fluorescent, and candle. The processing was done with the reconstructed hyperspectral video from patch-wise reconstruction network.

Rendering Video of Different Camera Models. Figure 17 shows the rendering video from different camera models with distinct spectral sensitivity for RGB colorization. The rendering video is computed by multiplying the spectral sensitivity function of each camera

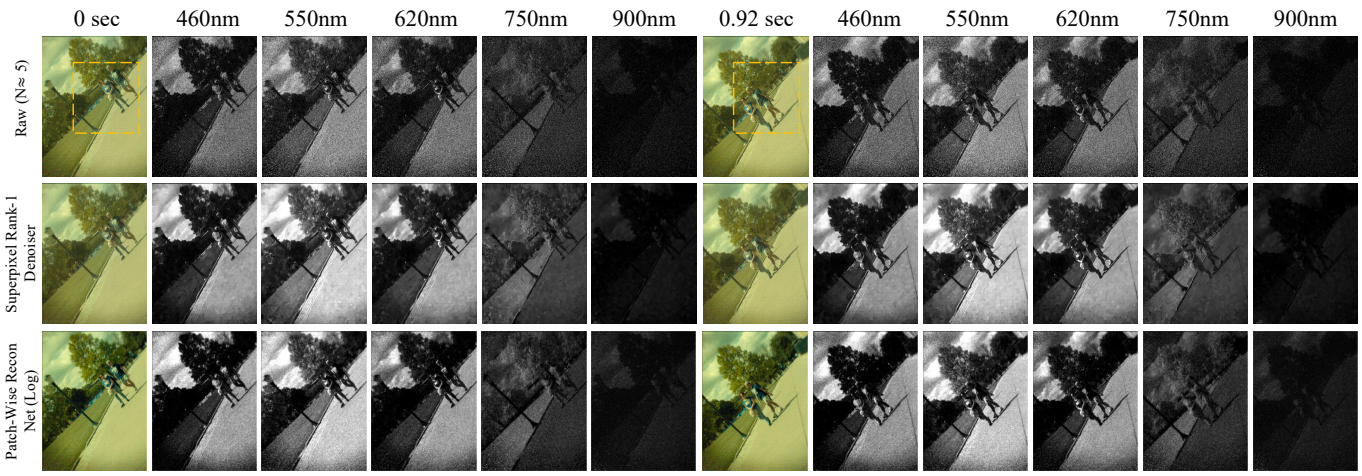


Fig. 15. Hyperspectral video frames captured by the Hi-SPAD camera and reconstructed hyperspectral video from super-pixel rank-1 denoising and patch-wise reconstruction network (with log-transformed spectrum). The video is captured with frame rate of 26 hyperspectral images per second. We exhibit two selected frames from the video and wavelength channels of the reconstructed hyperspectral video.

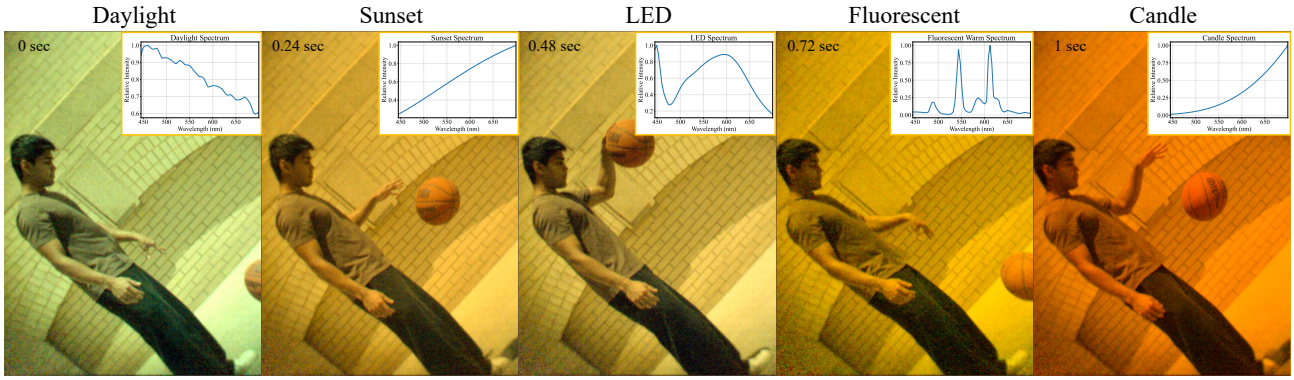


Fig. 16. Relighted HSV video frames from the captured hyperspectral video. We first captured the hyperspectral video of the scene at sunny day outdoor, then computed the spectral reflectance of the scene, and relighted the scene with different light source. The illumination control was done with the reconstructed hyperspectral video from patch-wise reconstruction network.

model in CAMSPEC dataset [Jiang et al. 2013] with the calibrated captured hyperspectral video. Spectral calibration of imaging system was performed by directly illuminating the calibration light source with known spectrum to the camera, capturing the hyperspectral image, and dividing the captured hyperspectral image with the known spectrum to obtain the spectral response of the imager. The rendering was done with the denoised hyperspectral video from patch-wise reconstruction network.

Video-rate Material Classification. We demonstrate the video-rate material classification capabilities of the Hi-SPAD camera. Figure 18 shows the material classification results from the hyperspectral video captured with the Hi-SPAD camera under sunny and cloudy outdoor conditions. The material classification was done with the cosine classifier on raw Hi-SPAD measurement with the exponentially transformed spectral dictionary of the material classes. The spectral dictionary was computed by averaging the spectra of the material classes in the training scene captured separately from the test scene. As shown in the figure, even with the raw Hi-SPAD measurement and simple cosine classifier, we could achieve promising material classification results. Figure 19 shows the confusion matrix of the material classification results for each scene.

6 Discussion

Hi-SPAD offers several advantages over conventional CMOS-based video-rate HSI systems, such as those using snapshot-based compressed sensing or spatially multiplexed spectral encoding methods [Lin et al. 2014; Saragadam et al. 2021; Wagadarikar et al. 2008]. These systems typically sacrifice spatial or spectral resolution to achieve snapshot acquisition. In contrast, Hi-SPAD employs a scanning-based strategy, leveraging SPAD sensors to overcome photon-noise and readout-noise limitations that traditionally constrain pushbroom systems in dynamic or low-light scenes. Despite capturing the full hyperspectral cube through sequential scanning,

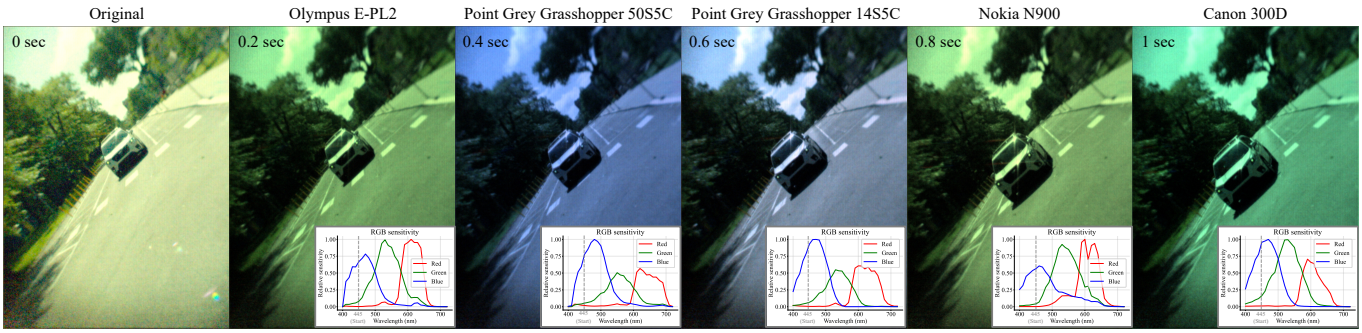


Fig. 17. Video-rate rendering of the different camera models. We calibrated spectral sensitivity of our imaging system first, then computed normalized hyperspectral video by dividing the captured video by the imager's spectral sensitivity function, then rendered the color video with the spectral sensitivity function of the different camera models by multiplying the normalized hyperspectral video with the spectral sensitivity function of the different camera models for RGB channels. The rendering was done with the reconstructed hyperspectral video from patch-wise reconstruction network.

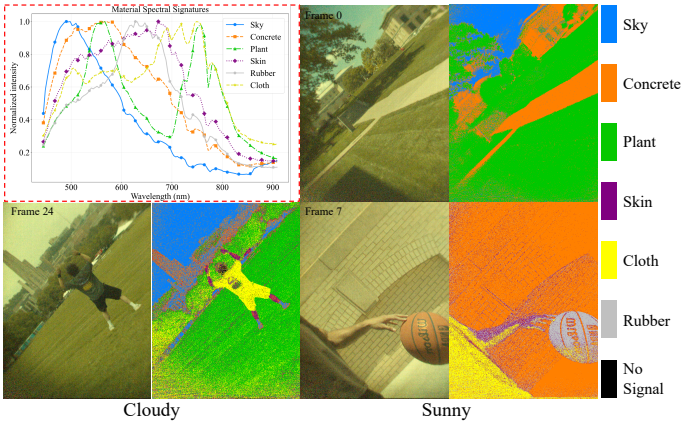


Fig. 18. Video-rate material classification results. Top left figure shows spectral dictionary of the material classes. We captured the hyperspectral video of the scene for both sunny and cloudy day. Classification was done with the cosine classifier on raw Hi-SPAD measurement with the exponentially transformed dictionary.

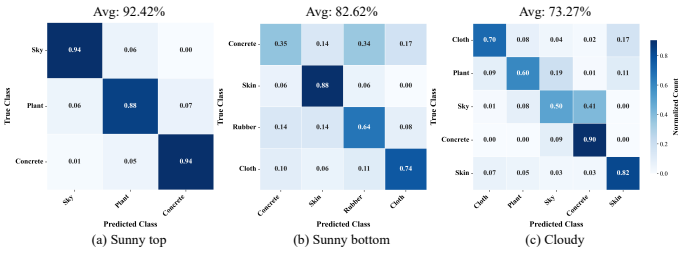


Fig. 19. Confusion matrix of the video-rate material classification results. We computed the confusion matrix by first deriving ground truth label map with manual inspection and comparing with the classification result. On the top of the figure, we show the overall classification accuracy.

our system achieves 26 frames per second across 123 spectral channels, while maintaining high spatial fidelity and signal integrity under ambient lighting conditions.

