SASSI — Super-Pixelated Adaptive Spatio-Spectral Imaging

Vishwanath Saragadam, Michael DeZeeuw, Richard G. Baraniuk, *Fellow, IEEE*,
Ashok Veeraraghavan, *Senior Member, IEEE*, and Aswin C. Sankaranarayanan

Abstract—We introduce a novel video-rate hyperspectral imager with high spatial, temporal and spectral resolutions. Our key hypothesis is that spectral profiles of pixels within each super-pixel tend to be similar. Hence, a scene-adaptive spatial sampling of a hyperspectral scene, guided by its super-pixel segmented image, is capable of obtaining high-quality reconstructions. To achieve this, we acquire an RGB image of the scene, compute its super-pixels, from which we generate a spatial mask of locations where we measure high-resolution spectrum. The hyperspectral image is subsequently estimated by fusing the RGB image and the spectral measurements using a learnable guided filtering approach. Due to low computational complexity of the superpixel estimation step, our setup can capture hyperspectral images of the scenes with little overhead over traditional snapshot hyperspectral cameras, but with significantly higher spatial and spectral resolutions. We validate the proposed technique with extensive simulations as well as a lab prototype that measures hyperspectral video at a spatial resolution of 600×900 pixels, at a spectral resolution of 10 nm over visible wavebands, and achieving a frame rate at 18fps.

Index Terms—Computational Photography, Hyperspectral Imaging, Adaptive Imaging, Hyperspectral Fusion, Superpixels

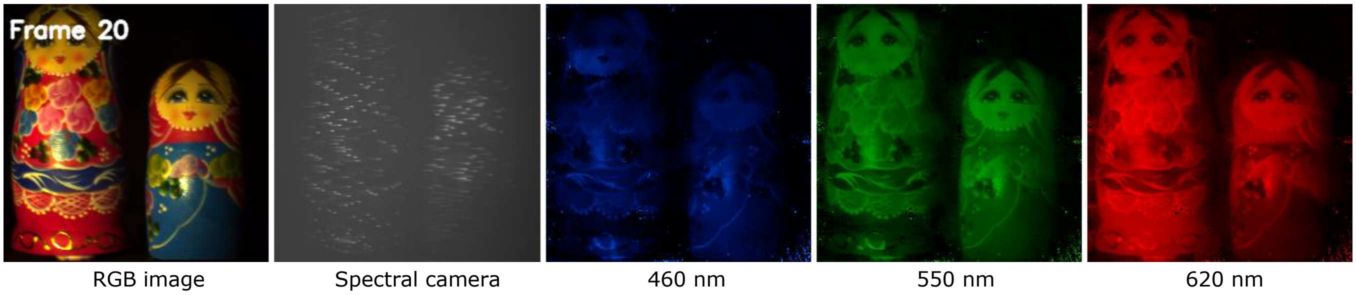


Fig. 1: **High resolution video rate hyperspectral imaging.** We propose a novel hyperspectral camera that is capable of capturing high spatial and spectral resolution images at video rate. We achieve this by a sparse, scene-adaptive spectral sampling, and then fusing it with an auxiliary RGB image. Full video can be accessed from the supplementary material.

1 Introduction

One of the classic approaches for plenoptic imaging is to sacrifice spatial resolution of a sensor to obtain resolution in other dimensions of light. This principle has been used for sensing color with filter arrays [1], polarization state with per-pixel polarizers [2], light fields with angle sensitive pixels [3] and integral imagers [4], high-speed imaging with staggered pixel exposures [5], and high dynamic range with per-pixel neutral density filters [6]. This trade-off is encapsulated in the concept of assorted pixels [7] where the pixels in the sensor also sample along non-spatial dimensions of the incident light — be it time, spectrum, angle, dynamic range or combinations thereof. The main challenge in sensing with assorted pixels has been the steep loss in spatial resolution, especially when we need to allocate a large portion of samples to other dimensions.

This paper provides a design template for novel hyperspectral cameras that work in the spirit of assorted pixels, i.e., trading spatial resolution of the sensor to obtain high spectral resolutions. The key differentiating aspect of our design is that it performs a *scene-adaptive* tiling of the spatio-spectral voxels onto the image sensor, which is in sharp contrast to prior work where the measurements are non-adaptive. Our approach relies on an assumption that, for most scenes, the spectrum of pixels is similar over small spatial neighborhoods that do not cross texture and material boundaries, such as superpixels. Hence, if we knew the regions with homogeneous spectra a priori, then we could sample one or more spectral profile for each region and propagate it to the remaining pixels, which would avoid blurring of the measurements as well as loss of spatial resolution. We refer to our approach as Super-pixelated Adaptive Spatio-Spectral Imager (SASSI).

SASSI relies on two components: a *guide RGB camera*, that is used to determine the spatial locations where we sample spectrum; and, a *spatio-spectral sampler*, that is capable of measuring the high-resolution spectrum at chosen spatial

- *Vishwanath Saragadam, Richard Baraniuk, and Ashok Veeraraghavan are with the ECE Department, Rice University, Houston, TX 77025.*
- *Michael DeZeeuw, and Aswin Sankaranarayanan are with the ECE Department, Carnegie Mellon University, Pittsburgh, PA 15213.*
- *Corresponding author email: saswin@andrew.cmu.edu*

locations by spreading their spectrum in space. The spatio-spectral sampler is identical to that of the single disperser coded aperture snapshot spectral imager (CASSI) architecture [8] with an exception of programmable spatial mask via a spatial light modulator (SLM). Both components share the same view point and are time synchronized.

To acquire the hyperspectral image (HSI) at a time instant t, we first perform super-pixel segmentation on the guide image acquired at the previous time instant, say $t - 1$, to obtain a clustering of compact regions that have similar RGB intensities; we use this as a proxy for the material map of the scene. We subsequently create a spatial mask satisfying two key criteria: first, the mask pixels are sufficiently far apart from each other so that their spectral spreads in the spatio-spectral sampler do not overlap; and second, the number of super-pixels selected is maximized along with the total sensor pixel count in the spatio-spectral sampler. We now acquire measurements at time t, with the adaptively-designed mask displayed on the SLM. Finally, we estimate the HSI by fusing the RGB image and the spatio-spectral samples acquired at time t. Since all measurements being fused are captured simultaneously in a time-synchronized manner, we avoid the need for registration of images across time instants.

Contributions. We propose SASSI, an adaptive HSI sensing approach, and make the following contributions.

- *Optical design.* We provide a compact, light efficient design for the SASSI camera.
- *Reconstruction of the HSI.* We propose two recovery approaches applicable to a variety of scenarios. The first one is a rank-1 reconstruction where each pixel within the superpixel is assigned a scaled version of the measured spectrum at the centroid. The second one is a per-wavelength neural network based reconstruction that relies on learnable guided filtering.
- *Lab prototype.* We build a prototype of the SASSI camera that is capable of acquiring HSIs with a spatial resolution of 600×900 over a visible waveband $400 - 700$ nm at a spectral resolution of 10 nm, operating at 18 fps.

We provide extensive evaluation using simulations and real data captured using the prototype. Our results come with a new HSI dataset with 50 scenes, including some microscopic samples. The dataset, training scripts, and pre-trained models can be downloaded from [9].

Limitations. The limitations of SASSI are three fold.

- *Thin spatial structures and clutter.* Our setup requires that sampling locations not be close to each other, which precludes measurement of highly complex spatial structures.
- *Scenes with rapid motion.* Our approach fuses RGB and spectral measurements from the same time instance; however, rapid motion between frames may create sampling locations that are off from the desired pattern.
- *Metamerism.* For scenes with metamerism, we run the risk of violating the assumption that the spectrum remains largely the same within each super-pixel.

2 Prior Work

We review some of the key prior art in hyperspectral imaging and adaptive sensing.

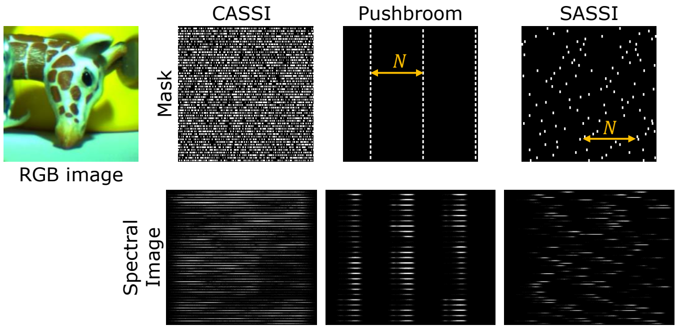


Fig. 2: **Various sampling schemes.** CASSI-type cameras sense with a dense spatial pattern requiring spectral demultiplexing. On the other hand, pushbroom cameras do not require demultiplexing but lead to severe loss in spatial resolution. SASSI provides a unique tradeoff where the sampling pattern is sparse enough to avoid any spectral multiplexing, while requiring only a single scene-adaptive spectral measurement.

2.1 Hyperspectral camera designs

The SASSI architecture is closely related to many existing designs for hyperspectral imaging; we discuss them next.

