

MantissaCam: Learning Snapshot High-dynamic-range Imaging with Perceptually-based In-pixel Irradiance Encoding

Haley M. So, Julien N. P. Martel, Piotr Dudek, and Gordon Wetzstein

Abstract—The ability to image high-dynamic-range (HDR) scenes is crucial in many computer vision applications. The dynamic range of conventional sensors, however, is fundamentally limited by their well capacity, resulting in saturation of bright scene parts. To overcome this limitation, emerging sensors offer in-pixel processing capabilities to encode the incident irradiance. Among the most promising encoding schemes is modulo wrapping, which results in a computational photography problem where the HDR scene is computed by an irradiance unwrapping algorithm from the wrapped low-dynamic-range (LDR) sensor image. Here, we design a neural network-based algorithm that outperforms previous irradiance unwrapping methods and we design a perceptually inspired “mantissa,” or log-modulo, encoding scheme that more efficiently wraps an HDR scene into an LDR sensor. Combined with our reconstruction framework, MantissaCam achieves state-of-the-art results among modulo-type snapshot HDR imaging approaches. We demonstrate the efficacy of our method in simulation and show benefits of our algorithm on modulo images captured with a prototype implemented with a programmable sensor.

Index Terms—computational photography, programmable sensors, in-pixel intelligence, end-to-end optimization

1 INTRODUCTION

HIGH Dynamic Range (HDR) imaging is crucial for a vast range of applications, including automotive vision systems [1], HDR display [2], and image processing [3], [4]. When capturing natural scenes, which can have an extreme high dynamic range [3], the level of detail is limited by the full well capacity and the quantization precision of the sensor. Unfortunately, the dynamic range offered by modern sensors is far smaller than that encountered in the wild [5], making specialized sensors or computational photography approaches to HDR imaging necessary.

Among the many HDR imaging techniques proposed in the literature, exposure bracketing [6], [7], [8], [9], [10], [11] and temporally varying exposures [12], [13], [14] can be successful, but fast motion introduces ghosting. Multi-sensor approaches [15], [16], [17] can overcome this limitation, but are expensive, bulky, and difficult to calibrate. Existing snapshot HDR imaging approaches hallucinate saturated image detail using neural networks [18], [19], [20], [21], [22], use spatially varying pixel exposures which trade spatial resolution for dynamic range [23], [24], [25], [26], [27], [28], [29], [30], or use optical encoding approaches that blur the sensor image [31], [32], [33]. Specialized sensors, for example recording logarithmic irradiance [34] or floating point extended dynamic range values [35] have also been proposed, but these either trade extended dynamic range for precision or require additional bandwidth.

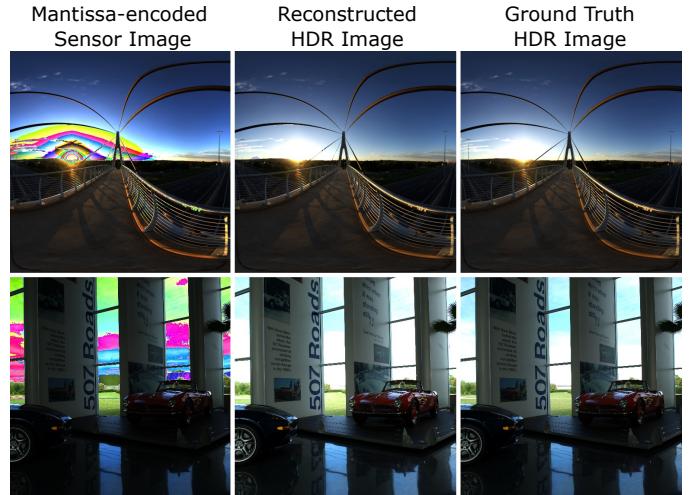


Fig. 1: MantissaCam electronically encodes the irradiance incident on the sensor into an LDR image by wrapping the intensity in a perceptually inspired manner (left). The proposed reconstruction algorithm estimates the HDR scene from this LDR image (center) and achieves accurate reconstructions compared to the ground truth (right).