Moreover, SPAD sensors inherently offer high dynamic range due to their nonlinear photon-response characteristics without a hard saturation limit [Ingle et al. 2019]. This allows Hi-SPAD to operate effectively across a broad spectrum of lighting environments—from extremely low to high photon flux—without requiring exposure adjustments or risking sensor saturation. This robustness is particularly valuable in outdoor environments or industrial settings where lighting can vary rapidly or unpredictably.

Beyond enabling video-rate HSI under low light, Hi-SPAD also establishes a new operational regime for spectral imaging: photon-counting hyperspectral video. This regime is distinct from traditional HSI systems and requires a rethink of reconstruction algorithms and noise models. As shown in our theoretical and empirical evaluations, reconstruction quality can be preserved even with as few as 5 one-bit frames per scan, which opens up possibilities for ultrafast, low-data-volume hyperspectral sensing that is impractical with CMOS alternatives.

Limitations. Despite the advantages of SPAD-based sensing, several practical limitations remain. Current SPAD arrays typically offer significantly fewer pixels than their CMOS counterparts, resulting in lower spatial resolution for captured hyperspectral images. Additionally, the larger pixel pitch and form factor of SPAD sensors require physically larger sensor areas to achieve comparable spectral coverage in line-scanning configurations. These hardware constraints, combined with the relatively high cost of SPAD devices, limit their immediate scalability for widespread adoption. However, given the rapid pace of development in SPAD technology [Morimoto et al. 2021; Mos et al. 2024; Ogi et al. 2021], we expect these limitations to be mitigated in future generations of sensors.

From an HSI system design perspective, several trade-offs emerge. First, because our system relies on line scanning with an effective acquisition time of approximately 38 ms per hyperspectral frame, it inherently suffers from rolling-shutter effects. This becomes especially problematic when capturing rapid dynamic scenes, as motion during scanning can introduce geometric distortions in the reconstructed hyperspectral frames. Given our system is scanning single-pixel-width column by column, the time per pixel column is $1/(26\,\text{fps} \times 768\,\text{pixels}) \approx 0.05\,\text{ms}$; the object speed (projected to

the image plane) must stay below the inverse of this, ≈ 20 pixel/ms, to avoid rolling-shutter artifacts. The rolling-shutter artifacts are visible in our real-world results: Figure 20 shows the noticeable distortion of the bouncing basketball and rotating fan. Importantly, while rolling shutter primarily induces spatial misregistration across scan lines, the spectral measurements within each line remain locally valid; thus, motion mainly affects geometric consistency rather than per-pixel spectral integrity.

Because our system has full controllability over both the scanning order and the per-column exposure time, rolling-shutter artifacts can be actively mitigated through programmable scan scheduling. Instead of scanning columns strictly sequentially from left to right, we can adopt an interlaced scanning strategy; e.g., first acquire every K-th column (for example, all even columns), and then acquire the remaining columns (odd columns). This effectively partitions a single hyperspectral frame into K temporally interleaved sub-frames, each with reduced temporal span. As shown in the coded rolling shutter framework of Gu et al. [2010], these sub-frames can be horizontally interpolated to full resolution and used to estimate optical flow between them, enabling interpolation of intermediate views and reconstruction of a artifact-free image.

Beyond artifact mitigation, the pushbroom flexibility of Hi-SPAD can be further exploited to enable additional functionalities. As shown in the coded rolling shutter framework of Gu et al. [2010], staggered column acquisition can increase effective temporal resolution for capturing fast dynamic events, while programmable per-column exposure times allow interlacing short and long exposures within a single hyperspectral frame to enable HDR imaging or motion-blur-aware reconstruction. Therefore, although rolling shutter introduces geometric distortion under rapid motion, the same programmable scanning mechanism that causes this effect can be repurposed as a computational imaging tool to improve temporal resolution, dynamic range, and motion robustness in Hi-SPAD. We believe this is a promising direction for future research.

Another limitation stems from our use of a diffraction grating (within a grism) to spectrally disperse the light. We currently extract only the first-order dispersion for reconstruction since it is the most dominant order and can be captured for whole columns with translation in the scanning direction. However, as shown in Figure 7, higher-order dispersions (e.g., second order or beyond) are clearly present in the captured images, especially at shorter wavelengths and for early scan columns. Leveraging these additional diffraction orders through careful calibration and joint reconstruction could improve light throughput and overall signal efficiency for the early columns. For the prototype we chose a grism with the groove density that maximizes the dispersion angle for the highest spectral resolution, but by using a smaller groove density—compromising spectral resolution—we could capture higher-order dispersion for the early columns. Employing specialized gratings, such as volume-phase holographic gratings [Barden et al. 1998] that concentrate light throughput into the first-order dispersion, could be a promising direction to improve light throughput and spectral resolution.

The other limitation arises from the stochastic nature of photon-counting measurements. In regions that are extremely bright or dark, SPAD measurements can become uninformative, manifesting as saturation to all ones in extremely bright regions, or as all-zero

(a) Dynamic scene HSVs (b) Static state

Fig. 20. Rolling-shutter artifacts in the reconstructed hyperspectral videos. (a) Selected RGB frames from HSVs capturing moving objects with fast motion. (b) The objects in static state. The artifacts are visible in the reconstructed RGB videos as distortions of the basketball and fan in the scenes.

(photon-starved) measurements in extremely dark regions. In our prototype, the saturation issue was largely avoided in practice because the combination of high spectral resolution and video-rate scanning places the system in a photon-limited operating regime. As a result, even under outdoor ambient lighting conditions, the majority of SPAD measurements did not saturate to all ones with the maximum exposure time of 10 µs per binary frame which ensures the sensor's maximum achievable frame rate of 100 kHz. In more extreme high-dynamic-range scenarios, however, saturation and photon-starvation may occur simultaneously. Such cases can be addressed by reducing the exposure time per binary frame to avoid all-ones measurements in bright regions, and compensating for reduced photon counts in dark regions by increasing the number of SPAD frames N. This trade-off effectively shifts acquisition time from per-frame exposure to temporal accumulation, at the cost of reduced hyperspectral video frame rate. Exploring adaptive strategies such as coded rolling shutter that jointly adjust exposure time and N for each column based on scene content is a promising direction for extending Hi-SPAD to challenging HDR scenes.

Stray Light. Since capturing full hyperspectral video typically occurs in the low-light regime, where only a few photons are detected per dimension, the optical system must be carefully designed to minimize stray light that does not originate from the scene. Although we optically shielded the prototype and applied background subtraction to remove stray-light artifacts, the reconstructed hyperspectral video still contains some residual artifacts. To fully address this challenge, future improvements will require a more complete optical enclosure as well as advanced post-processing methods, such as learning-based denoising techniques, to further suppress stray light after background subtraction.

Real-Time Reconstruction. Moreover, reconstructing full hyperspectral cubes from line-scanning systems is computationally intensive, presenting challenges for real-time applications. Potential solutions include integrating real-time reconstruction capabilities directly onto SPAD sensors, similar to software-defined SPAD camera approaches [Sundar et al. 2023]. This approach could significantly enhance practicality and real-time performance. For example, along the lines of Sundar et al. [2023], we can implement a "software-defined spectral camera" that implements certain desired linear projections on-chip to dramatically reduce readout bandwidth of the device. When coupled with sophisticated learning-based approaches to reconstruct HSIs or infer directly from these projections, we can benefit from the advantages that SPAD sensors provide without paying a significant penalty in readout bandwidth.

Acknowledgments

This work was financed in part by a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development. Lee was also supported in part by a Dean's Fellowship and a TCS Presidential fellowship. Gupta was supported by the National Science Foundation (CAREER Award #1943149), the Office of Naval Research (N000142412155), and a WARF Research Forward Initiative Award.

References

2025. Phantom High-Speed CMOS Camera TMX 7510 Model. https://www.phantomhighspeed.com/products/cameras/tmx/7510. Specification sheet accessed on February 19, 2025.

Muhammad Ahmad and Manuel Mazzara. 2024. Scsnet: Sharpened cosine similarity-based neural network for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 21 (2024), 1–4.

Reza Arablouei, Ethan Goan, Stephen Gensemer, and Branislav Kusy. 2016. Fast and robust pushbroom hyperspectral imaging via DMD-based scanning. In *Novel Optical Systems Design and Optimization XIX*, Vol. 9948. SPIE, 7–17.

Boaz Arad and Ohad Ben-Shahar. 2016. Sparse Recovery of Hyperspectral Signal from Natural RGB Images. In *European Conference on Computer Vision*. Springer, 19–34.

M. Kerem Aydin, Yi-Chun Hung, Jaclyn Pytlarz, Qi Guo, and Emma Alexander. 2025. Spectrum from Defocus: Fast Spectral Imaging with Chromatic Focal Stack. arXiv:2503.20184 [cs.CV] https://arxiv.org/abs/2503.20184

Samuel Charles Barden, James A Arns, and Willis S Colburn. 1998. Volume-phase holographic gratings and their potential for astronomical applications. In *Optical Astronomical Instrumentation*, Vol. 3355. SPIE, 866–876.

Jerome Baudot, W Dulinski, M Winter, R Barbier, E Chabanat, P Depasse, and N Estre. 2009. Photon detection with CMOS sensors for fast imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 604, 1-2 (2009), 111–114.

IS Bowen and AH Vaughan. 1973. " Nonobjective" Gratings. *Publications of the Astronomical Society of the Pacific* 85, 504 (1973), 174.

Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B Hullin. 2021. Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.