Classical hyperspectral cameras. Hyperspectral images capture information as a function of both space and wavelength, and can be represented as a 3D volume, $H(x, y, \lambda)$. Spectral profiles are often unique identifiers of materials and illuminants and have been used in several scientific [10–12] and computer vision [13–15] applications. Traditional cameras image either one wavelength at a time, or all wavelengths of one row of the image at a time, both of which require long exposure times. The pushbroom camera samples a single or multiple spatial columns of the HSI by smearing their spectral profiles onto the spatial sensor; this can be seen in Figure 2. Both pushbroom and the proposed SASSI avoid the spectral streaks from overlapping with each other, and hence, they both provide a sub-sampled Nyquist sampling of the HSI. The key difference lies in the uniform non-adaptive sampling in pushbroom versus the non-uniform adaptive sampling in SASSI.

Compressive sensing refers to a class of techniques that senses a signal from a set of random non-adaptive linear measurements, whose cardinality is smaller than the signal dimension. The coded aperture snapshot spectral imaging (CASSI) camera and its variants [8, 16–22] are examples of compressive imagers where a single spatio-spectrally multiplexed image is demultiplexed into a full-resolution HSI using various signal priors [23]. Such cameras lead to severe loss of spatial resolution, and require solving complex optimization problems. SASSI can be interpreted as an extension of the CASSI architecture. In addition to being adaptive, our method avoids multiplexing of spectrum across pixels. This enables recovery of high resolution spectra without compromising spatial resolution. Figure 2 shows the sampling masks for the CASSI system; the higher density of openings in CASSI leads to multiplexing of spectral measurements from different spatial locations, which is a key difference to the proposed technique.

Hybrid cameras. Hybrid cameras fuse a low-resolution hyperspectral imager and a high-resolution RGB camera [24, 25] to obtain a high spatial and spectral camera; this process is also referred to panchromatic sharpening [26, 27]. Cao et al. [28] used a CASSI system in place of a low-resolution hyperspectral camera. Instead of obtaining a random and dense spatio-spectral multiplexing, they rely on a uniform sparse mask with a transmission efficiency $< 0.03\%$ and then propagate the sparse set of spectral to all other locations by fusing it with an RGB image. However, a uniform sampling strategy is not suitable for scenes that have small, irregular objects.

2.2 Adaptive sampling strategies

Instead of fixing the sampling scheme across all scenes, it is intuitive that an adaptive sampling tuned to the specifics of an individual scene can provide higher accuracy [29–31]. In the context of snapshot hyperspectral imaging, Feng et al. [32] and Ma et al. [33] showed that an RGB image can be used to tailor the mask for CASSI, thereby creating a content-adaptive hyperspectral sampling strategy. This is a promising approach to snapshot hyperspectral imaging, and in many ways, is a key motivation for our own work.

There are some important differences between SASSI and the techniques mentioned above. Ma et al. [33] rely on temporal correction of errors, as well as a simple modification of bilateral filtering in the reconstruction procedure, which requires a very accurate estimation of correspondences across time-frames — a task that is often difficult in the absence of texture in the scene. SASSI instead relies on RGB image and spectral information from the *same* frame, thereby not requiring any image alignment. Further, we rely on a learned guided filtering based reconstruction technique that can recover HSIs with high accuracy.

2.3 Super-pixelation of images

Super pixels provide an over-segmentation of an image into homogeneous regions [34–36], and have been used extensively in vision problems. Such an oversegmentation can be used to guide spectral sampling by assuming that the spectra is similar within each super pixel. Super pixelation is computationally light-weight and hence is amenable to video-rate processing. We used the Simple Linear Iterative Clustering (SLIC) algorithm [34] — a fast light-weight technique that is especially suitable for real-time applications.

To verify our hypothesis of similar spectra within a superpixel, we estimated the similarity between spectrum at one location and all its neighbors within the superpixels for some widely used datasets [22, 37–39]. Figure 3 visualizes the dissimilarity within each super-pixel in terms of spectral angular mapping (SAM). The average SAM value for any given HSI was less than 10° — corresponding approximately to 30 dB — which states that super pixels are a reliable way of separating images into homogeneous regions. Intuitively, it should then suffice to sample at one location within each super pixel Such a strategy reduces complexity of sampling by providing a single shot hyperspectral imager with extremely fast reconstruction.

3 The SASSI Camera

We first describe the SASSI system in detail, starting with the optical design, and the key processing steps for reconstruction of snapshot HSIs.

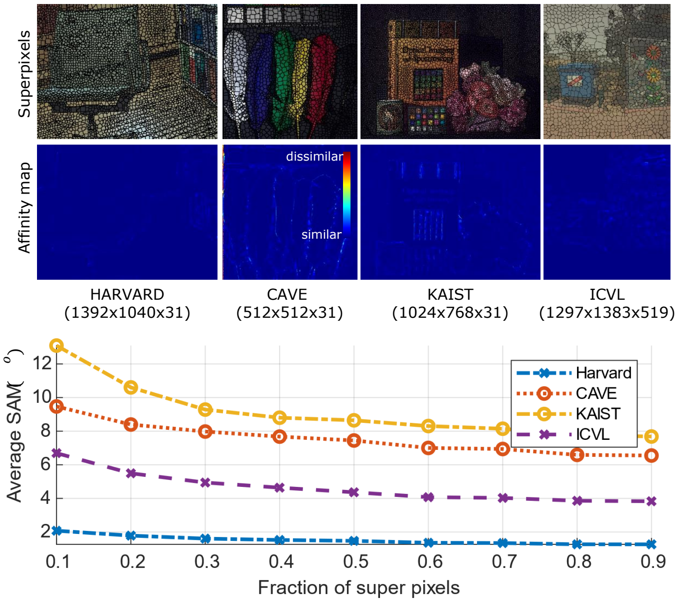


Fig. 3: **Homogeneity of super-pixels.** We hypothesize that super-pixels represent homogeneous regions of spectra. We estimated similarity between spectral profile at one location within a super-pixel and all its other members for commonly available datasets. We observed that spectral profiles inside a super-pixel are highly correlated, evident from the small Spectral Angular Mapping (SAM) value.

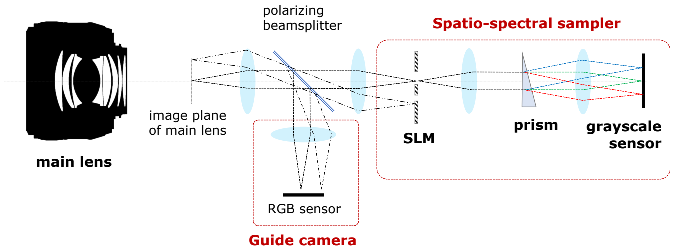


Fig. 4: **Schematic of the proposed setup.** Our optical setup consists of an RGB camera that guides the sampling system, and a spatio-spectral imager that consists of a spatial light modulator, a prism and a grayscale sensor. The guide image is utilized to generate a spatial mask that generates non-overlapping spectral profiles on the grayscale sensor. The guide image is then fused with the sparse spectral measurements to obtain a high spatial and spectral resolution HSI.

3.1 Optical schematic

Figure 4 shows a schematic of our optical setup, which consists of two sub-systems: the guide RGB camera and the spatio-spectral sampler that involves an SLM for spatial sampling, a prism for spectral dispersion, and finally a grayscale camera.

Given the HSI of the scene, $H(x, y, \lambda)$, we now describe the image formation process at both cameras. An image of the scene is focused by a main imaging lens and is then relayed to the RGB guide camera and SLM via a polarizing beamsplitter. The guide camera senses a color image of the scene, $I_{\text{RGB}}(x, y)$, whose intensities relate to $H(x, y, \lambda)$ via the sensor's spectral response. The SLM implements a transmission mask $M(x, y)$ which results in a modified HSI, $\widetilde{H}(x, y, \lambda) = H(x, y, \lambda)M(x, y)$. This signal then propagates

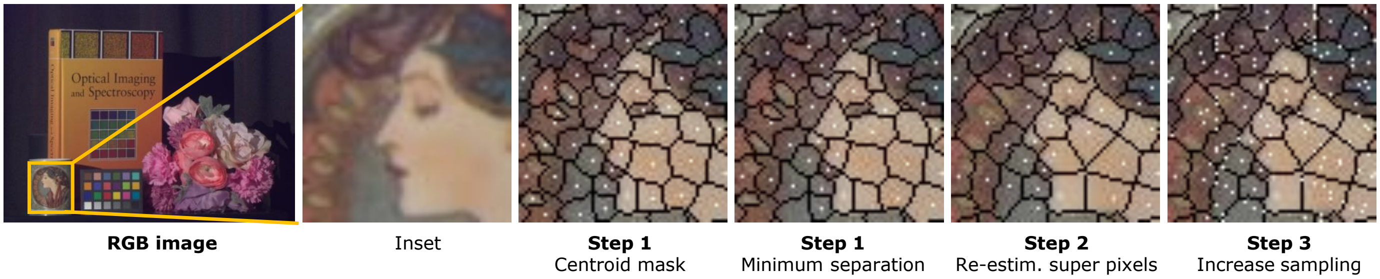


Fig. 5: **Super-pixel sampling strategy.** We follow a computationally simple, three step sampling strategy for estimating the mask. Given a super-pixel segmentation, we first create a mask with centroids of super-pixels as sampling locations. Then, we enforce minimum separation along horizontal direction by moving/removing sampling locations. Next, we re-estimate the super-pixels with the new sampling locations as centroids. Finally, we increase light throughput by creating sampling locations everywhere with minimum separation between neighbors. Such a sampling strategy requires less than 2 ms on a modern computer, ensuring real time mask generation.

through the beamsplitter and the prism which generates the following spectrally-spread image,

$$I_{\text{spec}}(x, y) = \int_{\lambda} \widetilde{H}(x - f(\lambda), y, \lambda)\; c(\lambda)\; d\lambda, \qquad (1)$$

where $f(\cdot)$ is a spectral shift induced by the prism and $c(\lambda)$ is spectral response of the camera. We assume the maximum spatial spread of spectrum to be N pixels. Note that (1) is the image formation model for the single-disperser CASSI system for a static attenuation mask.