- H. M. So, J. N. P. Martel, and Professor G. Wetzstein are in the Department of Electrical and Engineering at Stanford University, Stanford, CA, 94305, United States.
- Professor P. Dudek is in the School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom.
- <https://www.computationalimaging.org/publications/mantissacam>

Our work (Fig. 1) is inspired by the idea of electronically applying a modulo encoding of the irradiance on the sensor followed by an intensity unwrapping algorithm [36], [37]. This idea is beneficial over other snapshot approaches because it does not degrade a low-dynamic-range (LDR) image, as optical encoding approaches do, it does not hallucinate detail but recovers them, it does not decrease image resolution, or increase the required bandwidth. As we show

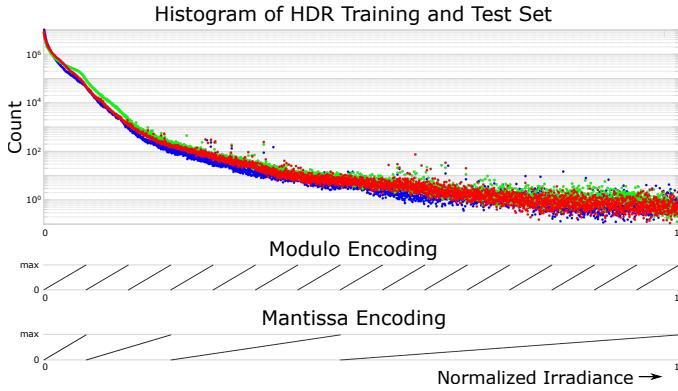


Fig. 2: Log histogram of normalized irradiance values of all pixels in our training and test sets of HDR images for all color channels (top). This histogram is highly biased towards low-intensity values, indicating that irradiance values of natural images are not uniformly distributed. Yet, the modulo encoding subdivides this intensity range uniformly and wraps each of these areas into the available dynamic range of the sensor, as shown for a 1D ramp (center). The proposed mantissa encoding wraps the same 1D ramp in a perceptually more uniform manner in log space, which is observed as non-uniform wrapping in irradiance space (bottom).

in this paper, there are several downsides to the modulo camera, as proposed in prior work. Specifically, modulo wrapping is done directly in irradiance space, which allocates precision and number of wraps linearly in this domain. However, the human visual system is perceptually approximately linear in the log-domain, so a conventional modulo encoding wastes precision for detail that we do not perceive. Moreover, the irradiance distribution of natural scenes is heavily skewed towards darker values (see log-histograms in Fig. 2), so it makes sense to nonlinearly distribute the irradiance wraps in order to minimize their number, because they have to be computationally unwrapped again.

We address these challenges by proposing a perceptually inspired modulo-type wrapping scheme that operates in the log-irradiance domain. This idea intuitively combines the principles of operation of both log [34] and modulo [36] cameras. Indeed, the signal we propose to measure is essentially a generalization of the mantissa used by the IEEE Standard for Floating-Point Arithmetic [38], or the log base 2 of the intensity modulo the well capacity. So instead of taking the modulo of the signal, we first take the log, and then the modulo. We demonstrate that such a log-modulo or *mantissa* camera allocates precision in a perceptually meaningful manner and it nonlinearly distributes the wraps in irradiance space to better match the distribution of irradiance values in natural scenes (see Fig. 2, top). This directly leads to fewer wraps of natural scenes (see Figs. 2, center and bottom, and 3), which make the inverse problem of 2D irradiance unwrapping easier to solve. To solve the unwrapping problem, we introduce a neural network architecture that is more robust than prior work using graph cut algorithms [36] or other network architectures [37]. Finally, we prototype a modulo camera

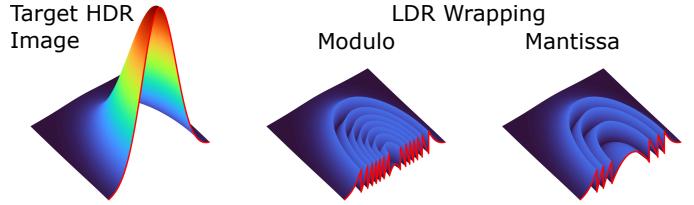


Fig. 3: Example showing an HDR Gaussian function wrapped using the modulo and mantissa encoding in an LDR image. For this example, the modulo encoding requires more wraps than the mantissa encoding, which makes its reconstruction via computational unwrapping more challenging.

using a SCAMP-5 programmable sensor [39] which allows for flexible re-configuration of the in-pixel irradiance encoding in software. These types of programmable sensors are expected to be widely available in the near future.

Specifically, we make the following contributions

- We introduce MantissaCam as a new snapshot approach to HDR imaging, combining perceptually motivated irradiance encoding and decoding.
- We develop a neural network architecture that outperforms existing unwrapping methods for modulo cameras and that demonstrates state-of-the-art performance with our mantissa encoding.
- We build a prototype modulo camera and show improved results over previous methods.

Overview of Limitations.

The SCAMP sensor we have does not include the log circuitry needed for capturing mantissa images, but we still demonstrate the benefits of the proposed reconstruction algorithm on captured modulo images.