Yurong Chen, Hui Zhang, Yaonan Wang, Yimin Yang, and Jonathan Wu. 2024. FlexDLD: Deep Low-Rank Decomposition Model With Flexible Priors for Hyperspectral Image Denoising and Restoration. *IEEE Transactions on Image Processing* 33 (2024), 1211–1226. doi:10.1109/TIP.2024.3360902

CC Cheung and M Christophersen. 2023. Mosaic-pixelated Fabry-Pérot filters for snapshot spectral imagers. In *Imaging Spectrometry XXVI: Applications, Sensors, and Processing*, Vol. 12688. SPIE, 54–61.

Deglr6328. 2005. Fluorescent lighting spectrum with emission peaks numbered. https://commons.wikimedia.org/wiki/File:Fluorescent_lighting_spectrum_peaks_labelled.gif. Original spectrum measured by the author using an Ocean Optics HR2000 spectrometer; image hosted on Wikimedia Commons. Accessed: 2026-01-25.

Ahmet T Erdogan, Richard Walker, Neil Finlayson, Nikola Krstajić, Gareth Williams, John Girkin, and Robert Henderson. 2019. A CMOS SPAD line sensor with per-pixel histogramming TDC for time-resolved multispectral imaging. *IEEE Journal of Solid-State Circuits* 54, 6 (2019), 1705–1719.

Zackory M. Erickson, Nathan Luskey, S. Chernova, and C. Kemp. 2018. Classification of Household Materials via Spectroscopy. *IEEE Robotics and Automation Letters* 4 (2018), 700–707. doi:10.1109/LRA.2019.2892593

Huixin Fan, Jie Li, Qiangqiang Yuan, Xinxin Liu, and M. Ng. 2019. Hyperspectral image denoising with bilinear low rank matrix factorization. *Signal Process.* 163 (2019), 132–152. doi:10.1016/J.SIGPRO.2019.04.029

Abdulrahman Galal, Hesham Hassan, and Ibrahim F Imam. 2012. A novel approach for measuring hyperspectral similarity. *Applied Soft Computing* 12, 10 (2012), 3115–3123.

Abhiram Gnanasambandam, Omar Elgendy, Jiaju Ma, and Stanley H Chan. 2019. Megapixel photon-counting color imaging using quanta image sensor. *Optics express* 27, 12 (2019), 17298–17310.

Alexander D Griffiths, Haochang Chen, Johannes Herrnsdorf, David Li, Robert K Henderson, Michael J Strain, and Martin D Dawson. 2018. Hyperspectral imaging under low illumination with a single photon camera. In *2018 IEEE British and Irish Conference on Optics and Photonics (BICOP)*. IEEE, 1–4.

Jinwei Gu, Yasunobu Hitomi, Tomoo Mitsunaga, and Shree Nayar. 2010. Coded rolling shutter photography: Flexible space-time sampling. In *2010 IEEE International Conference on Computational Photography (ICCP)*. 1–8. doi:10.1109/ICCPHOT.2010.5585094

Anant Gupta, Atul Ingle, and Mohit Gupta. 2019. Asynchronous single-photon 3D imaging. In *Proceedings of the IEEE/CVF international conference on computer vision*. 7909–7918.

Nathan Hagen and Michael W Kudenov. 2013. Review of snapshot spectral imaging technologies. *Optical Engineering* 52, 9 (2013), 090901–090901.

Samuel W Hasinoff, Frédo Durand, and William T Freeman. 2010. Noise-optimal capture for high dynamic range photography. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 553–560.

E Keith Hege, Dan O'Connell, William Johnson, Shridhar Basty, and Eustace L Dereniak. 2004. Hyperspectral imaging for astronomy and space surveillance. In *Imaging Spectrometry IX*, Vol. 5159. SPIE, 380–391.

Rebecca Ilehag, A. Schenk, Yilin Huang, and S. Hinz. 2019. KLUM: An Urban VNIR and SWIR Spectral Library Consisting of Building Materials. *Remote. Sens.* 11 (2019), 2149. doi:10.3390/rs11182149

Atul Ingle, Andreas Velten, and Mohit Gupta. 2019. High flux passive imaging with single-photon sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6760–6769.

Daniel S. Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. 2019. Compact Snapshot Hyperspectral Imaging with Diffracted Rotation. *ACM Transactions on Graphics (Proc. SIGGRAPH 2019)* 38, 4 (2019), 117:1–13. doi:10.1145/3306346.3322946

Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. 2013. What is the space of spectral sensitivity functions for digital color cameras?. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 168–179.

Raymond F Kokaly, Roger N Clark, Gregg A Swayze, K Eric Livo, Todd M Hoefen, Neil C Pearson, Richard A Wise, William Benzel, Heather A Lowers, Rhonda L Driscoll, et al. 2017. *USGS spectral library version 7*. Technical Report. US Geological Survey.

Nikola Krstajić, James Levitt, Simon Poland, Simon Ameer-Beg, and Robert Henderson. 2015. 256× 2 SPAD line sensor for time resolved fluorescence spectroscopy. *Optics express* 23, 5 (2015), 5653–5669.

Andras Kufcsák, Ahmet Erdogan, Richard Walker, Katjana Ehrlich, M Tanner, Alicia Megia-Fernandez, Emma Scholefield, Philip Emanuel, Kanwaldeep Dhaliwal, Mark Bradley, et al. 2017. Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications. *Optics express* 25, 10 (2017), 11103–11123.

Junichi Kurihara, Tetsuro Ishida, and Yukihiro Takahashi. 2020. Unmanned Aerial Vehicle (UAV)-based hyperspectral imaging system for precision agriculture and forest management. *Unmanned Aerial Vehicle: Applications in Agriculture and Environment* (2020), 25–38.

Zeqiang Lai, Chenggang Yan, and Ying Fu. 2023. Hybrid spectral denoising transformer with guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13065–13075.

Haejoon Lee and Aswin C Sankaranarayanan. 2024. Spectral Subsurface Scattering for Material Classification. In *European Conference on Computer Vision*. Springer, 108–124.

Jongho Lee, Atul Ingle, Jenu V Chacko, Kevin W Eliceiri, and Mohit Gupta. 2023. CASPI: collaborative photon processing for active single-photon imaging. *Nature Communications* 14, 1 (2023), 3158.

Hoong-Ta Lim and Vadakke Matham Murukeshan. 2015. Pushbroom hyperspectral imaging system with selectable region of interest for medical imaging. *Journal of biomedical optics* 20, 4 (2015), 046010–046010.

Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. 2014. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 1–11.

Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Brushini, Edoardo Charbon, and Mohit Gupta. 2020. Quanta Burst Photography. *ACM Transactions on Graphics (TOG)* 39 (7 2020). doi:10.1145/3386569.3392470

Sizhuo Ma, Varun Sundar, Paul Mos, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. 2023. Seeing photons in color. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–16.

Hanning Mai, Anneliese L. Jarman, A. Erdogan, Conor Treacy, N. Finlayson, R. Henderson, and S. Poland. 2023. Development of a high-speed line-scanning fluorescence lifetime imaging microscope for biological imaging. *Optics letters* 48 8 (2023), 2042–2045. doi:10.1364/ol.482403

Yuki Maruyama, Jordana Blacksberg, and Edoardo Charbon. 2013. A 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS. *IEEE Journal of Solid-State Circuits* 49, 1 (2013), 179–189.

Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and L. Waller. 2020. Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array. *ArXiv* abs/2006.08565 (2020). doi:10.1364/optica.397214

Kazuhiro Morimoto, J Iwata, M Shinohara, H Sekine, A Abdelghafar, H Tsuchiya, Y Kuroda, K Tojima, W Endo, Y Maehashi, et al. 2021. 3.2 megapixel 3D-stacked charge focusing SPAD for low-light imaging and depth sensing. In *2021 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 20–2.

Paul Mos, Michael A Wayne, Andrei Ardelean, Arin C Ulku, Edoardo Charbon, and Claudio Bruschini. 2024. SwissSPAD2/3: a family of natively digital, time gated SPAD cameras with continuous streaming at up to 100 kpfs and picosecond system-level synchronization for quantum imaging applications. In *Quantum Sensing, Imaging, and Precision Metrology II*, Vol. 12912. SPIE, 189–196.

Cristiano Niclass, Claudio Favi, Theo Kluter, Marek Gersbach, and Edoardo Charbon. 2008. A 128 × 128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array. *IEEE Journal of Solid-State Circuits* 43, 12 (2008), 2977–2989. doi:10.1109/JSSC.2008.2006445

Cristiano Niclass, Alexis Rochas, P-A Besse, and Edoardo Charbon. 2005. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits* 40, 9 (2005), 1847–1854.

Jun Ogi, Takafumi Takatsuka, Kazuki Hizu, Yutaka Inaoka, Hongbo Zhu, Yasuhisa Tochigi, Yoshiaki Tashiro, Fumiaki Sano, Yusuke Murakawa, Makoto Nakamura, et al. 2021. A 124-dB dynamic-range SPAD photon-counting image sensor using subframe sampling and extrapolating photon count. *IEEE Journal of Solid-State Circuits* 56, 11 (2021), 3220–3227.

Oxford Instruments Andor. 2023. *Sensitivity and Noise of CCD, EMCCD and sCMOS Sensors*. https://andor.oxinst.com/learning/view/article/sensitivity-and-noise-of-ccd-emccd-and-scmos-sensors Accessed: 2025-08-17.

Xuanyu Qian, Wei Jiang, Ahmed Elsharabasy, and M Jamal Deen. 2023. Modeling for single-photon avalanche diodes: State-of-the-art and research challenges. *Sensors* 23, 7 (2023), 3412.

Vishwanath Saragadam, Michael DeZeeuw, Richard G Baraniuk, Ashok Veeraraghavan, and Aswin C Sankaranarayanan. 2021. SASSI—super-pixelated adaptive spatio-spectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 7 (2021), 2233–2244.

Theodor Scheimpflug. 1904. Improved method and apparatus for the systematic alteration or distortion of plane pictures and images by means of lenses and mirrors for photography and for other purposes. *GB patent* 1196 (1904), 12.