Systems such as CASSI utilize a dense mask which leads to mixing of spectra at neighboring locations leading to an ill-conditioned inverse problem. However, if $M(x, y)$ is designed with a minimum spacing of N pixels between neighboring openings in the mask, we can ensure that each pixel in the image I_{spec} only measures light from a single spatial pixel of the SLM. The guide image acquired at time $t - 1$ is used to determine the mask at time t, $M(x, y, t)$, via super-pixelation and a judicious sampling of pixels given the super-pixels. The mask $M(x, y, t)$ is displayed on the SLM when acquiring the images at time t; these images, along with the mask, are fused to obtain the hyperspectral image at time t. We describe each step in detail next.

3.2 Super-pixel generation

Given the guide image, we perform super-pixelation with the computationally efficient simple linear iterative clustering algorithm (SLIC) [34, 40]. The SLIC algorithm essentially has two parameters, namely the number of super-pixels, Q and the *compactness*, C. Compactness controls the regularity of super-pixels; a larger value of C promotes more regularly shaped super-pixels, while smaller value ensures that the super-pixel boundaries adhere to edges in the image. Such a simple and effective parametrization is favorable to our mask generation scheme, and hence very crucial. The size and shape (compactness) of super-pixels depends on number of pixels over which spectrum is spread at each spatial point and hence can be effectively set.

3.3 Adaptive spatial sampling

Given the super-pixel image, we next efficiently generate a scene-adaptive spatial mask that enables and efficient sampling of the scene's HSI without spectral multiplexing. The generated mask aims to satisfy two key requirements:

1) Each super-pixel is sampled at least once, and
2) The sampling locations are horizontally separated by N pixels to ensure no overlap between the spectral smears arising from two different spatial locations.

We further seek to increase the light throughput by maximizing the number of sampling locations on the SLM. Generating an optimal mask is a computationally expensive problem. We instead rely on a simple multi-step approach to ensure real-time computations.

Step 1 – Initial centroid-based mask. We start by setting the sampling locations to super-pixel centroids, $\{(x_c^k, y_c^k)\}_{k=1}^{Q}$. This generates a mask $M_1(x, y)$ that has Q openings in total, which may be small but ensures that all super-pixels are sampled. To enforce the second constraint, we then modify the initial mask by moving/removing sampling locations till every location is separated by at least N pixels.

Step 2 – Resegmentation of super-pixels. The resultant mask after step 1 satisfies the second constraint but may violate the first constraint of one sample per superpixel – especially if the super-pixels are small. In order to satisfy the first constraint, we use the updated sampling locations from first step as the centroids and reestimate the super-pixels. In case of SLIC super-pixels, this involves running a simple local K-means clustering which is fast and efficient.

Step 3 – Increasing light transmission. Given a spatial resolution of $H \times W$ and a spectral spread of N pixels, the mask can have a maximum of $Q_{\max} = \frac{HW}{N}$ sampling locations. We followed a greedy approach to achieve this. Starting from left side of the mask, we open every location that is not sampled in a $2N - 1$ horizontal neighborhood. This is then repeated for all rows of the image till no more pixels can be opened. Figure 5 visualizes each step of the sampling strategy. Given a super-pixel segmentation, the whole process takes less than 2 ms on a mid-range PC.

3.4 Reconstruction techniques

We propose two approaches to estimate HSIs.

Local rank-1 approximation. We follow a "rank-1" approximation for each super-pixels, where we assume that spectrum at each spatial location is a scaled version of the sampled spectra within its super-pixel neighborhood. Let $I_{\text{gray}}(x, y)$ be the grayscale image of the scene obtained from the guide camera. We illustrate the reconstruction details for one super-pixel; the method is the same for all other super-pixels. Let $\{(x_p, y_p)\}_p$ be sampled locations within the l^{th} super-pixel, and let $S_p(\lambda) = H(x_p, y_p, \lambda)$ the sampled spectrum. The spectrum at an unsampled location (x_u, y_u) within this super-pixel is then computed as,

$$\widehat{H}(x_u, y_u, \lambda) = \frac{I_{\text{gray}}(x_u, y_u)}{\sum_{p=1}^{N_l} I_{\text{gray}}(x_p, y_p)} \sum_{p=1}^{N_l} S_p(\lambda) \qquad (2)$$

Such a reconstruction strategy is computationally inexpensive, and can be easily implemented on GPUs, making it an appealing approach for real-time reconstruction.

Learned guided filtering. When capturing measurements at video rate, the SNR is expected to be low, which gives rise to artifacts around superpixel boundaries. To overcome this, we rely on a data-driven approach for high quality reconstruction even under severely noisy measurements. The technique is similar in spirit to the well-known guided reconstruction techniques such as guided filtering [41], and more recently, a neural-network variant of the same [42].

Our neural network based reconstruction has two distinct components: first, a guided filter with learnable filters, and second, a simple 2-layer neural network.

Step 1 —– Guided filter. The first step estimates the spectral image as an affine scaling of the guide, i.e. $I^k_\lambda(x, y) \approx \alpha_k I^k_{\text{guide}} + \beta_k$ for the k^{th} image patch. The values of α_k and β_k are estimated on the fly by solving an optimization problem of the form,

$$\min_{\alpha_k, \beta_k} \|F \odot M^k \odot \left(I^k_{\text{input}} - \alpha_k I^k_{\text{guide}} - \beta_k\right)\|^2, \qquad (3)$$

where M^k is the spatial mask for the k^{th} patch, and F is a box function in traditional guided filtering. We observe that the box function is not sufficient when working with sampled data, and hence we make the parameters of F trainable as part of the neural network.

Step 2 —– Refinement. We then use the outputs of the guided filter as inputs to a three-layer neural network, which acts as a refinement layer to remove any artifacts. Other architectures based on guided filtering are possible, but we found this simple architecture sufficed for our purposes. Figure 6 visualizes the reconstruction pipeline.

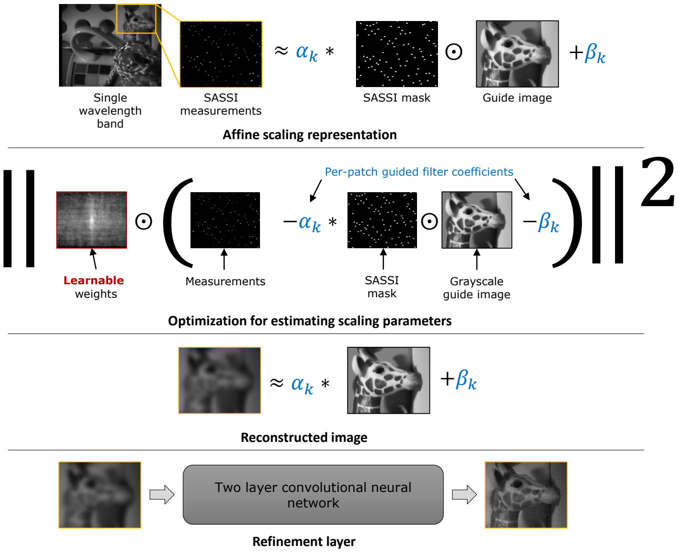


Fig. 6: **Visualization of guided filtering layer.** We rely on a simple modification to traditional guided filtering. We represent the measurements as an affine scaling of the grayscale guide image. Then we solve a *weighted* least squares problem to estimate the coefficients, which is then used to reconstruct image at each wavelength band. This is followed by a two-layer refinement stage to get the final output. The learnable weights are trained along the two-layer neural network with a composite loss consisting of SSIM and MSE loss functions.