2 RELATED WORK

HDR Imaging. The limited dynamic range of conventional camera sensors has been addressed by a number of computational imaging techniques. Exposure bracketing, for example, fuses several low-dynamic-range (LDR) photographs into a single HDR image [6], [7], [8], [9], [10], [11]. Temporally varying exposures can also be processed to obtain HDR videos [12], [13], [14]. Yet, slight movements in the scene will create ghosting artifacts, which are challenging to be removed [40]. The widely used Google HDR+ [11] burst photography of underexposed images approach increases read noise. So, newer HDR+ variants incorporate bracketed exposures. Another class of approaches involves multiple sensors to capture these LDR images simultaneously [15], [16], [17]. Although successful, these systems are expensive, bulky, and often difficult to calibrate.

Several approaches have been developed to estimate an HDR image from a single input image. Reverse tone mapping approaches aim at inverting a tone mapping operator [41], [42], [43], which is an ill-posed inverse problem. Convolutional neural networks can also be directly applied to an LDR image to hallucinate the HDR image [18], [19], [20], [21], [22]. Neither of these approaches, however, has the capability to recover true image details. Bright highlights

can also be optically encoded in an LDR image [31], [32], [33], but this approach relies on the required deconvolution to clean up even an LDR scene perfectly to compete with the quality of conventional sensors, which is challenging. Spatially varying pixel exposures are a promising direction but, similar to color filter arrays, they trade spatial resolution for dynamic range [23], [24], [25], [26], [27], [28], [29], [30].

Among these, our approach to snapshot HDR imaging is most closely related to the modulo camera [36], which combines a modulo-type encoding of the irradiance on the sensor combined with a reconstruction algorithm that solves a 2D unwrapping problem. A conventional modulo operation, however, makes it difficult to distinguish between wrapping boundaries and high-frequency image detail. We introduce a perceptually motivated intensity wrapping technique for this class of computational cameras, which better preserves high-frequency image detail and dynamic range, and we also improve upon existing 2D unwrapping algorithms developed for related tasks.

Unwrapping Algorithms. Phase unwrapping is a problem often encountered in optical interferometry, where the surface profile of some optical element or scene can be indirectly imaged as the wrapped phase of a coherent reference beam. A number of algorithms to unwrap these interferograms has been developed, as surveyed in [44]. When working with wrapped intensities of natural images, instead of optical phase values, the complex interplay of high spatial frequencies and drastically varying light intensity has to be accounted for. Unwrapping techniques for natural images have been analyzed [45] and tailored algorithms developed [46], [47], [48], but these require multiple input images. Most recently, the UnModNet network architecture was introduced to unwrap a single intensity image with state-of-the-art quality [37]. Our network architecture improves upon this method for HDR imaging for modulo cameras but shows best results when used with the proposed mantissa encoding scheme.

Floating Point and Emerging HDR Sensors from the early 2000s allow for capturing high dynamic range with multiple sampling [49], [35] and variations with overlapping integration intervals [50], or choosing optimum integration time [51]. Floating point sensors have great potential, however they require additional bandwidth. Newer image sensors such as Sony’s IMX490 read out multiple images at different gains which then can be merged similarly to bracketed exposures. However, these strategies also require additional bandwidth. In Sony’s sensor, they read out 4 times as many bits as we do, which is significant overhead compared to our approach. Our work reconstructs an HDR image from a single captured LDR image, the same bit depth as a conventional LDR sensor, and instead, utilize the programmability of new sensors for in-pixel irradiance encoding together computational post-processing of that data.

Exotic Sensors for HDR Imaging. Specialized sensor circuits have been developed to support spatially varying pixel and adaptive exposures [35], [52], [53], [54], [55] as well as logarithmic [34] or modulo [36], [56], [57] irradiance encoding. Emerging photon-counting sensors can facilitate HDR imaging, but they are best suited for low-light applications [58] or observe response functions that are similar to logarithmic sensors [59]. These systems can be inflexible,

because they are not programmable at the pixel level. Near-focal-plane sensor-processors [60] include some amount of computing capabilities in the sensor and related systems have become programmable [39], [61], [62], [63], [64], [65]. In this work, we use one of these platforms, SCAMP-5 [39], to prototype modulo encoding and the proposed neural network-based HDR reconstruction algorithm experimentally.

3 PERCEPTUALLY-BASED HDR IMAGING

The MantissaCam framework comprises an electronic in-pixel irradiance encoding scheme and a neural network-based decoding algorithm, which solves the 2D unwrapping problem to reconstruct the irradiance incident on the sensor. We discuss these aspects next.