Suhyun Shin, Seungwoo Yoon, Ryota Maeda, and Seung-Hwan Baek. 2025. Dense Dispersed Structured Light for Hyperspectral 3D Imaging of Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Varun Sundar, Andrei Ardelean, Tristan Swedish, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. 2023. Sodacam: Software-defined cameras via single-photon imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8165–8176.

Toshihiro Takamatsu, Ryodai Fukushima, Kounosuke Sato, Masakazu Umezawa, Hideo Yokota, Kohei Soga, Abian Hernandez-Guedes, Gustavo M Callico, and Hiroshi Takemura. 2024. Development of a visible to 1600 nm hyperspectral imaging rigid-scope system using supercontinuum light and an acousto-optic tunable filter. *Optics Express* 32, 9 (2024), 16090–16102.

Alessandro Tontini, L. Gasparini, and M. Perenzoni. 2020. Numerical Model of SPAD-Based Direct Time-of-Flight Flash LIDAR CMOS Image Sensors. *Sensors (Basel, Switzerland)* 20 (2020). doi:10.3390/s20185203

Darren Turner, Arko Lucieer, Matthew McCabe, Stephen Parkes, and I Clarke. 2017. Pushbroom hyperspectral imaging from an unmanned aircraft system (uas)–geometric processingworkflow and accuracy assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2017), 379–384.

Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolović, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. 2018. A 512× 512 SPAD image sensor with integrated gating for widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics* 25, 1 (2018), 1–12.

Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. 2008. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics* 47, 10 (2008), B44–B51.

Pengchong Wang and Zhonghua Zhang. 2016. Double-filtering method based on two acousto-optic tunable filters for hyperspectral imaging application. *Optics Express* 24, 9 (2016), 9888–9895.

Lisha Zhang, Jingjing Ge, Xiaoming Zhong, Fang Xue, and Yanli Liu. 2023. Snapshot spectral imaging based on SPAD and CS for space targets. In *Sensors, Systems, and Next-Generation Satellites XXVII*, Vol. 12729. SPIE, 301–307.

Xuanyu Zhang, Bin Chen, Wenzhen Zou, Shuai Liu, Yongbing Zhang, Ruiqin Xiong, and Jian Zhang. 2024. Progressive content-aware coded hyperspectral snapshot compressive imaging. *IEEE Transactions on Circuits and Systems for Video Technology* (2024).

Lina Zhuang, Xiyou Fu, Michael K. Ng, and J. Bioucas-Dias. 2021. Hyperspectral Image Denoising Based on Global and Nonlocal Low-Rank Factorizations. *IEEE Transactions on Geoscience and Remote Sensing* 59 (2021), 10438–10454. doi:10.1109/TGRS.2020.3046038

Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitany, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini, et al. 2020. Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation. *Scientific Reports* 10, 1 (2020), 20986.

Sheng Zou, Paul Gader, and Alina Zare. 2019. Hyperspectral tree crown classification using the multiple instance adaptive cosine estimator. *PeerJ* 7 (2019), e6405.

A Proof of the Error Bounds

A.1 Material Classification

A.1.1 Material Classification with Unaltered Spectrum Dictionary.

Preliminaries. We have a dictionary of M reference spectra corresponding to M materials. That is,

$$\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_M], \tag{36}$$

with

$$\mathbf{d}_i \in \mathbb{R}^L, \quad L : \text{number of spectral channels}, \tag{37}$$

$$\mathbf{d}_i[n] \geq 0, \quad n \in \{1, \ldots, L\}, \tag{38}$$

and

$$\|\mathbf{d}_i\| = 1. \tag{39}$$

Measurement Notation. Let

$$\mathbf{y} \in \mathbb{R}^L. \tag{40}$$

We model the measurement at each spectral channel n as

$$\mathbf{y}[n] \sim \text{Binomial}\Big(N,\ 1 - e^{-\Phi_0\, \mathbf{d}_k[n]}\Big), \tag{41}$$

where

- N is the number of frames captured with the SPAD sensor,
- $\Delta t'$ is the exposure time per frame,
- $\mathbf{d}_k$ is the normalized spectrum of the k-th material at the point,
- Φ is a constant representing the overall light level (unit: photon/sec),
- and we define

$$\Phi_0 = \Phi\, \Delta t'. \tag{42}$$

Classifier. We use cosine similarity as a classifier. In particular, a measurement vector $\mathbf{y}$ is assigned to the l-th material if

$$l = \arg\max_j \langle \mathbf{y}, \mathbf{d}_j \rangle. \tag{43}$$

Correct classification occurs if

$$\langle \mathbf{y}, \mathbf{d}_k \rangle > \langle \mathbf{y}, \mathbf{d}_j \rangle \quad \text{for all } j \neq k. \tag{44}$$

Defining

$$s_j = \langle \mathbf{y}, \mathbf{d}_j \rangle \quad \text{and} \quad e_j = S_k - S_j, \tag{45}$$

misclassification occurs when $e_j < 0$. Our goal is to bound $P(e_j \leq 0)$.

Chernoff Bounds. If we use a Chernoff bound, then

$$P(e_j \leq 0) \leq \mathcal{M}(s) e^{-s\cdot 0}, \quad \text{for } s < 0, \tag{46}$$

so that, since this bound holds for every negative s,

$$P(e_j \leq 0) \leq \inf_{s<0} \mathcal{M}(s), \tag{47}$$

where the moment–generating function (MGF) is

$$\mathcal{M}(s) = \mathbb{E}\Big(e^{s\, e_j}\Big). \tag{48}$$

Finding the MGF. Recall that

$$S_j = \langle \mathbf{y}, \mathbf{d}_j \rangle = \sum_n \mathbf{y}[n]\, \mathbf{d}_j[n], \tag{49}$$

and thus

$$e_j = S_k - S_j = \sum_n \mathbf{y}[n] \Big(\mathbf{d}_k[n] - \mathbf{d}_j[n]\Big). \tag{50}$$

Defining

$$e_{j,n} = \mathbf{y}[n] \Big(\mathbf{d}_k[n] - \mathbf{d}_j[n]\Big), \tag{51}$$

the MGF factorizes as

$$\mathcal{M}(s) = \prod_n \mathbb{E}\Big(e^{s\, e_{j,n}}\Big) = \prod_n \mathcal{M}_n(s), \tag{52}$$

where, since

$$\mathbf{y}[n] \sim \text{Binomial}\{N, P_n\} \quad \text{with} \quad P_n = 1 - e^{-\Phi_0\, \mathbf{d}_k[n]}, \tag{53}$$

and letting

$$a_n = \mathbf{d}_k[n] - \mathbf{d}_j[n], \tag{54}$$

the MGF for each channel is

$$\mathcal{M}_n(s) = \mathbb{E}\Big(e^{s\, a_n\, \mathbf{y}[n]}\Big) = (1 - P_n + P_n\, e^{s\, a_n})^N. \tag{55}$$

Thus,

$$\mathcal{M}(s) = \prod_n (1 - P_n + P_n\, e^{s\, a_n})^N. \tag{56}$$

This expression can be equivalently written as

$$\mathcal{M}(s) = \left\{ \prod_n \Big[1 - P_n + P_n\, e^{s\, a_n} \Big] \right\}^N, \tag{57}$$

or, writing it in more explicit form,

$$\begin{aligned} \mathcal{M}(s) = \Bigg\{ \prod_n \Bigg[1 - \Big(1 - e^{-\Phi_0\, \mathbf{d}_k[n]}\Big) \\ + \Big(1 - e^{-\Phi_0\, \mathbf{d}_k[n]}\Big)\, e^{s\, \left(\mathbf{d}_k[n] - \mathbf{d}_j[n]\right)} \Bigg] \Bigg\}^N. \end{aligned} \tag{58}$$

For convenience, we now introduce the following definitions:

$$\alpha_n \triangleq \mathbf{d}_k[n], \quad \beta_n \triangleq \mathbf{d}_j[n], \tag{59}$$

$$\boldsymbol{\alpha},\ \boldsymbol{\beta} \in \mathbb{R}^L, \quad \|\boldsymbol{\alpha}\| = \|\boldsymbol{\beta}\| = 1, \quad \alpha_n, \beta_n \geq 0 \quad \forall\, n. \tag{60}$$

Defining the function

$$g(s) \ = \ \prod_n \Big(e^{-\Phi_0\, \alpha_n} \ + \ (1 - e^{-\Phi_0\, \alpha_n})\, e^{s\, (\alpha_n - \beta_n)} \Big), \tag{61}$$

we can express the overall MGF as

$$M(s) \ = \ \big[g(s) \big]^N. \tag{62}$$

Our goal is to show

$$g(s) \ < \ 1 \quad \text{for some } s < 0, \tag{63}$$

noting that

$$g(s) \ > \ 0 \quad \text{for all } s < 0. \tag{64}$$

Proof Regarding $g(s)$. Consider the natural logarithm of $g(s)$. Define

$$\begin{aligned} h(s) &= \ln\big[g(s)\big] \\ &= \sum_n \ln\Big(e^{-\Phi_0\,\alpha_n} \;+\; \big[1-e^{-\Phi_0\,\alpha_n}\big]\,e^{s\,(\alpha_n-\beta_n)}\Big). \end{aligned} \tag{65}$$

At $s=0$,

$$\begin{aligned} h(0) &= \sum_n \ln\Big(e^{-\Phi_0\,\alpha_n} + \big[1-e^{-\Phi_0\,\alpha_n}\big]\Big) \\ &= \sum_n \ln(1) \\ &= 0. \end{aligned} \tag{66}$$

Next, we differentiate $h(s)$ with respect to s:

$$h'(s) \;=\; \sum_n \frac{(\alpha_n-\beta_n)\,\big[1-e^{-\Phi_0\,\alpha_n}\big]\,e^{s\,(\alpha_n-\beta_n)}}{e^{-\Phi_0\,\alpha_n}+\big[1-e^{-\Phi_0\,\alpha_n}\big]\,e^{s\,(\alpha_n-\beta_n)}}. \tag{67}$$

In particular, at $s=0$,

$$h'(0) \;=\; \sum_n (\alpha_n-\beta_n)\,\big[1-e^{-\Phi_0\,\alpha_n}\big]. \tag{68}$$

Case 1: Φ_0 is small. When Φ_0 is small, we use the first–order approximation

$$1-e^{-\Phi_0\,\alpha_n} \;\approx\; \Phi_0\,\alpha_n. \tag{69}$$

Then,

$$\begin{aligned} h'(0) &\approx \sum_n \Phi_0\,\alpha_n\,(\alpha_n-\beta_n) \\ &= \Phi_0\Big(\sum_n \alpha_n^2 - \sum_n \alpha_n\,\beta_n\Big) \\ &= \Phi_0\,\Big(1-\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle\Big) \\ &> 0. \end{aligned} \tag{70}$$

Since $h(0)=0$ and $h'(0)>0$, $h(s)$ is positive for small positive s. By continuity, there must exist some negative s for which $h(s)<0$. Therefore,

$$g(s) \;=\; e^{h(s)} \;<\; 1 \quad \text{for some } s<0, \tag{71}$$

leading to an exponential decrease in the misclassification bound with increasing N.