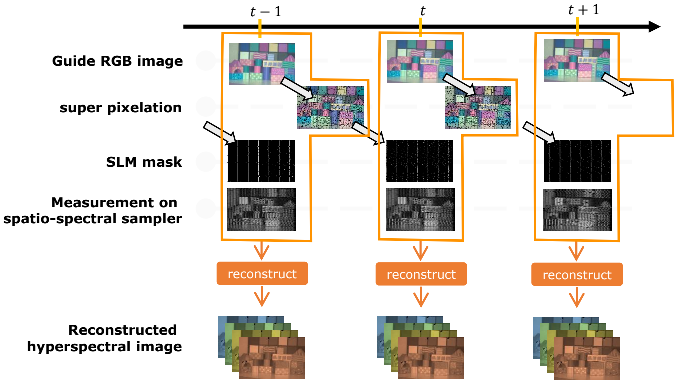


Fig. 7: **System pipeline.** Our method avoids temporal registration as it relies only on the information captured at time instance t. Specifically, the guide image at $t-1$ is used to create a mask, which is used to capture another guide image and the spatio-spectral image at time t. Instead of fusing guide image from $t-1$, we fuse images only from time instance t which ensures that there are no motion artifacts.

3.5 Reconstruction approach for video sequence

Since our approach requires fusion of two images, it is important that they capture measurements of the *same* scene. When capturing a video sequence, our optical system generates color images $I^t_{RGB}(x, y)$ which generates masks $M^t(x, y)$ at time instance t. This mask is used to measure the RGB and spectral images at time instance $t+1$, which is likely different from the scene at t due to motion. Instead of fusing $I^t_{\text{RGB}}(x, y)$ and $I^{t+1}_{\text{spec}}(x, y)$, which may lead to erroneous reconstruction, we combine $I^{t+1}_{\text{RGB}}(x, y)$ and $I^{t+1}_{\text{spec}}(x, y)$ — which ensures that both images are in lock-step. We note here that step 2 in the mask generation process, where the super-pixels are re-estimated is an important contributor to accurately fusing the two images. By generating mask from time instance t and using it to estimate super-pixels in time instance $t+1$, we avoid temporal registration which is hard to perform for an arbitrary scene. The timing of the system pipeline is illustrated in Figure 7.

4 Experiments

We first validate SASSI with simulations on a commonly available datasets, and then demonstrate results with the optical setup we built in our lab.

4.1 Simulations

We simulated our approach on standard HSI datasets including CAVE dataset [37], KAIST dataset [22], Harvard Dataset [38], and ICVL dataset [39], each with 31 spectral bands from $400 - 700$nm. For video rate results, we used data from [43] which comprised of 31 video frames, with each frame consisting of 33, 480×752 images sampled from $400 - 720$nm.

Modeling noise. We modeled our sensing camera to have a readout noise standard deviation of $5e^-$, a value typical to scientific cameras. We assumed various light levels, starting from $100e^-$ per pixel to $10,000e^-$ per pixels with the aim of showing performance variations with noise levels.

Training details. Our neural network consisted of 8 guided filter layers of 51×103 dimension each, followed by two convolutional layers along with ReLU non-linearity, each with 32 layers of 3×3 filters. Further details about the training process can be found in the supplementary material.

4.1.1 Number of superpixels

The number of super-pixels is upper bounded by $\frac{HW}{N}$; however, the specific number of super pixels that maximizes reconstruction accuracy depends on spatial complexity of the scene, and the noise levels. Intuitively, smaller super-pixels ensure that highly textured scenes are well sampled. However, each super-pixel will then get fewer samples, which is not desirable in low light conditions.

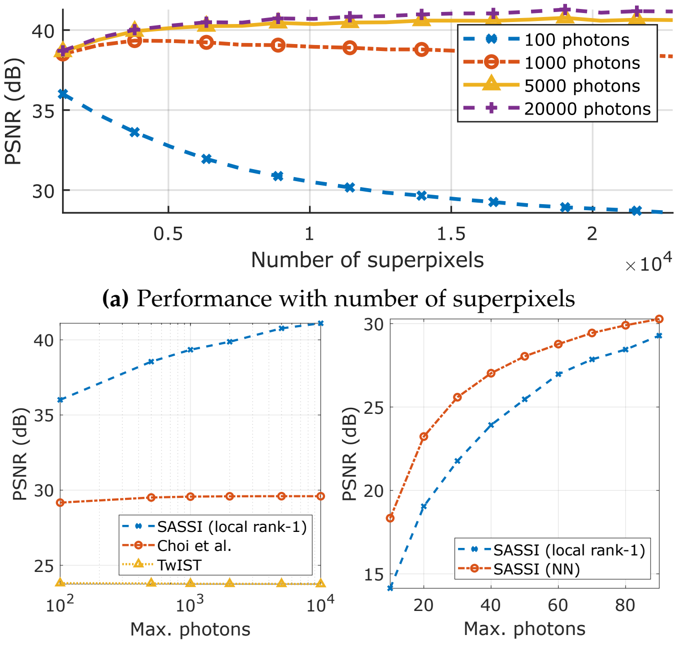
We empirically evaluated the effect of number of super-pixels under varying noise conditions on sample HSIs. Figure 8a showed accuracy as a function of number of super-pixels under varying light conditions. At low light, larger super-pixels are advantageous, while at higher light levels, smaller super-pixels lead to accurate results. In our real experiments, we found that $Q = \frac{1}{4}\frac{HW}{N}$ gave the best results for short exposure times of 50 ms or lower.

4.1.2 Comparisons to prior art

We simulated densely-sampled mask based reconstruction approaches with traditional total variation (TV) penalized reconstruction [44] and a more recent neural network based approach [22], denoted by TwIST and [Choi et al.], respectively. Figure 8 compares SASSI against [22] and reconstruction with TwIST [44] across varying light levels, and Fig. 9 visualizes the reconstructions. These methods do not utilize an extra guide image. We also compared against guide image based reconstruction approach by [Cao et al.] [28] where the HSI is uniformly and sparsely sampled and reconstruction is done through a modified bilateral filtering. Among methods that do not utilize a guide image, [22] had superior performance in terms of visual quality as well as reconstruction SNR. When using a guide image, SASSI outperforms other approaches by more than 2dB.

4.2 Experiments with a lab prototype

We first provide details of our optical setup and then demonstrate results in various settings.



(a) Performance with number of superpixels

(b) Average to high intensity

(c) Low light levels

Fig. 8: **Performance with number of superpixels and noise.** (a) As noise increases, a small number of superpixels is advantageous, as the spectra within a superpixel get averaged. (b) Our rank-1 reconstruction performs well against prior work under average to high light intensity. (c) At lower light levels, our guided filtering approach outperforms the rank-1 approach.

4.2.1 Optical setup

Figure 10 shows an image of the optical setup. We used a FLIR Blackfly BFLY-U3-16S2C-CS color camera as guide camera, and Hamamatsu Orca Flash 4.0 as spatio-spectral camera, Holoeye HES 6001 LCoS display as SLM operating at 60fps, and a a 30° prism for dispersion. Additional details can be found in the supplementary material.

System properties. Our setup was optimized to image from $400 - 700$ nm, with spectrum spread over 68 pixels on the spatio-spectral camera. Due to non-linear dispersion of the prism, we obtain a spectral resolution of 2 nm at 400nm and 10 nm at 700 nm. All the HSIs captured in the upcoming sections were captured at a spatial resolution of 600×900 pixels. Detailed calibration steps required to scan HSIs are provided in the supplementary material.

Removing offset in spectral image. Due to finite contrast ratio of SLMs, the spatial mask can be effectively written as $M_{\text{real}} = (1 - \epsilon)M_{\text{ideal}} + \epsilon$, where ϵ is the non-zero throughput when the SLM pixel is set to 0. This leads to a large background offset in the measured spatio-spectral image, as illustrated in Fig. 11(a). This can be compensated by capturing an image with an all-zero mask, like Fig. 11(f); however this requires an additional capture. Instead we observed that the offset image is smooth and estimated it from the spatio-spectral image, by first identifying and removing the spectral streaks, and interpolating the missing pixels using cubic interpolation. This process is visually illustrated in Figure 11.