3.1 In-pixel Irradiance Encoding

The image formation model of the MantissaCam is

$$I_{\text{sensor}}(x, y) = q(\text{mod}(\log_{\tilde{\alpha}}(I(x, y)), I_{\max})) + \eta, \quad (1)$$

where I describes the spatially varying irradiance (i.e., the target HDR image) on the sensor, I_{sensor} is the measured LDR sensor image, $\log_{\tilde{\alpha}}(I(x, y)) = I_{\max} \cdot \log_{\alpha}(I(x, y)/I_{\max})$, and η is zero-mean additive Gaussian noise. The parameter α models a family of logarithmic irradiance response functions. For example, the special case $\alpha = 2$ of our encoding scheme is similar to the mantissa encoding of the IEEE 754 standard for floating point arithmetic. Sensor quantization is modeled by the function $q(\cdot)$. I_{\max} is the maximum allowed irradiance value before the intensity wraps. This could be the well capacity of a pixel or a user-defined value that is slightly lower than that. For more details on implementing the formation model, please refer to the supplement.

3.2 Irradiance Decoding

The proposed decoding scheme is implemented by two neural networks. The first takes the wrapped sensor image as input and predicts the wrap edges, effectively separating them from the texture edges. The second network predicts the winding number (i.e., the number of times intensity has wrapped or the number of times the pixel saturated and reset) of each pixel from these wrap edges.

To predict either modulo or mantissa wrap edges from a sensor image, we directly use the “modulo edge separator” proposed as part of the UnModNet architecture [37]. This edge separator is a residual-type convolutional neural network (CNN) that takes as input a concatenation of the LDR sensor image and a Laplacian-filtered copy of the same. We illustrate our network in Figure 4 and refer the interested reader to [37] for additional details.

Given the wrap edges and the sensor image as input, our second network predicts the winding number for each pixel, $W(x, y)$. For this purpose, features are extracted from both input images using the lightweight CNN-based feature extraction layers from [37]. These are fed into an attention UNet [66] with four downsampling and four upsampling blocks, with each downsampling block using a strided convolutional layer and a residual bottleneck block, and each upsampling block mirroring it but with the addition

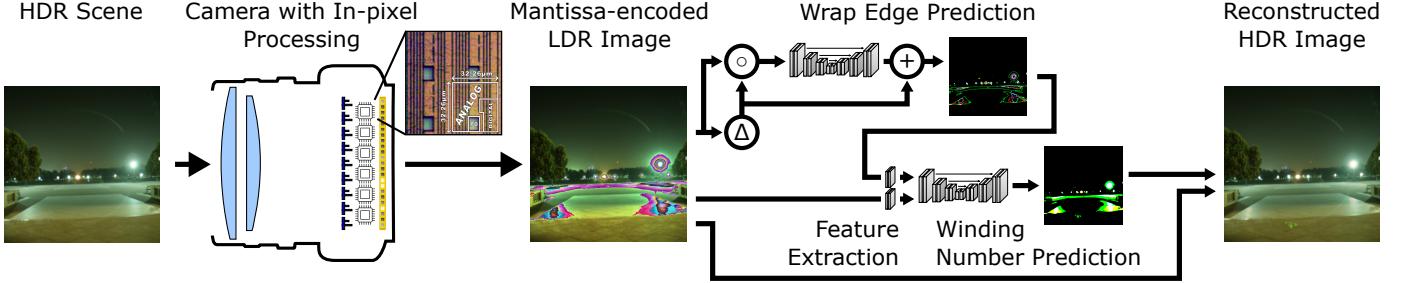


Fig. 4: MantissaCam pipeline. An HDR scene is imaged by a camera with in-pixel processing capabilities, implementing the proposed irradiance encoding scheme (left). The resulting LDR sensor image encodes lower irradiance values similar to a conventional camera, but bright image regions, including the lamp and the reflections on the ground, are wrapped rather than saturated (center). The mantissa-encoded image is first processed by a network that predicts the wrap edges and then by another network that predicts the winding number (center right), the number of times the pixel saturates and resets. The per-pixel winding number, together with the mantissa-encoded image, are used to reconstruct the HDR image (right). The symbols Δ , \circ , and $+$ denote channel-wise Laplacian operators, channel concatenation, and addition, respectively.

of attention gates. This is a standard neural network architecture, but its application to directly predicting the winding number of irradiance-wrapped images is new. Note that this part of our algorithm is substantially different from the iterative, graph-cuts inspired unwrapping procedure proposed in [37]. Their method aims at unwrapping the HDR image layer by layer, which is prone to propagating errors, whereas our approach directly predicts the number of wraps, i.e., the winding number, using a single pass through the UNet. We discuss additional details of this network architecture in the supplement and outline the training procedure of both networks in Section 3.4 and the supplement.

Given the predicted winding number for each pixel as well as the raw sensor I_{sensor} , we formulate the reconstruction of the HDR image \tilde{I} as

$$\tilde{I}(x, y) = I_{\max} \cdot \alpha^{I_{\text{sensor}}/I_{\max} + W(x, y)}. \quad (2)$$

In our implementation, we choose $\alpha = 2$.