Case 2: Φ_0 is large. When Φ_0 is large, the factor $1-e^{-\Phi_0\,\alpha_n}$ is close to 1. Then

$$h'(0) \;\approx\; \sum_n (\alpha_n-\beta_n). \tag{72}$$

Here, the sign of $h'(0)$ depends on the specific vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, in this case, we couldn't guarantee that $h'(0)>0$, and the proof of $g(s)<1$ for some $s>0$ is not guaranteed. As a result, the misclassification bound may not decrease exponentially with increasing N when Φ_0 is large.

A.1.2 Classification with Cosine on Exponential Transformed **D**. We now define an alternative classifier based on an exponential transform. Define the measurement

$$y[n] \sim \text{Binomial}\Big(\mathcal{N},\; 1-e^{-\Phi_0\,d_k[n]}\Big), \tag{73}$$

with

$$\Phi_0 = \Phi\,\Delta t', \tag{74}$$

and consider the spectrum dictionary

$$\mathbf{D} = \{\mathbf{d}_1,\ldots,\mathbf{d}_M\}, \quad \|\mathbf{d}_i\| = 1 \quad \forall\, i. \tag{75}$$

We then form the exponential transform

$$\mathbf{P} = \{\mathbf{p}_1,\ldots,\mathbf{p}_M\}, \tag{76}$$

where

$$\mathbf{p}_i = \frac{1-e^{-\Phi_0\,\mathbf{d}_i}}{\|1-e^{-\Phi_0\,\mathbf{d}_i}\|} \in \mathbb{R}^L, \tag{77}$$

so that

$$\|\mathbf{p}_i\| = 1 \quad \forall\, i. \tag{78}$$

The measurement can then be rewritten as

$$\mathbf{y} \sim \text{Binomial}\Big(\mathcal{N},\; \alpha_k\,\mathbf{p}_k\Big), \tag{79}$$

where

$$\alpha_k = \|1-e^{-\Phi_0\,\mathbf{d}_k}\|, \tag{80}$$

with $0 \le \alpha_k \le L$ (for large Φ_0). For classification we compute

$$S_j = \langle \mathbf{y}, \mathbf{p}_j\rangle, \tag{81}$$

and the misclassification event is given by

$$\Pr\Big\{e_j = S_k - S_j = \langle \mathbf{y}, \mathbf{p}_k - \mathbf{p}_j\rangle < 0\Big\}. \tag{82}$$

Chernoff Bound. By the Chernoff bound,

$$\Pr\{e_j<0\} \le M(s), \quad s<0, \tag{83}$$

where

$$M(s) = \mathbb{E}\Big(e^{s\,e_j}\Big) = \prod_n \mathbb{E}\Big(e^{s\,e_{jn}}\Big). \tag{84}$$

In our case, one can write (as we did in the previous section)

$$M(s) = \left\{\prod_n \Big[1-\alpha_k\,\mathbf{p}_k[n]+\alpha_k\,\mathbf{p}_k[n]\,e^{s\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])}\Big]\right\}^{\mathcal{N}}. \tag{85}$$

Proof Requirements and Bounding $f(s)$. We need to show that the function

$$f(s) = \prod_n \Big(1-\alpha_k\,\mathbf{p}_k[n]+\alpha_k\,\mathbf{p}_k[n]\,e^{-s\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])}\Big) \tag{86}$$

satisfies

$$f(s) < 1 \quad \text{for some } s>0. \tag{87}$$

Taking logarithms yields

$$g(s) = \ln f(s) = \sum_n \ln\Big(1-\alpha_k\,\mathbf{p}_k[n]+\alpha_k\,\mathbf{p}_k[n]\,e^{-s\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])}\Big), \tag{88}$$

with $g(0)=0$. Its derivative is

$$g'(s) = \sum_n \frac{-\alpha_k\,\mathbf{p}_k[n]\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])\,e^{-s\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])}}{1-\alpha_k\,\mathbf{p}_k[n]+\alpha_k\,\mathbf{p}_k[n]\,e^{-s\,(\mathbf{p}_k[n]-\mathbf{p}_j[n])}}. \tag{89}$$

Evaluating at $s = 0$ gives

$$g'(0) = -\alpha_k \sum_n \mathbf{p}_k[n]\Big(\mathbf{p}_k[n] - \mathbf{p}_j[n]\Big) = -\alpha_k\Big(1 - \cos\theta\Big), \quad (90)$$

where $\cos\theta = \langle \mathbf{p}_k, \mathbf{p}_j \rangle$. Since $\alpha_k > 0$ and $1 - \cos\theta > 0$, we have $g'(0) < 0$. Hence, there exists some $s > 0$ for which $g(s) < 0$ (i.e. $f(s) = e^{g(s)} < 1$), leading to an exponential decrease in the misclassification probability with $\mathcal{N}$.

Bounding $f(s)$. We start by rewriting $f(s)$ as

$$f(s) = \prod_n \Big(1 - \alpha\, B_n \left[1 - e^{-s\,(B_n - C_n)}\right]\Big), \quad s > 0, \quad (91)$$

where $B_n = \bar{p}_k[n]$ and $C_n = \bar{p}_j[n]$. Notice that

$$1 - e^{-s\,(B_n - C_n)} = s\,(B_n - C_n) - \frac{s^2 (B_n - C_n)^2}{2!} + \cdots. \quad (92)$$

Since the higher–order terms are always negative for

$$s \le \frac{1}{|B_n - C_n|},$$

we obtain the inequality

$$f(s) \le \prod_n \Big(1 - \alpha\, B_n \cdot s\,(B_n - C_n)\Big). \quad (93)$$

Taking the natural logarithm yields

$$\begin{aligned} \ln f(s) &\le \ln\left[\prod_n \Big(1 - \alpha\, B_n \cdot s\,(B_n - C_n)\Big)\right] \\ &= \sum_n \ln\Big(1 - \alpha\, s\, B_n\,(B_n - C_n)\Big). \end{aligned} \quad (94)$$

For brevity, we define

$$k_n \triangleq \alpha\, s\, B_n\,(B_n - C_n).$$

If we assume that

$$s \le \frac{1}{L\,\max_n |B_n - C_n|},$$

then $|k_n| \le 1$ (since $0 \le \alpha \le L$) and we may expand using the Taylor series:

$$\ln(1 - k_n) = -\sum_{m=1}^{\infty} \frac{k_n^m}{m}. \quad (95)$$

Thus, from (94) we have

$$\ln f(s) \le -\sum_n k_n - \sum_n \frac{k_n^2}{2} - \sum_n \frac{k_n^3}{3} - \cdots. \quad (96)$$

Discarding the higher–order (negative) terms gives the simple bound

$$\ln f(s) \le -\sum_n k_n = -\alpha\, s \sum_n B_n\,(B_n - C_n). \quad (97)$$

If we denote by θ the angle between the vectors $\mathbf{p}_k$ and $\mathbf{p}_j$, then

$$\sum_n B_n\,(B_n - C_n) = \langle \mathbf{p}_k,\, \mathbf{p}_k - \mathbf{p}_j \rangle = 1 - \langle \mathbf{p}_k, \mathbf{p}_j \rangle = 1 - \cos\theta.$$

Hence,

$$\ln f(s) \le -\alpha\, s\,(1 - \cos\theta). \quad (98)$$

Exponentiating both sides yields

$$f(s) \le \exp\Big\{-\alpha\, s\,(1 - \cos\theta)\Big\}. \quad (99)$$

Since the misclassification probability satisfies

$$\Pr\{e_j < 0\} \le M(s) \le \big[f(s)\big]^{\mathcal{N}}, \quad (100)$$

we deduce that

$$\Pr\{e_j < 0\} \le \exp\Big\{-\mathcal{N}\,\alpha\, s\,(1 - \cos\theta)\Big\}, \quad (101)$$

where

$$\alpha = \|1 - e^{-\Phi_0\,\mathbf{d}_k}\|.$$

Since $g(s) = \ln f(s)$ is linear in s, the tightest bound is obtained by choosing

$$s = \frac{1}{L\,\max_n |B_n - C_n|}.$$

Thus, the final form of the bound is

$$\Pr\{e_j < 0\} \le \exp\left\{-\frac{\mathcal{N}\,\|1 - e^{-\Phi_0\,\mathbf{d}_k}\|\,(1 - \cos\theta)}{L\,\max_n\big(|\bar{p}_k[n] - \bar{p}_j[n]|\big)}\right\}. \quad (102)$$

A.2 Theory for Full Spectrum Reconstruction

Let

$$\mathbf{d} = (d(1), d(2), \ldots, d(L)) \in \mathbb{R}^L, \quad d(n) \ge 0, \quad (103)$$

be the true normalized spectrum. Two cases are considered:

(1) **Scaling Estimator (Small $\Phi_0\, d(n)$):** When $\Phi_0\, d(n)$ is small so that

$$1 - e^{-\Phi_0\, d(n)} \approx \Phi_0\, d(n), \quad (104)$$

the measurement at wavelength n is modeled as

$$b(n) \sim \text{Binomial}\Big(N,\ \Phi_0\, d(n)\Big), \quad (105)$$

and the natural unbiased estimator is

$$\hat{d}(n) = \frac{b(n)}{N\,\Phi_0}, \quad (106)$$

since

$$E\big[\hat{d}(n)\big] = \frac{E\big[b(n)\big]}{N\,\Phi_0} = \frac{N\,\Phi_0\, d(n)}{N\,\Phi_0} = d(n). \quad (107)$$

(2) **Log–Transform Estimator (General $\Phi_0\, d(n)$):** For the general case (without assuming $\Phi_0\, d(n)$ is small), the measurement is modeled as

$$b(n) \sim \text{Binomial}\Big(N,\ 1 - e^{-\Phi_0\, d(n)}\Big), \quad (108)$$

and we define the log–transform estimator by

$$\hat{d}(n) = -\frac{1}{\Phi_0} \ln\Big(1 - \frac{b(n)}{N}\Big). \quad (109)$$

If $\frac{b(n)}{N}$ exactly equals $p(n) = 1 - e^{-\Phi_0\, d(n)}$, then

$$\hat{d}(n) = -\frac{1}{\Phi_0} \ln\Big(1 - p(n)\Big) = d(n). \quad (110)$$

A.2.1 Scaling Estimator (Small $\Phi_0\, d(n)$).

Measurement Model and Estimator. Assume that at wavelength n the true spectral value is

$$d = d(n) > 0. \tag{111}$$

When $\Phi_0\, d(n)$ is small, we have

$$1 - e^{-\Phi_0\, d(n)} \approx \Phi_0\, d(n). \tag{112}$$

Thus the measurement is modeled as

$$b(n) \sim \text{Binomial}\Big(N,\ \Phi_0\, d(n)\Big), \tag{113}$$

and the unbiased estimator is

$$\hat{d}(n) = \frac{b(n)}{N\,\Phi_0}, \tag{114}$$

since

$$E\big[\hat{d}(n)\big] = \frac{E\big[b(n)\big]}{N\,\Phi_0} = \frac{N\,(\Phi_0\, d(n))}{N\,\Phi_0} = d(n). \tag{115}$$

Chernoff Bound via the MGF.. Define the centered variable

$$X = b(n) - N\,\Phi_0\, d(n), \tag{116}$$

so that the estimation error is

$$\hat{d}(n) - d(n) = \frac{X}{N\,\Phi_0}. \tag{117}$$

Our goal is to bound

$$P\Big(|\hat{d}(n) - d(n)| \geq \epsilon\Big) = P\Big(|X| \geq N\,\Phi_0\,\epsilon\Big) \tag{118}$$

for any $\epsilon > 0$.

MGF of X. Since $b(n) \sim \text{Binomial}\Big(N,\ \Phi_0\, d(n)\Big)$, its MGF is

$$M_{b(n)}(s) = E\Big(e^{s\,b(n)}\Big) = \Big(1 - \Phi_0\, d(n) + \Phi_0\, d(n)\, e^{s}\Big)^{N}. \tag{119}$$

Thus, the MGF of

$$X = b(n) - N\,\Phi_0\, d(n) \tag{120}$$

is

$$\begin{aligned} M_X(s) &= E\Big(e^{sX}\Big) = E\Big(e^{s\left(b(n) - N\,\Phi_0\, d(n)\right)}\Big) \\ &= e^{-s\,N\,\Phi_0\, d(n)}\, M_{b(n)}(s) \\ &= \Big(e^{-s\,\Phi_0\, d(n)}\Big(1 - \Phi_0\, d(n) + \Phi_0\, d(n)\, e^{s}\Big)\Big)^{N}. \end{aligned} \tag{121}$$

Upper Tail Bound. For any $s > 0$, Chernoff's inequality gives

$$P\Big(X \geq N\,\Phi_0\,\epsilon\Big) \leq \frac{M_X(s)}{e^{s\,N\,\Phi_0\,\epsilon}}. \tag{122}$$

That is,

$$\begin{aligned} P\big(\hat{d}(n) - d(n) \geq \epsilon\big) &= P\big(X \geq N\,\Phi_0\,\epsilon\big) \\ &\leq \left[e^{-s\,\Phi_0\left(d(n)+\epsilon\right)}\Big(1 - \Phi_0\, d(n) \right. \\ &\qquad \left. + \Phi_0\, d(n)\, e^{s}\Big)\right]^{N}. \end{aligned} \tag{123}$$

Defining

$$f(s) = e^{\,-s\,\Phi_0\left(d(n)+\epsilon\right)}\Big(1 - \Phi_0\, d(n) + \Phi_0\, d(n)\, e^{s}\Big), \tag{124}$$

we have

$$P\Big(\hat{d}(n) - d(n) \geq \epsilon\Big) \leq \Big[f(s)\Big]^{N}. \tag{125}$$

Proof that $f(s) < 1$ for Some $s > 0$. Notice that when $s = 0$,

$$f(0) = e^{0}\Big(1 - \Phi_0\, d(n) + \Phi_0\, d(n)\Big) = 1. \tag{126}$$

Differentiating, writing

$$\begin{aligned} &f(s) = A(s)\, B(s), \quad \text{where} \\ &A(s) = 1 - \Phi_0\, d(n) \ + \ \Phi_0\, d(n)\, e^{s}, \\ &B(s) = e^{-s\,\Phi_0\,\left(d(n)+\epsilon\right)}. \end{aligned} \tag{127}$$

we obtain

$$f'(s) = A'(s)B(s) + A(s)B'(s), \tag{128}$$

with

$$A'(s) = \Phi_0\, d(n)\, e^{s}, \quad B'(s) = -\Phi_0\Big(d(n) + \epsilon\Big)e^{\,-s\,\Phi_0\left(d(n)+\epsilon\right)}. \tag{129}$$

Evaluating at $s = 0$ (using $e^0 = 1$) yields

$$f'(0) = \Phi_0\, d(n) - \Phi_0\Big(d(n) + \epsilon\Big) = -\Phi_0\,\epsilon < 0. \tag{130}$$

Because $f(s)$ is continuously differentiable and $f(0) = 1$ with $f'(0) < 0$, by continuity there exists a $\delta > 0$ such that for all $s \in (0, \delta)$,

$$f(s) < 1 \tag{131}$$

and by Chernoff's inequality,

$$P\big(\hat{d}(n) - d(n) \geq \epsilon\big) \ \leq \ \big[f(s)\big]^{N} \ \leq \ \exp\big(-\kappa_n\, N\big), \tag{132}$$

where $\kappa_n = -\ln\big[f(s)\big] > 0$. A similar derivation (using $s < 0$) establishes the same exponential decay for the lower tail. Hence,

$$P\big(|\hat{d}(n) - d(n)| \ \geq \ \epsilon\big) \ \leq \ 2\, \exp\big(-\kappa_n\, N\big). \tag{133}$$

A.2.2 Extension to the Full Spectrum. Consider the full spectral vector

$$\mathbf{d} = \big(d(1), d(2), \ldots, d(L)\big) \in \mathbb{R}^{L}. \tag{134}$$

For each channel n, suppose

$$b(n) \sim \text{Binomial}\Big(N,\ \Phi_0\, d(n)\Big). \tag{135}$$

Define the estimator

$$\hat{d}(n) = \frac{b(n)}{N\,\Phi_0}. \tag{136}$$

Then, for each n, the bound

$$P\Big(\big|\hat{d}(n) - d(n)\big| \geq \epsilon\Big) \leq 2\, \exp\Big(-\kappa_n\, N\Big) \tag{137}$$

holds.

Applying a union bound over the L channels,

$$\begin{aligned} P\Big(\max_{1\leq n\leq L}\big|\hat{d}(n) - d(n)\big| \geq \epsilon\Big) &\leq \sum_{n=1}^{L} 2\, \exp\Big(-\kappa_n\, N\Big) \\ &\leq 2L\, \exp\Big(-\kappa\, N\Big), \end{aligned} \tag{138}$$

where

$$\kappa = \min_{1\leq n\leq L} \kappa_n. \tag{139}$$

In particular, if the error is measured in the ℓ_2 norm, then

$$\|\hat{\mathbf{d}} - \mathbf{d}\|^2 \le L \max_{1\le n\le L} \left|\hat{d}(n) - d(n)\right|^2. \tag{140}$$

Thus, the overall estimation error decays exponentially with N.

A.2.3 Log Transform Estimator for General $\Phi_0\, d(n)$.