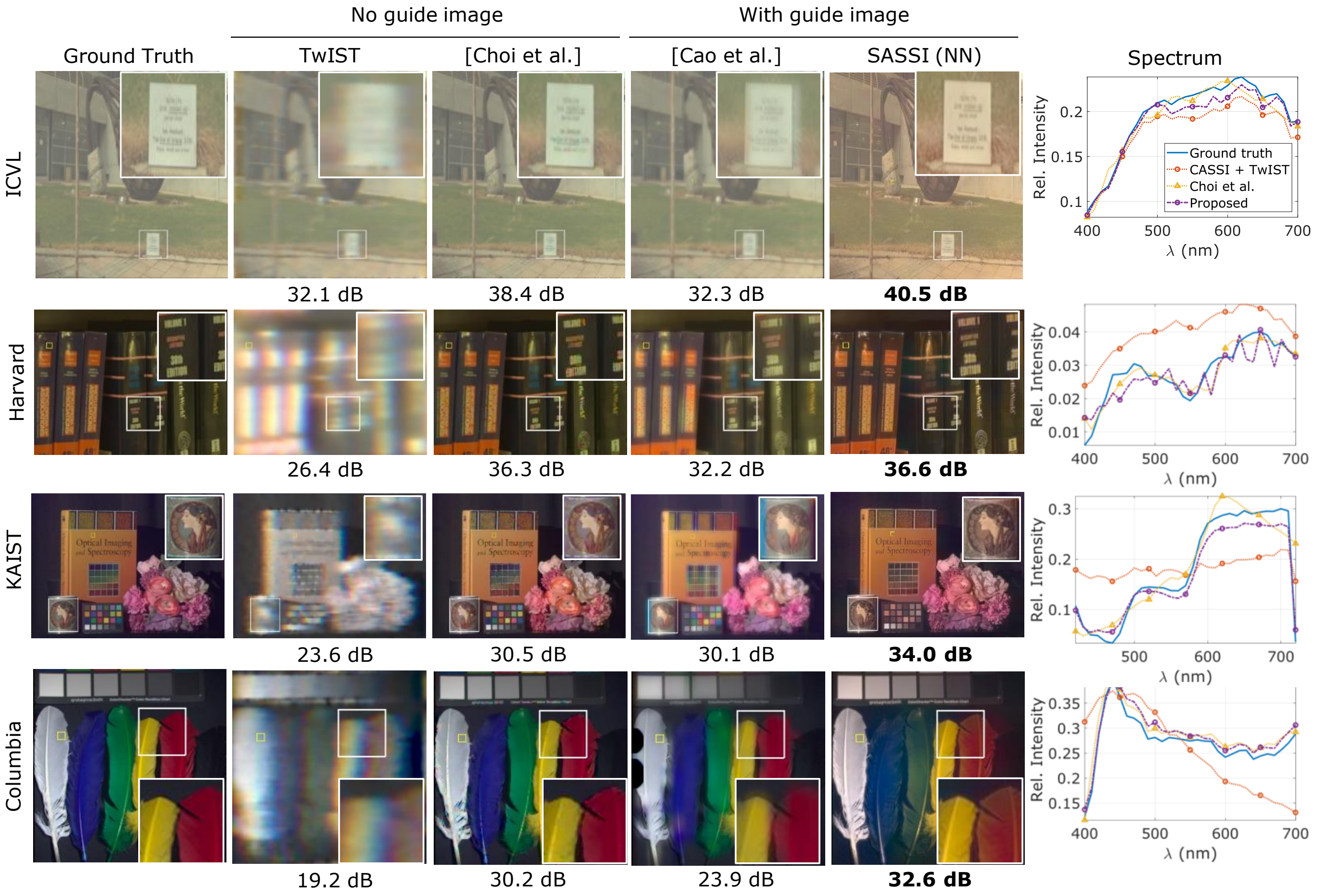


Fig. 9: **Comparisons with CASSI techniques.** We compared reconstruction approaches which do not use a guide image [22, 44], and ones that use guide image [28]. SASSI outperforms other techniques across the board.

Nyquist scanning. To obtain a full scan of scene's HSI, we implement a parallelized version of pushbroom scan where multiple pinhole arrays are displayed on the SLM. The horizontal spacing of the pinhole array was kept at 100 pixels, and the verticle spacing was 5 pixels. We then scanned a total of 500 images, with 100 horizontal shifts and 5 vertical shifts. Figure 2 shows an example of the mask that is used as well as the corresponding measurement in this sampling process.

4.2.2 Dataset details

We collected several hyperspectral images of various objects, under varying illumination conditions. A snapshot of the training data is shown in Fig. 12. Each capture included a full scan of the scene, along with a guide RGB image, and the corresponding super-pixel sampled image. We retrained the guided filter on this dataset for finetuning it to the specifics of the prototype. Further details about the training parameters can be found in the supplementary material.

4.2.3 Results

Unless specified, each measurement involved a single image captured by the guide and the spatio-spectral cameras.

Comparison of sampling strategies. Figure 13 compares adaptive, random, and uniform sampling on a real scene. SASSI captured the intricate spatial details such as the lips on the face, and small dots on the dress, making a compelling case for adaptive sampling.

Comparison of reconstruction approaches. Figure 14 shows guide image, super pixelation, captured spatio-spectral image and an RGB visualization of reconstruction with rank-1 and guided filters approach. Images were captured at an exposure rate of 33ms. The second row shows spectra at the marked location, and the third row visually compares ground truth (right) and guided filters reconstruction (left). The accuracy of rank-1 reconstruction was 41.5dB with an SSIM of 0.98, whereas the guided filters approach had an accuracy of 42.6dB with SSIM of 0.99. We observed the performance to be similar across all our real experiments. This is expected, as the guided filters approach is particularly well suited for low light levels. The last row shows a reconstruction at 501nm for both ground truth (left) and our method (right).

Static scenes. Figure 15 visualizes reconstruction across a variety of geometries, spectral profiles, and illumination conditions. None of the scenes shown in the figure overlap with the training data for neural networks. We observe the importance of per-band reconstruction in the third row – since we only learn on images, and not on the full HSI, we were able to reconstruct complex spectra such as a scene illuminated with LED lamp. We also tested our system on

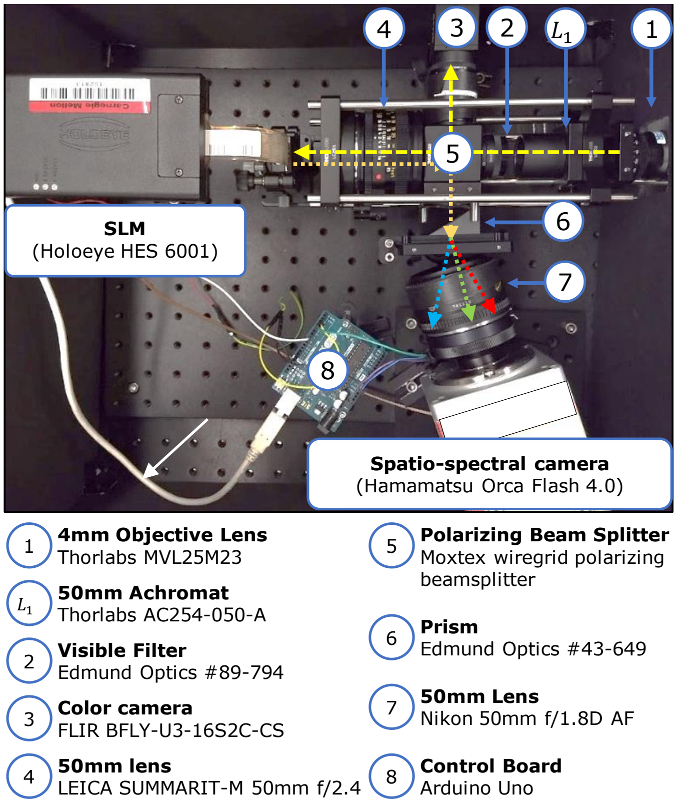


Fig. 10: **Lab prototype.** The image above shows a photograph of our lab setup with key components marked. We also showed an overlay of ray tracing for easy understanding.

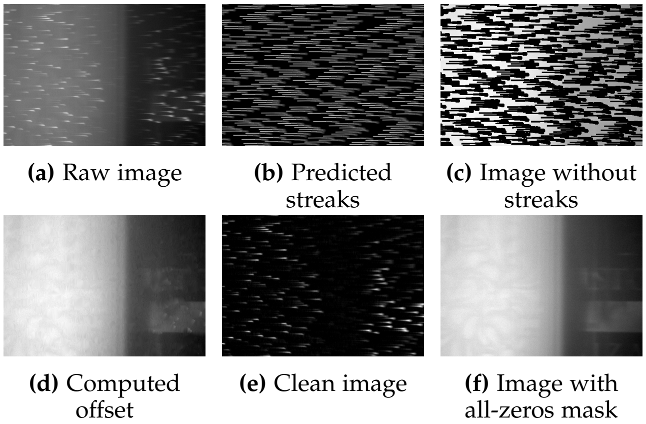

(a) Raw image
(b) Predicted streaks
(c) Image without streaks
(d) Computed offset
(e) Clean image
(f) Image with all-zeros mask

Fig. 11: **Removing background offset.** Since the SLM has finite contrast ratio, the resultant mask has non-zero values everywhere, leading to a background offset shown in (a). We rely on (b) calibration information along with interpolation to predict the offset image in (d), leading to a cleaned image in (e). Notice that the offset image is close to an image captured with all-zeros spatial mask.

microscopic specimens, specifically a small tissue of a dog's esophagus. Notice the difference in intensities of the cellular wall across various wavelengths. Figure 16 shows a plot of error vs. distance from nearest sampling point, showing a graceful decrease in accuracy. Across the board, we observed that our sampling and reconstruction approach gave high quality results, with PSNR exceeding 30dB compared to a full scan of the scene.