3.3 Understanding the Relation between Resolution and Dynamic Range

The theory addressing the ability to perfectly reconstruct a signal with MantissaCam falls within the framework of unlimited sampling recently developed in [45], [67]. Here, rather than formally treating the reconstruction problem, we attempt to highlight the advantages of a mantissa over a modulo encoding and develop an understanding of the tradeoffs between those.

Let us consider the 1D band-limited irradiance function $I(x)$, with maximal frequency f_{\max} . The irradiance is encoded on the sensor by the wrapping function \mathcal{W} of the imaging model:

$$\mathcal{W} : I \in \mathbb{R}_+ \mapsto \mathcal{W}(I) \in [0, I_{\max}]. \quad (3)$$

In particular, we consider the two wrapping functions:

$$\mathcal{W}_{\text{mod}}(I) = I - W(I(x)) \cdot I_{\max}, \quad (4)$$

and

$$\mathcal{W}_{\text{mant}}(I) = \log_{\tilde{\alpha}}(I) - W(\log_{\tilde{\alpha}} I(x)) \cdot I_{\max}, \quad (5)$$

with $W(\cdot) = \lfloor \frac{\cdot}{I_{\max}} \rfloor$ and $\lfloor \cdot \rfloor$ being the floor function.

In order to avoid aliasing on our discrete sensor array, we assume the sampling of I respects the Nyquist sampling criterion $f_s > 2 \cdot f_{\max}$, with the sampling frequency f_s related to the inverse pixel pitch $T_s = \frac{1}{f_s}$ (i.e. the resolution or pixel density, for instance expressed in line pairs per millimeter) of the sensor array.

Recoverability of irradiance from modulo and mantissa encodings To get an intuition about the irradiance fields $\mathcal{W}(I)$ that can be perfectly reconstructed, let us consider the discretized irradiance $I[n] = I(n \cdot T_s)$ as seen by a pixel n .

If a wrap of $\mathcal{W}(I)$ occurs within a pixel, information is lost and it is impossible to reconstruct the incident irradiance field. Therefore, a set of conditions to recover the field is:

$$\begin{cases} |\mathcal{W}(I[n+1]) - \mathcal{W}(I[n])| \leq I_{\max}, \\ |W(I[n+1]) - W(I[n])| \leq 1, \end{cases} \quad (6)$$

where the first condition derives from the Euclidean Division Lemma and makes sure we cannot wrap “within” a pixel, the second condition allows at most one wrap between two pixels.

For the modulo encoding those conditions translate into

$$|I[n+1] - I[n]| \leq I_{\max}, \quad (7)$$

and for the mantissa encoding we have that

$$|\log_{\tilde{\alpha}}(I[n+1]) - \log_{\tilde{\alpha}}(I[n])| \leq I_{\max}. \quad (8)$$

This shows that while the modulo encoding can reconstruct any irradiance with arithmetic growth of I_{\max} , a mantissa encoding can reconstruct a larger class of functions with geometric growth of I_{\max} .

Dynamic range. For both types of encoding, these results imply an interesting tradeoff between the dynamic range of the sensor and its spatial resolution. With two sensors of the same size, using different pixel pitches T_s and T'_s such that $T'_s > T_s$, the sensor with a smaller pixel pitch T_s (i.e., of higher resolution) can reconstruct faster spatial variations of irradiance ($\frac{I_{\max}}{T_s} > \frac{I_{\max}}{T'_s}$ in the modulo case). Therefore, there is a relationship between the maximum dynamic range recoverable for a sensor given its resolution.

For two sensors of fixed size with N pixels, the maximum recoverable irradiance is a ramp starting at pixel $n = 0$ and ending at pixel $n = N - 1$. In this setting, the sensor with modulo encoding can reconstruct a maximum dynamic range of $\text{DR} \approx 10 \log(N \cdot I_{\max}) \text{dB}$ while the one with a mantissa encoding can recover a much wider dynamic range of $\text{DR} \approx 10 \cdot N \log(I_{\max}) \text{dB}$.

Quantization. The ultra-high dynamic range of the mantissa encoding comes at the expense of loss of precision. In practice, no sensor has infinite bit depth but is quantized to 8–12 bits. As shown in the bottom graphs of Figure 2, the same number of levels are distributed on a much wider range as the winding number W increases. This means a MantissaCam cannot resolve irradiance with the same precision ModuloCam can at high irradiance levels—the quantization error is higher for our encoding. Yet, early psychophysics studies [68] noted that perceived light intensity is proportional to the logarithm of the light intensity. This is known as Fechner-Weber law. Likewise when displaying an HDR image, displays have limited dynamic range, so tonemapping is performed which allocates more precision to lower irradiance values. Together, these imply that the coarser quantization of MantissaCam at high irradiance levels might not be perceptually important.