Measurement Model and Estimator. At wavelength n let the true spectral value be

$$d = \mathbf{d}(n) > 0. \tag{141}$$

Now assume the measurement is given by

$$\mathbf{y}(n) \sim \text{Binomial}\Big(N,\ \mathbf{p}(n)\Big), \quad \text{with} \quad \mathbf{p}(n) = 1 - e^{-\Phi_0\, d(n)}. \tag{142}$$

Noting that

$$E\Big[\frac{\mathbf{y}(n)}{N}\Big] = \mathbf{p}(n), \quad \text{and} \quad -\ln\Big(1 - \mathbf{p}(n)\Big) = \Phi_0\, \mathbf{d}(n), \tag{143}$$

we define the log–transform estimator as

$$\hat{\mathbf{d}}(n) = -\frac{1}{\Phi_0} \ln\Big(1 - \frac{\mathbf{y}(n)}{N}\Big). \tag{144}$$

If $\frac{\mathbf{y}(n)}{N}$ exactly equals $\mathbf{p}(n)$, then

$$\hat{\mathbf{d}}(n) = -\frac{1}{\Phi_0} \ln\Big(1 - \mathbf{p}(n)\Big) = \mathbf{d}(n). \tag{145}$$

Relating the Error to the Proportion. Define

$$Y = \frac{\mathbf{y}(n)}{N}. \tag{146}$$

Then,

$$\hat{\mathbf{d}}(n) = -\frac{1}{\Phi_0} \ln(1 - Y). \tag{147}$$

For an error tolerance $\epsilon > 0$, the event

$$\hat{\mathbf{d}}(n) - \mathbf{d}(n) \ge \epsilon \tag{148}$$

is equivalent to

$$-\frac{1}{\Phi_0} \ln(1 - Y) - \mathbf{d}(n) \ge \epsilon. \tag{149}$$

Multiplying by $-\Phi_0$ (and reversing the inequality) gives

$$\ln(1 - Y) \le -\Phi_0\Big(\mathbf{d}(n) + \epsilon\Big), \tag{150}$$

so that, after exponentiating,

$$1 - Y \le e^{-\Phi_0\left(\mathbf{d}(n)+\epsilon\right)} \quad \Longrightarrow \quad Y \ge 1 - e^{-\Phi_0\left(\mathbf{d}(n)+\epsilon\right)}. \tag{151}$$

Defining

$$q_+ = 1 - e^{-\Phi_0\left(\mathbf{d}(n)+\epsilon\right)}, \tag{152}$$

and noting that $\mathbf{p}(n) = 1 - e^{-\Phi_0\, \mathbf{d}(n)}$ with $q_+ > p(n)$, we write

$$q_+ = (1+\delta)p(n) \quad \Longrightarrow \quad \delta = \frac{q_+}{p(n)} - 1 > 0. \tag{153}$$

Thus, the event $\{\hat{\mathbf{d}}(n) - \mathbf{d}(n) \ge \epsilon\}$ is equivalent to

$$Y \ge (1+\delta)\mathbf{p}(n). \tag{154}$$

Chernoff Bound. Since $\mathbf{y}(n) \sim \text{Binomial}(N, p(n))$ with mean $Np(n)$, its MGF is

$$M_{y(n)}(s) = (1 - p(n) + p(n)e^s)^N. \tag{155}$$

Thus, applying the Chernoff bound,

$$\begin{aligned} P\Big(Y \ge (1+\delta)p(n)\Big) &= P\Big(\mathbf{y}(n) \ge (1+\delta)Np(n)\Big) \\ &\le \frac{M_{\mathbf{y}(n)}(s)}{e^{s(1+\delta)Np(n)}} \\ &= \left[\frac{1 - p(n) + p(n)e^s}{e^{s(1+\delta)p(n)}}\right]^N. \end{aligned} \tag{156}$$

Defining

$$f(s) = \frac{1 - p(n) + p(n)e^s}{e^{s(1+\delta)p(n)}}, \tag{157}$$

we have

$$P\Big(Y \ge (1+\delta)p(n)\Big) \le \big[f(s)\big]^N. \tag{158}$$

Direct Proof that $f(s) < 1$ for Some $s > 0$. Observe that at $s = 0$,

$$f(0) = \frac{1 - p(n) + p(n)}{1} = 1. \tag{159}$$

Differentiating, writing

$$\begin{aligned} &f(s) = A(s)B(s), \quad \text{with} \\ &A(s) = 1 - p(n) + p(n)e^s, \\ &B(s) = e^{-s(1+\delta)p(n)}. \end{aligned} \tag{160}$$

so that

$$f'(s) = A'(s)B(s) + A(s)B'(s), \tag{161}$$

with

$$A'(s) = p(n)e^s, \quad B'(s) = -(1+\delta)p(n)e^{-s(1+\delta)p(n)}. \tag{162}$$

Evaluating at $s = 0$ (with $e^0 = 1$) yields

$$f'(0) = p(n) - (1+\delta)p(n) = -\delta\, p(n) < 0. \tag{163}$$

Thus, there exists $\eta > 0$ such that for $0 < s < \eta$ we have $f(s) < 1$, and with the similar proof for the lower tail, we obtain

$$P\Big(|\hat{d}(n) - d(n)| \ge \epsilon\Big) \le 2\exp(-\kappa_n N), \tag{164}$$

with $\kappa_n = -\ln f(s) > 0$.

A.2.4 Extension to the Full Spectrum. For the spectral vector

$$\mathbf{d} = (d(1), d(2), \ldots, d(L)) \in \mathbb{R}^L, \tag{165}$$

assume that at each wavelength n

$$b(n) \sim \text{Binomial}\Big(N,\ 1 - e^{-\Phi_0\, d(n)}\Big), \tag{166}$$

and the estimator is given by

$$\hat{d}(n) = -\frac{1}{\Phi_0} \ln\Big(1 - \frac{b(n)}{N}\Big). \tag{167}$$

Then, for each n,

$$P\Big(|\hat{d}(n) - d(n)| \ge \epsilon\Big) \le 2\exp(-\kappa_n N), \tag{168}$$

Table 3. List of components in our lab prototype

Name	Company	Item
DMD	Vialux	DLP7000
SPAD	Pi Imaging	SPAD 512S
Grating (Grism)	Edmund Optics	70 grooves/mm (#46-068)
Prism (Grism)	Edmund Optics	5.22° wedge angle (#84-863)
4f relay lens	Thorlabs	75mm (AC254-075-AB)
DMD-to-SPAD relay	Nikon	105mm f/2.8D AF Micro

and by a union bound,

$$P\Big(\max_{1\le n\le L}|\hat{d}(n)-d(n)|\ge\epsilon\Big)\le 2L\exp(-\kappa N), \tag{169}$$

with

$$\kappa=\min_{1\le n\le L}\kappa_n. \tag{170}$$

In particular, if the error is measured in the ℓ_2 norm,

$$\|\hat{\mathbf{d}}-\mathbf{d}\|^2\le L\max_{1\le n\le L}|\hat{d}(n)-d(n)|^2, \tag{171}$$

so that the overall estimation error decays exponentially with N.

B Details of Hi-SPAD lab-built prototype

In this section, we provide the details of the Hi-SPAD lab-built prototype.

B.1 Hardware Design

Our hardware prototype is built on a pushbroom camera architecture with a Digital Micromirror Device (DMD) for programmable line scanning and uniquely integrates a SPAD sensor to enable video-rate hyperspectral imaging under photon-limited conditions. Figure 6 showcases our lab-built prototype of the Hi-SPAD camera.

The line-scanning HSI acquisition process with a DMD is as follows. The light coming from the scene is focused on the DMD through an objective lens and a 4f relay system. The 4f relay system is composed of two lenses with focal length of 75 mm. With this arrangement, we can easily manipulate the scene-to-sensor magnification by changing the objective lens, without affecting the rest of the optics. Since we are implementing a pushbroom camera, the DMD selectively reflects a single column of the focused image by flipping a corresponding single column of micromirrors toward the imaging arm. The specific DMD model that we used, Vialux DLP7000, has micromirrors that flip along their diagonal, which normally would reflect the optical axis off the horizontal plane; to resolve this, we tilted the DMD and the SPAD sensor by 45°. Finally, as is common in all DMD-based imaging design, the optical axis is no longer perpendicular to the DMD post-reflection; hence, the SPAD camera is tilted so as to satisfy the Scheimpflug principle [Scheimpflug 1904]. We place a spectral disperser between the DMD and the SPAD sensor; specifically, we designed a grism, a combination of a transmission grating and a prism [Bowen and Vaughan 1973], to simplify the overall alignment and avoid additional skew in the optical axis. With this, the reflected light from the DMD first reaches the grating, where the dispersion is performed, and the prism redirects the first-order dispersion to the desired location on the sensor. A closer look at the captured column-wise dispersed images is shown in the first row of Figure 7. By sequentially displaying columns on the DMD and capturing the spectrally dispersed images of each column, we can acquire the data for reconstructing the hyperspectral cube.

The spatial and spectral dimensions of the captured hyperspectral cube are jointly determined by the resolutions of the DMD and the SPAD sensor, as well as the optical magnification. The DMD has a resolution of 1024×768 micromirrors and serves as the spatial scanning interface, while the SPAD sensor has a resolution of 512×512 pixels and captures the spectrally dispersed light from each selected DMD column. To ensure that the full width of the DMD is mapped onto the vertical dimension of the SPAD sensor, the optical magnification was chosen such that all 1024 DMD rows are imaged onto the 512-pixel height of the SPAD sensor. The horizontal axis of the SPAD sensor is reserved for capturing the spectrally dispersed signal. In our prototype, 123 horizontal pixels were allocated to accommodate the full spectral dispersion range on the SPAD sensor, resulting in 123 spectral channels in the reconstructed hyperspectral cube. To prevent overlap between the first- and second-order diffraction, we computed the angular separation introduced by the spectral disperser and restricted the usable spectrum using a bandpass filter. As a result, the operational spectral range of our HSI prototype spans from 445 nm to 910 nm.

It is important to note that while the horizontal spatial dimension (corresponding to the DMD columns) is preserved through sequential scanning, the vertical spatial dimension is limited by the SPAD sensor resolution. Specifically, the 768 DMD columns are scanned sequentially and thus fully preserved in the reconstructed hyperspectral cube. However, along the vertical dimension, the 1024 DMD rows are optically mapped onto the 512-pixel height of the SPAD sensor within a single capture. As a result, the effective vertical spatial resolution of the reconstructed hyperspectral cube is 512 pixels rather than 1024.

As a final step, hyperspectral video is captured by repeating the above process for each hyperspectral frame, exploiting the high frame rates of the DMD and SPAD sensor. The DMD and SPAD sensor are synchronized by connections on the breadboard. The DMD sends a trigger signal to the SPAD sensor when it displays a new pattern, guiding the sensor to capture a new image. The maximum speed of the DMD is 20,000 patterns per second and the SPAD sensor is 100,000 binary frames per second; theoretically, the fastest (and noisiest) capture per single column is five 1-bit frames at 20,000 fps. With this operation, the system can acquire 26 fps hyperspectral video with 123 spectral channels in the wavelength range of 445 nm to 910 nm.