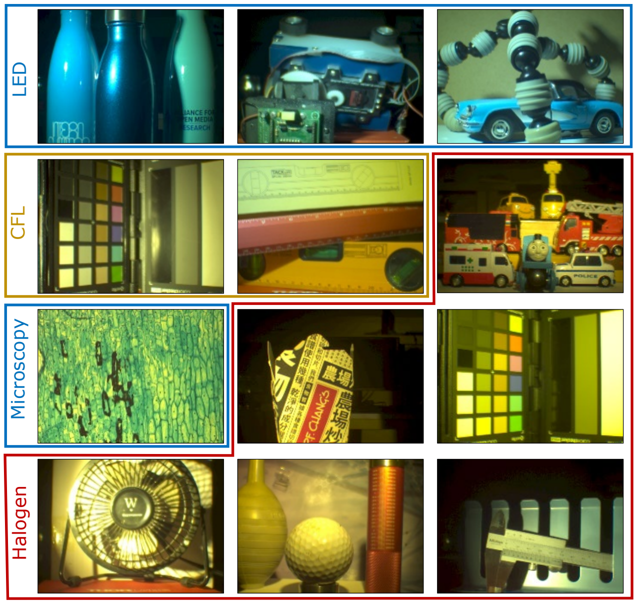


Fig. 12: **Training data for learning guided filters.** We collected more than 50 hyperspectral images with various objects and illuminants. We then used 18 for training the neural network. None of the training images were used in the testing phase.

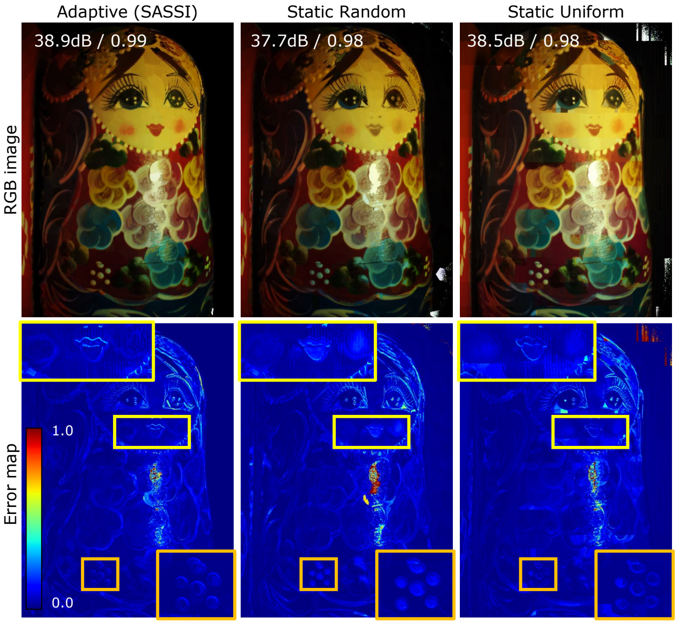
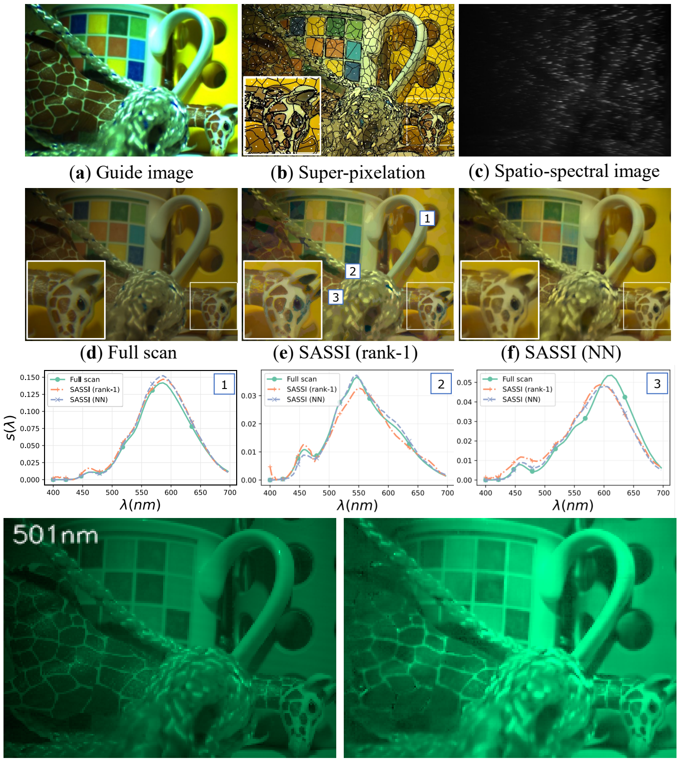


Fig. 13: **Effect of sampling strategy.** Adaptive sampling captures intricate spatial profiles such as the lips, and the small dots in the figurine.

Dynamic scenes. SASSI is most impactful when capturing video rate HSIs of scenes with complex geometry and a rich gamut of spectral profiles. Our setup captured HSIs at a rate of 18 frames per second. This capture rate was primarily limited by the vagaries associated with our SLM, and not an upper limit on the method itself. Timing analysis of each step of our algorithm is listed in Tab. 1.

Figure 1 shows an example of two Matroshka dolls rotating on a turnstile. We showed a frame from the video sequence of three spectral bands over 100 time frames. Full


(a) Guide image (b) Super-pixelation (c) Spatio-spectral image

(d) Full scan (e) SASSI (rank-1) (f) SASSI (NN)

Fig. 14: **Comparison of rank-1 and NN reconstructions.** Reconstruction results on a scene containing color full object. Shown are (a) guide image, which is used to generate (b) super pixels and sampling pattern (white dots), and subsequently (c) the spatio-spectral measurements. In (d-e) and the plots, we compare reconstructions with a rank-1 and NN approaches along with full pushbroom scan. The last row shows image at 501nm for the full scan and the NN reconstruction. Visualization of the full hyperspectral image can be accessed from supplementary material.

TABLE 1: **Timing per frame.** We implemented a highly optimized pipeline for video hyperspectral imaging. Shown below are the timing for the key steps in our acquisition pipeline. The numbers provided here were observed on a Dell Alienware computer with 9^{th} Gen Intel Core i7, 32GB RAM, and NVIDIA GeForce GTX 1660 Ti.

Step	Time taken	Remarks
Capture guide image	15ms	Depends on exposure time of guide camera
Super pixelation	15ms	Depends on number of super pixels. Can be made faster with GPU
Mask generation	< 1.5ms	Fast C/Python implementation
Displaying on SLM	< 1 ms	Bottlenecked by SLM refresh rate
Capture spatio-spectral image	< 1ms	Camera simultaneously exposes during other computations

video sequence can be accessed from the supplementary material. The blue, and red colors on the dolls are distinctly visible in the spectral bands, along with accurate texture reconstruction. Figure 17 shows frames from two more video reconstructions. Both scenes show highly dynamic motions that are challenging to capture; our reconstructions indicate successful reconstructions of the hyperspectral video.

Spectral reflectance. The high resolution and snapshot capabilities of SASSI are useful when we want to measure higher dimensional slices of the plenoptic function such as the spectral bidirectional reflectance distribution function. The setup used for this purpose is shown in Fig. 18 (a), which consisted of the SASSI camera and a dome of twelve white LED sources well spread over the hemisphere. We imaged a butterfly, which exhibits rich structural coloration as the angle of illumination changed. The spatial images as well as spectral reflectances are shown in Fig. 18 (c) through (f). Certain illumination conditions, such as view 2, 4 showed strong blue peaks, whereas view 5 produced a brown colored image, which reveals the eye-like structure under the wing.

Using this data, we decomposed the measured spectrum into a basis that separated the blue iridescent reflectance from the brown underlying reflectance using non-negative matrix factorization. This provided us with a set of spectral feature vectors and corresponding vector weights combining these features at every point. We first applied this only to the spectral dimension, with a 2-dimensional decomposition. The results are shown in Fig. 19(a). We observed that the dominant spectral features consisted of a high intensity blue spectrum, and a low intensity brown spectrum. These angles match the view points in Fig. 18 that exhibited deep blue color and are consistent across different parts of the butterfly that show iridescence at the same angles.

We then decomposed the HSIs along spatial and illumination directions. By separating these features into iridescent and non-iridescent components, we were able to decompose an image into two separate images corresponding to these components. The results of this decomposition are shown in Fig. 19 (b). Using only one image per illumination angle, we are were able to separate the blue iridescent reflectance from the brown underlying reflectance, revealing the patterns on the butterfly's wing.

5 Conclusion

Our paper showed that adapting the sampling patterns to the specific instance of a scene can significantly reduce measurement time, without sacrificing spatial, or spectral resolutions. By making hyperspectral imaging faster and robust, SASSI opens up novel applications in material identification, biological imaging, and computer vision tasks. It also opens up the possibility for measuring higher dimensions of light involving spectrum, angle and polarimetry that has hitherto been hard to sense. We hence believe the ideas presented in this paper will push the boundaries of plenoptic imaging and its associated applications.

6 Acknowledgments

The authors acknowledge support via the National Geospatial Intelligence Agency's Academic Research Program (Award No. HM0476-17-1-2000), the NSF SaTC award 1801382, and the NSF Expeditions in Computing Award 1730147, 1730574.