3.4 Dataset and Implementation Details

For a fair comparison, the dataset used to train and evaluate our model was the same dataset created by UnModNet [37]. We randomly split the images into 400 training images and 193 testing images. We used the same process to augment the training dataset, over-exposing and cropping images to yield a total of 5,945 training images.

We train our networks in three stages. First, we train the wrap edge prediction network by itself for 400 epochs, taking simulated sensor images as input, using a binary cross entropy loss with the ground truth wrap edge images obtained via simulation. Second, we train the winding number prediction network by itself for 200 epochs, taking simulated sensor images and ground truth wrap edges as input, using a mean-squared error (MSE) loss on the ground truth winding number. Third, we train both networks jointly for another 200 epochs using the same MSE loss on ground truth winding number. Additional implementation details are found in the supplement.

4 EXPERIMENTS

4.1 Evaluation on Synthetic Data

Figure 5 qualitatively and quantitatively compares modulo and mantissa encoding schemes combined with different reconstruction algorithms. Using a single modulo-wrapped image as input, graph cuts perform poorly [36]. The UnModNet network [37] does reasonably well in some cases, but struggles to reconstruct the large bright parts of the first example scene and the lights on the bridge of the third scene. Their iterative unwrapping procedure sometimes fails in stopping to unwrap, which results in extremely high irradiance values lowering their PSNR and obscuring fine image detail. Our algorithm achieves a better quality than these methods on the same modulo-encoded images,

as evaluated by the peak signal-to-noise (PSNR or P), structural similarity (SSIM or S), multiscale structural similarity (MS-SSIM), learned perceptual image patch similarity (LPIPS), and quality (Q) score of the perceptual HDR Visual Difference Predictor (HDR-VDP-3) [69] metrics. We use standard definitions of PSNR and SSIM (see supplement). The Learned Perceptual Image Patch Similarity (LPIPS) metric evaluates the distance between images. Lower numbers indicate higher perceptually similarity. The HDR-VDP3 Q-Score predicts image quality degradation with respect to the reference image, the max quality being 10. Moreover, when combined with the proposed mantissa irradiance encoding scheme, our framework achieves the best results among all of these methods.

Table 1 also quantitatively compares all of these approaches using several different metrics on the test set of the dataset described in Sec. 3.4. In addition to the above methods, we also include a comparison to a CNN operating directly on a conventional LDR sensor image to hallucinate the HDR scene [19] in Table 1. Not shown are the results from the combination of the UnModNet architecture with the mantissa encoding. The average PSNR was less than 10 dB due to UnModNet’s iterative unwinding. It is prone to propagating errors and with the mantissa encoding, the errors are “exponentially” propagated. As shown in Table 1, the proposed mantissa encoding scheme combined with our reconstruction framework achieves the best results using all metrics, outperforming the state of the art, i.e., UnModNet, by almost 8 dB of PSNR. We also include spatial quality maps in Fig. 6 as well as additional maps in the supplement.

All simulations with synthetic data are run on noise-free images to study the upper bound of all of these algorithms. However, we do include results of simulations with simulated sensor noise in the supplement and also evaluate the best-performing algorithms on noisy captured data in the following.

4.2 Prototyping a Modulo Camera using SCAMP-5

We build a physical prototype using an example of an emerging class of sensors, dubbed focal-plane sensor-processors [60], that embed small processing circuits in-

Encoder Decoder	Modulo			Mantissa	None
	Graph Cuts [36]	UnModNet [37]	Ours	Ours	CNN [19]
PSNR (\uparrow)	21.4	29.5	32.2	37.4	22.7*
Q Score (\uparrow)	5.89	9.08	8.94	9.30	—
SSIM (\uparrow)	0.80	0.79	0.84	0.97	0.72*
MS-SSIM (\uparrow)	0.82	0.91	0.93	0.99	0.76*
LPIPS (\downarrow)	0.29	0.12	0.10	0.03	—

TABLE 1: Quantitative evaluation of modulo and mantissa in-pixel encoding combined with various reconstruction algorithms for simulated data. Our irradiance unwrapping network performs better than existing algorithms on the modulo encoding, as evaluated by several metrics. Combined with the proposed mantissa encoding, our approach achieves state-of-the-art results. We also show the quality of a CNN working with conventional LDR images using the same dataset. Values marked with * are reproduced from [37].