B.2 Spectral Calibration and HSI Reconstruction

Figure 7 shows the captured spectrally dispersed images of the sequence of DMD patterns and captured images when a monochromatic light source is illuminated onto the DMD with a dot-column pattern used for calibration. We define a mapping function F that relates the opened DMD pixel coordinates $(x_{\text{dmd}}, y_{\text{dmd}})$ and wavelength (λ) to the corresponding SPAD sensor coordinates:

$$(x_{\text{img}},\ y_{\text{img}})=F(x_{\text{dmd}},\ y_{\text{dmd}},\lambda). \tag{172}$$

From the images in the second row of Figure 7, we detected pixel locations of dots corresponding to first-order dispersion and built a dataset of the DMD pixel coordinates and the corresponding SPAD sensor coordinates under different wavelengths:

$$\mathcal{M} = \{(x^k_{\text{dmd}}, y^k_{\text{dmd}}, x^k_{\text{img}}, y^k_{\text{img}}, \lambda_k)\}_{k=1}^{Q}, \tag{173}$$

where Q is the number of calibration points. A polynomial approximation $\widehat{F}$ of F is then obtained by minimizing the least-squares cost

$$E = \sum_{k=1}^{Q} \left\| (x^k_{\text{img}}, y^k_{\text{img}}) - \widehat{F}(x^k_{\text{dmd}}, y^k_{\text{dmd}}, \lambda_k) \right\|^2. \tag{174}$$

Once the mapping function is calibrated, the hyperspectral cube is reconstructed by remapping the SPAD sensor measurements to images at the corresponding wavelengths.

B.3 Background Subtraction for Stray Light Removal

Since the SPAD sensor is highly sensitive to photon-level signals, stray light can easily contaminate measurements—especially under low-light conditions typical of video-rate hyperspectral imaging, where only a few photons are detected per spectral band. To mitigate this, we enclosed the entire Hi-SPAD camera system with black cloth and cardboard shielding during data acquisition to minimize ambient light intrusion.

Despite this shielding, light from the scene passing through the objective lens can still reflect off the DMD surface and reach the SPAD sensor, even when an all-zero pattern is displayed (i.e., when all micromirrors are directed away from the imaging path). As shown in the top right of Figure 8, this back-reflected light persists in high-stray-light conditions and triggers consistent sensor activations. During line scanning, this appears as spatially translating bright spectral peaks, resulting in rainbow-like streaks in the reconstructed RGB image or persistent bright pixels in individual monochromatic bands. These artifacts are especially prominent in low-light settings, where signal sparsity increases their relative visibility.

To correct for this, we perform stray-light correction via background subtraction. Specifically, we compute a background image by identifying and masking out the region corresponding to first-order spectral dispersion (mode-1), using the learned mapping function from the calibration step. We then average the remaining pixels, under the assumption that consistently firing (stray light–affected) pixels dominate this region. To account for the masked-out signal region, we scale the averaged background image by the ratio of the full DMD width to the width of the excluded dispersion region. This scaled background image is subtracted from each column-wise raw measurement to yield the corrected image. Through this method, we avoid capturing a separate background image for each acquisition, which would add additional time to the capture process and slow down the frame rate. The bottom row of Figure 8 illustrates the effect: after background subtraction, rainbow streaks and spurious bright pixels are substantially reduced, even under extreme low-light conditions captured using the smallest aperture setting.

C Gallery of the Hi-SPAD Hyperspectral Video

In this section, we exhibit the gallery of the Hi-SPAD hyperspectral video for indoor, sunny outdoor, and cloudy outdoor scenes.

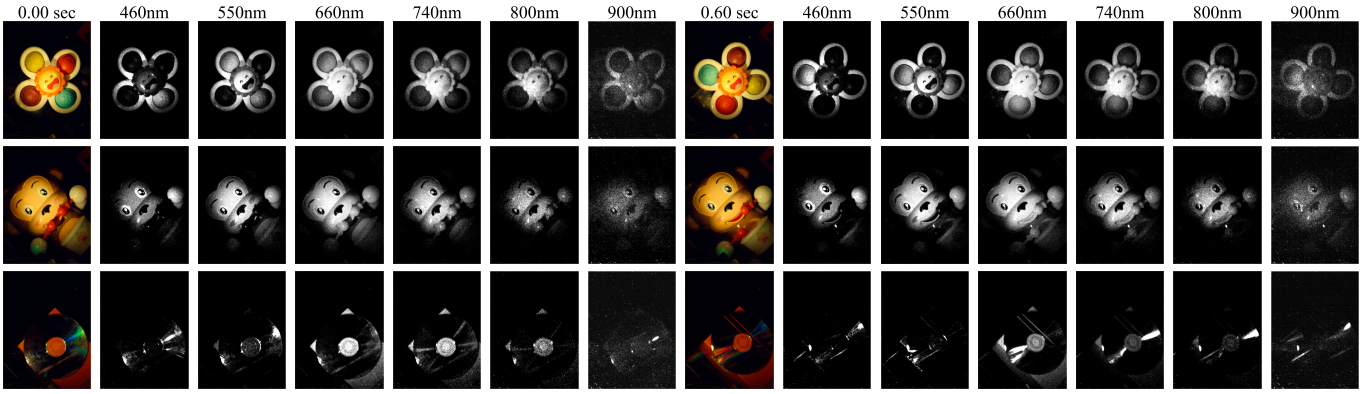


Fig. 21. Gallery of the Hi-SPAD hyperspectral video for an indoor scene. We exhibit selected bands of the hyperspectral video with the reconstructed RGB frame. The video was reconstructed from the raw Hi-SPAD measurement with a patch-wise reconstruction network.

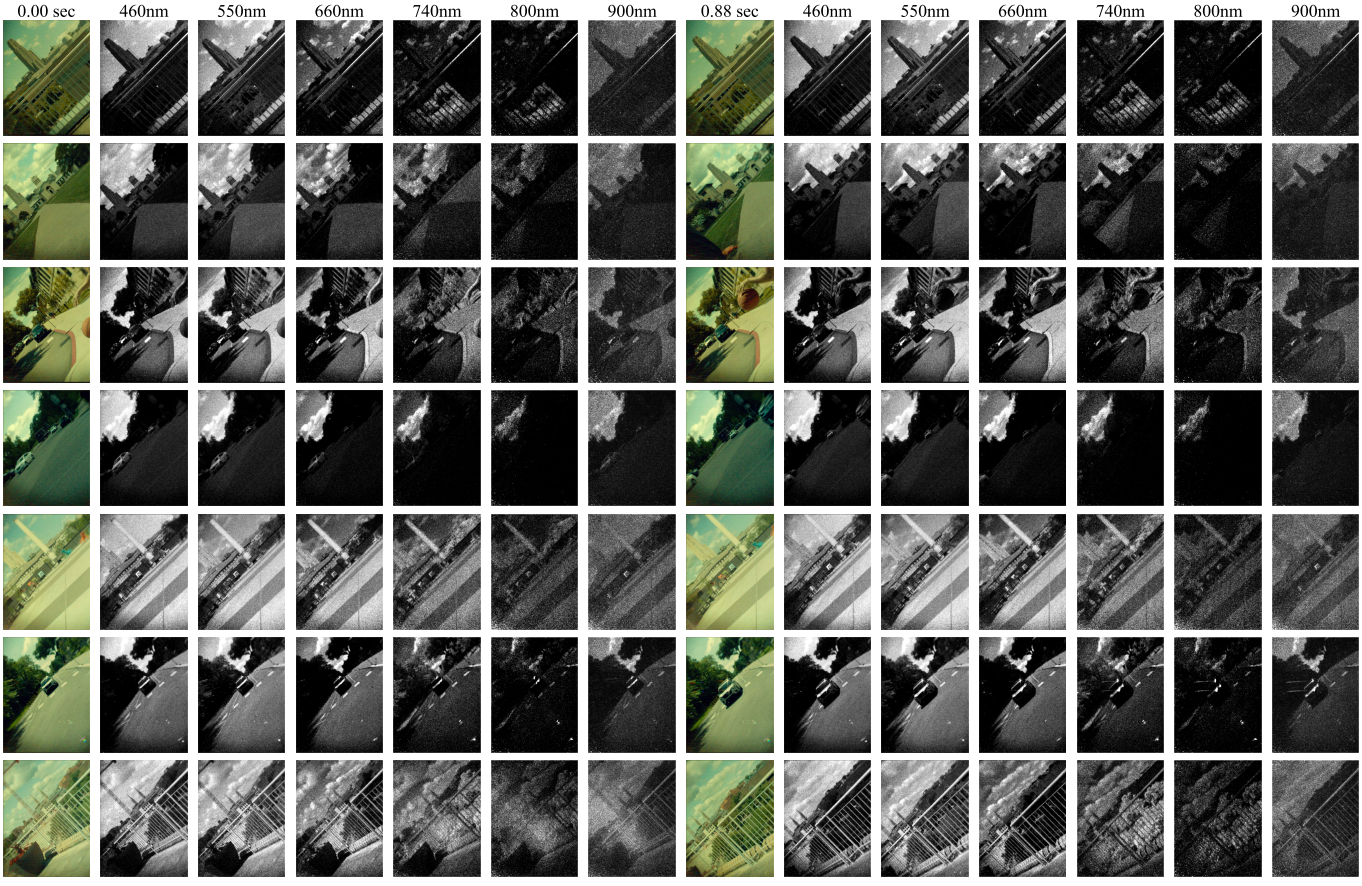


Fig. 22. Gallery of the Hi-SPAD hyperspectral video for a sunny outdoor scene. We exhibit selected bands of the hyperspectral video with the reconstructed RGB frame. The video was reconstructed from the raw Hi-SPAD measurement with a patch-wise reconstruction network.

Fig. 23. Gallery of the Hi-SPAD hyperspectral video for a cloudy outdoor scene. We exhibit selected bands of the hyperspectral video with the reconstructed RGB frame. The video was reconstructed from the raw Hi-SPAD measurement with a patch-wise reconstruction network.