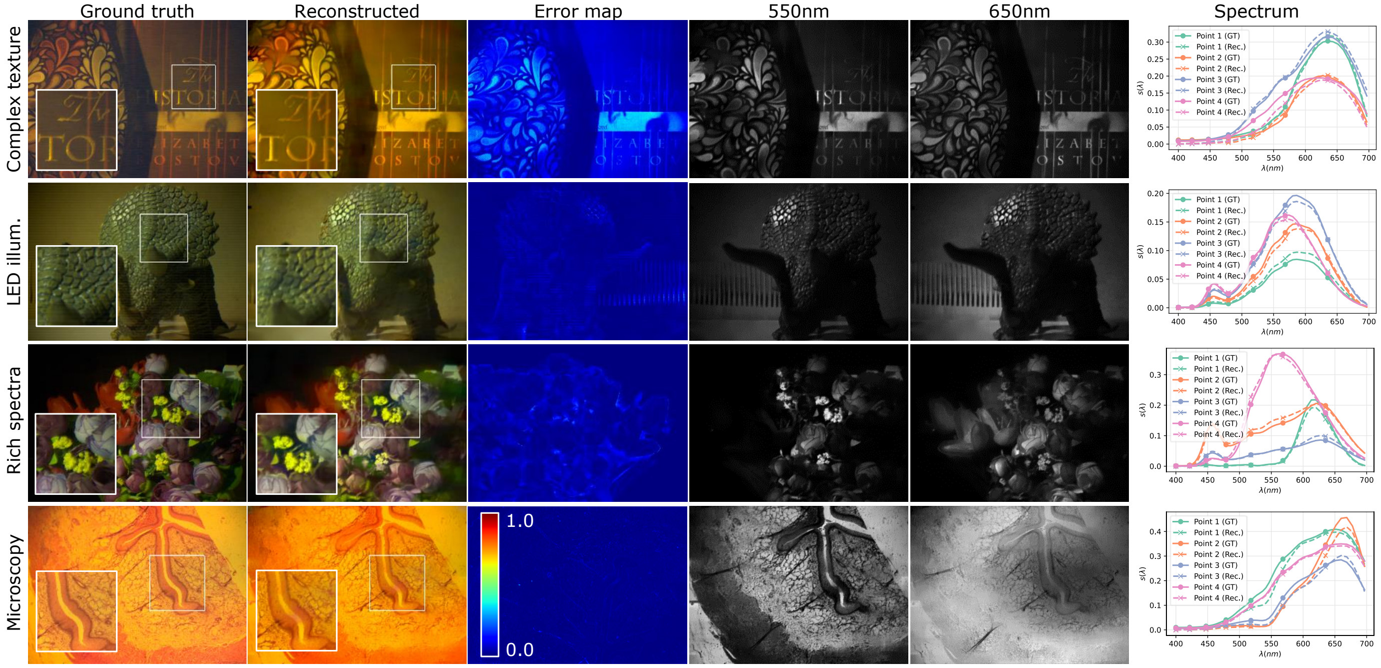


Fig. 15: **Real experiments.** The figure showcases results across a variety of scenes with varying scene, spectrum, and illumination complexity. Across the board, our sampling and reconstruction approach produces accurate results with reconstruction accuracy exceeding 32dB.

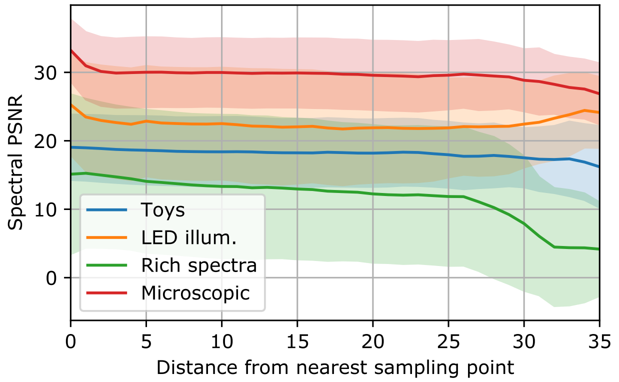


Fig. 16: **Accuracy vs. distance.** Since we sample sparsely, we expect the accuracy to be lower with increasing distance between spatial point and nearest sampling point. This is reflected across all our experiments.

References

[1] B. E. Bayer, "Color imaging array," 1976, US Patent 3,971,065.

[2] V. Gruev, R. Perkins, and T. York, "CCD polarization imaging sensor with aluminum nanowire optical filters," *Optics Express*, vol. 18, no. 18, pp. 19 087–19 094, 2010.

[3] M. Hirsch, S. Sivaramakrishnan, S. Jayasuriya, A. Wang, A. Molnar, R. Raskar, and G. Wetzstein, "A switchable light field camera architecture with angle sensitive pixels and dictionary-based sparse coding," in *IEEE Intl. Conf. Computational Photography*, 2014.

[4] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford Tech Report CTSR 2005-02, Tech. Rep., 2005.

[5] G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, "Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging," *Nature Methods*, vol. 7, no. 3, p. 209, 2010.

[6] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2000.

[7] S. G. Narasimhan and S. K. Nayar, "Enhancing resolution along multiple imaging dimensions using assorted pixels," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 518–530, 2005.

[8] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[9] V. Saragadam, M. DeZeeuw, R. G. Baraniuk, A. Veeraraghavan, and A. C. Sankaranarayanan, "SASSI: Super-pixelated adaptive spatial spectral imager," https://github.com/Image-Science-Lab-cmu/SASSI.

[10] E. Cloutis, "Review article hyperspectral geological remote sensing: Evaluation of analytical techniques," *International J. Remote Sensing*, vol. 17, no. 12, pp. 2215–2242, 1996.

[11] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature Methods*, vol. 2, no. 12, pp. 910:1–11, 2005.

[12] T. Zhi, B. R. Pires, M. Hebert, and S. G. Narasimhan, "Multispectral imaging for fine-grained recognition of powders on complex backgrounds," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2019.

[13] M. Goel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, G. Borriello, and S. N. Patel, "Hypercam: Hyperspectral imaging for ubiquitous computing applications," in *ACM Intl. Joint Conf. Pervasive and Ubiquitous Computing*, 2015, pp. 145–156.

[14] V. Saragadam and A. C. Sankaranarayanan, "Programmable spectrometry–per-pixel classification of materials using learned spectral filters," *IEEE Intl. Conf. Computational Photography*, 2020.

[15] Z. Hui, K. Sunkavalli, S. Hadap, and A. C. Sankaranarayanan, "Illuminant spectra-based source separation using flash photography," in *IEEE Intl. Conf. Computer*

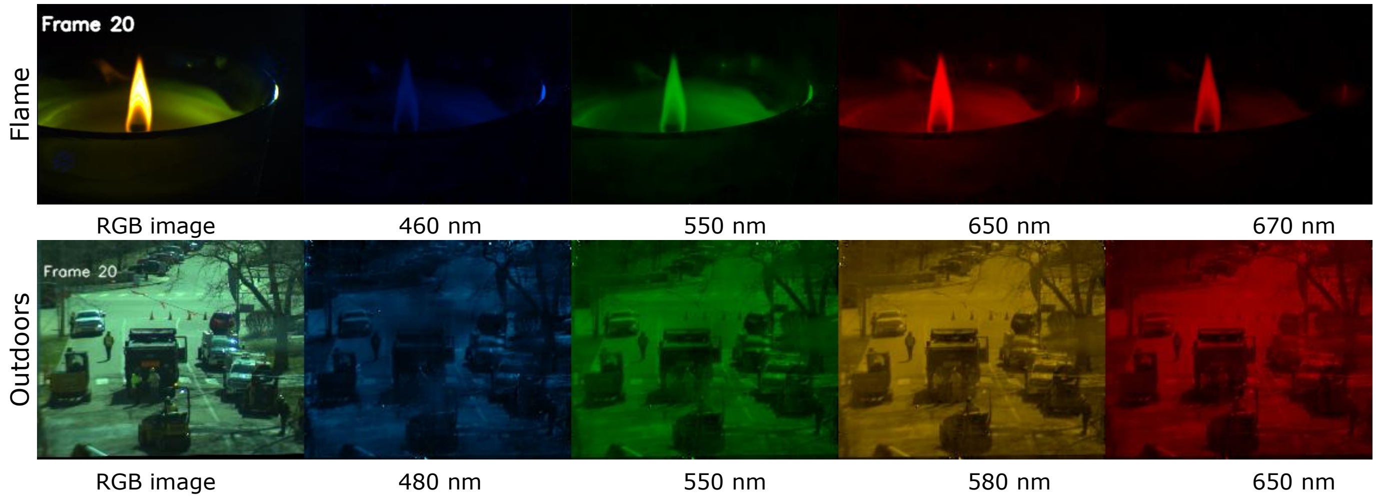
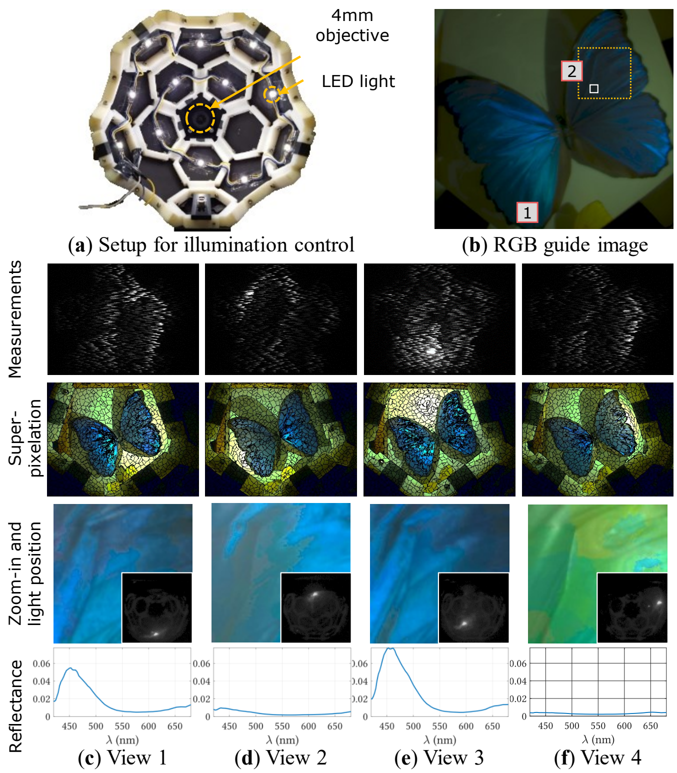