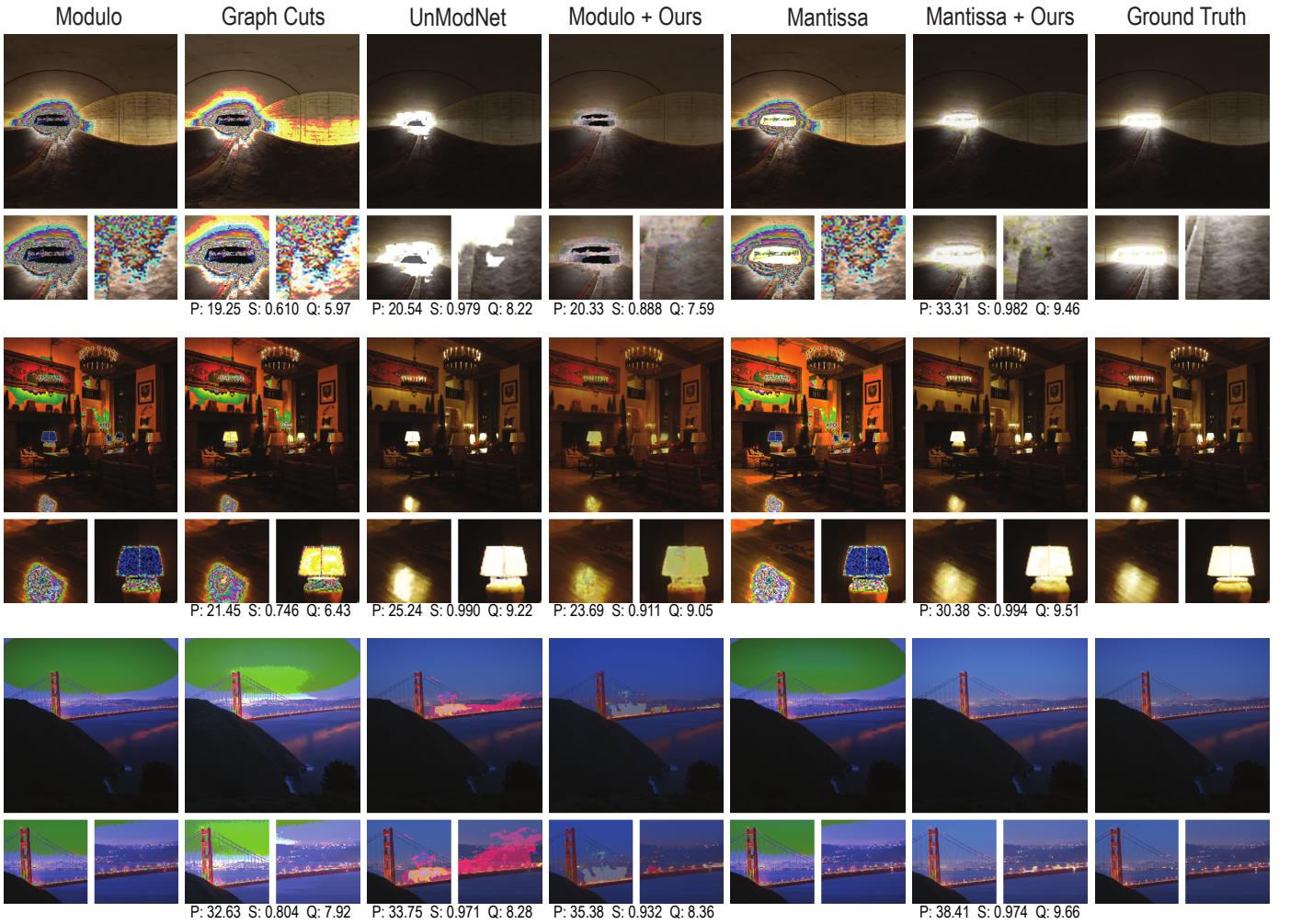


Fig. 5: Evaluation of encoding and decoding schemes in simulation. A conventional modulo encoding wraps the irradiance of a scene into an LDR sensor image (column 1). A graph cuts-based reconstruction algorithm [36] usually performs poorly (column 2) whereas the recently proposed UnModNet architecture [37] often estimates reasonable HDR images (column 3). Yet, the proposed reconstruction framework works best among these methods (column 4). Moreover, the proposed mantissa encoding scheme (column 5) induces fewer irradiance wraps making it easier to reconstruct the HDR image using our framework (column 6). Our approach achieves reconstructions closest to the ground truth (column 7). ‘P’, ‘S’, and ‘Q’ indicate the PSNR, SSIM and Q-score for each reconstruction method.

side each pixel. We use SCAMP-5 [39], whose processing elements (PE) are programmable in a single instruction multiple data (SIMD) fashion, similar to a GPU where the same instruction is performed for all processing elements simultaneously on some local piece of data. Specifically, a PE is equipped with a few analog and digital memories. Instructions can be performed as light is being collected by the pixel’s photo-sensitive element, thus enabling to change the way integration is performed, as required for our implementation. We include a circuit diagram of a single PE in Fig. 7b. For more information on the SCAMP-5, we refer the reader to [39]. In other SCAMP versions, there is log circuitry that would allow us to take mantissa images,

however, our version does not have this capability. We are still able to implement the modulo camera and show the benefits of our reconstruction method over previous state-of-the-art methods.