Fig. 17: **Video rate reconstruction.** We imagined challenging scenes such as a live flame, and an outdoor scene, both consisting of complex texture, spectrum, and temporal dynamics. Our system was able to accurately recover the HSIs at each time-frame. Full video sequence can be accessed from supplementary material.



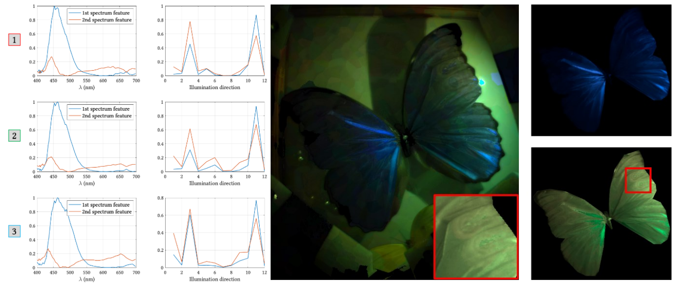
(**a**) Setup for illumination control (**b**) RGB guide image

(**c**) View 1 (**d**) View 2 (**e**) View 3 (**f**) View 4

Fig. 18: **Sampling spectral BRDF.** SASSI allows us to sample higher dimensional visual signals such as spectral BRDF. We combined our setup with a light dome to estimate spectral BRDF of a butterfly, illuminated in different directions by a white LED light source. We then captured HSIs for each illumination condition (see inset image for direction of illumination) and estimated spectrum on a small patch. We observed that the spectrum is dependent on illumination condition, as evident from deep blue color in (e) view 4 and brown shade in (f) view.

Vision and Pattern Recognition, 2018.

[16] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Optics*, vol. 49, no. 36, pp. 6824–6833, 2010.



(**a**) Spectral bases (**b**) Spectral decomposition

Fig. 19: **Decomposition of spectral BRDF.** (a) shows a plot of rank-2 decomposition of spectra at the three marked points in Fig. 18, with the blue plot representing the iridescent blue spectrum, and the red plot representing the background, broadband spectrum. A spectral and angular decomposition of the object enabled us to separate the scene into iridescent and non-iridescent components, all achieved with only one image captured per angle.

[17] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graphics*, vol. 33, no. 6, pp. 233:1–11, 2014.

[18] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 105–115, 2014.

[19] H. Arguello and G. R. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," *IEEE Trans. Image Proc.*, vol. 22, no. 3, pp. 941–954, 2013.

[20] H. Rueda, H. Arguello, and G. R. Arce, "Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays," *Appl. Optics*, vol. 55, no. 33, pp. 9584–9593, 2016.

[21] ——, "High-dimensional optimization of color coded apertures for compressive spectral cameras," in *European Signal Proc. Conf.*, 2017.

[22] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graphics*, vol. 36, no. 6, pp. 218:1–13, 2017.

[23] C. Yang, F. Cao, D. Qi, Y. He, P. Ding, J. Yao, T. Jia, Z. Sun,

and S. Zhang, "Hyperspectrally compressed ultrafast photography," *Physical Review Letters*, vol. 124, no. 2, p. 023902, 2020.

[24] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2015, pp. 4942–4950.

[25] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery via non-iterative fusion of dual-camera compressive hyperspectral imaging," *arXiv Preprint arXiv:2012.15104*, 2020.

[26] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[27] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2011.

[28] X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2011.

[29] I. Kauvar, S. J. Yang, L. Shi, I. McDowall, and G. Wetzstein, "Adaptive color display via perceptually-driven factored spectral projection." *ACM Trans. Graphics*, vol. 34, no. 6, pp. 165:1–10, 2015.

[30] M. O'Toole and K. N. Kutulakos, "Optical computing for fast light transport analysis." *ACM Trans. Graphics*, vol. 29, no. 6, pp. 164:1–12, 2010.

[31] V. Saragadam and A. Sankaranarayanan, "KRISM—krylov subspace-based optical computing of hyperspectral images," *ACM Trans. Graphics*, vol. 38, no. 5, pp. 148:1–14, 2019.

[32] X. Fang, J. Feng, and Y. Wang, "Texture-adaptive hyperspectral video acquisition system with a spatial light modulator," in *Optoelectronic Imaging and Multimedia Technology III*, 2014.

[33] C. Ma, X. Cao, R. Wu, and Q. Dai, "Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system," *Optics letters*, vol. 39, no. 4, pp. 937–940, 2014.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[35] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Intl. J. Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[36] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.

[37] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Proc.*, vol. 19, no. 9, pp. 2241–2253, 2010.

[38] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2011.

[39] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conf. Computer Vision*, 2016.

[40] A. Kim, "FastSLIC: Optimized slic superpixels," https://github.com/Algy/fast-slic.

[41] K. He, J. Sun, and X. Tang, "Guided image filtering," in *European Conf. Computer Vision*, 2010.

[42] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Fast end-to-end trainable guided filter," in *IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2018.

[43] A. Mian and R. Hartley, "Hyperspectral video restoration using optical flow and sparse coding," *Optics Express*, vol. 20, no. 10, pp. 10 658–10 673, 2012.

[44] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Proc.*, vol. 16, no. 12, pp. 2992–3004, 2007.

Vishwanath Saragadam received his Ph.D. from Carnegie Mellon University, Pittsburgh and is currently a postdoctoral researcher at the DSP group at Rice University. His research interests include hyperspectral imaging, thermal imaging, computational photography, compressive sensing and computer vision. He is the recipient of the Future Faculty Fellowship at Rice in 2020, A.G. Jordan outstanding thesis award in 2020, the Prabhu and Poonam Goel graduate fellowship in 209, the outstanding TA award from the ECE Department at CMU in 2018, and Siemens award for highest GPA in the EE department in IIT Madras in 2014.

Michael De Zeeuw received his B.S. in Electrical and Computer Engineering from Calvin University in 2018. He is currently pursuing his Ph.D. in ECE at Carnegie Mellon University. He received the Pittsburgh Chapter Award from the Achievement Rewards for College Scientists Foundation in 2018. His research interests include efficient shape estimation, reflectance capture and modeling.

Richard Baraniuk received the B.S. from the University of Manitoba, Canada (1987), the M.S. from the University of Wisconsin-Madison (1988), and the Ph.D. degree from the University of Illinois at Urbana-Champaign (1992), all in electrical engineering. He is currently the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University and the Founding Director of OpenStax (openstax.org). His research interests lie in new theory, algorithms, and hardware for sensing, signal processing, and machine learning. He is a Fellow of the American Academy of Arts and Sciences, National Academy of Inventors, American Association for the Advancement of Science, and IEEE. He has received the DOD Vannevar Bush Faculty Fellow Award (National Security Science and Engineering Faculty Fellow), the IEEE James H. Mulligan, Jr. Education Medal, and the IEEE Signal Processing Society Technical Achievement, Education, Best Paper, Best Magazine Paper, and Best Column Awards. He holds 35 US and 6 foreign patents.

Ashok Veerarghavan is a Professor of Electrical and Computer Engineering at Rice University. His research interests are broadly in the areas of imaging, vision, and machine learning and their applications to health and neuroengineering.

Aswin C. Sankaranarayanan (SM'17) is an Associate Professor in the ECE Department at Carnegie Mellon University (CMU). He earned his Ph.D. from University of Maryland, College Park and was a post-doctoral researcher at the DSP group at Rice University before joining CMU. Aswin's research spans topics in imaging, vision, and image processing. He is the recipient of the CVPR 2019 best paper award, the CIT Dean's Early Career Fellowship in 2018, the NSF CAREER award in 2017, the Eta Kappa Nu (Sigma chapter) Excellence in Teaching award in 2017, the 2016 Herschel M. Rich invention award, and the distinguished dissertation fellowship from the ECE Department at University of Maryland in 2009.