In our implementation of a modulo camera, an exposure consists of N iterations of length δt (that is programmable and sets the total exposure as $\Delta T = N \cdot \delta t$). At each iteration, the pixel compares whether the accumulated irradiance PIX at the photosensitive element is above a threshold I_{\max} . Whenever the threshold is exceeded, the pixel is reset and a 6-bits pixel-wise digital counter is incremented. This is summarized in the pseudo code of Algorithm 1.

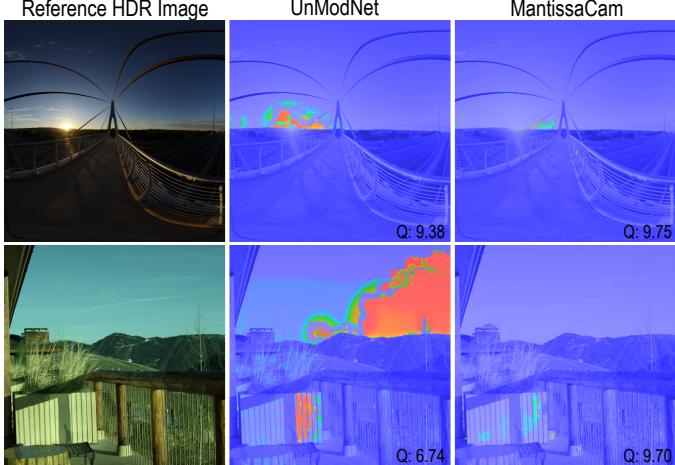


Fig. 6: We show the tonemapped HDR image and the HDR-VDP-3 Q scores and quality maps of UnModNet’s reconstruction and our reconstruction. The maps shows the contrast-normalized per-pixel difference weighted by the probability of detection. Red corresponds to a large perceived difference and blue a low perceived difference.

Algorithm 1: Pseudo code describing our implementation of a Modulo Camera. The for-loop highlighted in grey can be thought of as “parallel-for” executed by all the pixels simultaneously. PIX is a variable representing the photosensitive element whose value is at any instant the integrated value of irradiance from the last pixel reset (one can alternatively think about every inner loop incrementing PIX by the irradiance $I_i \cdot \delta t$ where δt is the time it took to perform the inner loop).

```

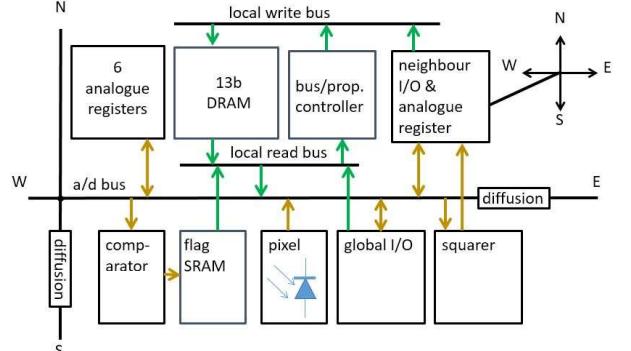
repeat for all frames
    n ← 0
    gaini ← 1
    for all pixels i do
        PIXi ← 0
        while n < N do
            if PIXi ≥ Imax then
                PIXi ← 0
                cnti ← cnti + 1
            n ← n + 1
        readout winding number in cnti
        readout modulo in PIXi
    
```

4.3 Experimental Results

We use SCAMP-5 to prototype a modulo camera and capture HDR scenes outside (see Fig. 7a). This sensor records grayscale images with a resolution of 256×256 pixels. For this experiment, we retrained both UnModNet and our network on modulo images using the same training procedure described in Section 3.4, but on grayscale images captured with SCAMP-5. For this purpose, we collected a dataset of 14,810 modulo and corresponding reference HDR images using the SCAMP-5 prototype. We split this dataset into 13,329 training images and 1,481 test images. No artificial data augmentation was performed. We trained a



(a)



(b)

Fig. 7: (a) Prototype camera capturing an outdoor HDR scene. (b) A circuit diagram for each PE of SCAMP-5, courtesy of Piotr Dudek.

single edge predictor network that we used for UnModNet’s iterative unwrapping approach and also as part of our own pipeline. This network was trained for 150 epochs using the experimentally captured dataset.

Figure 8 shows captured modulo images, the tonemapped reconstructions, and a tonemapped reference HDR image. The captured images include sensor noise, which is especially noticeable around the irradiance wraps. The graph cuts and UnModNet algorithms usually fail to estimate reasonable HDR images, likely due to the noise in the sensor images. For more recognizable results, we limited the number of unwrappings for UnModNet to a maximum of five iterations. Otherwise, the reconstructions end up completely white. The dynamic range of this scene is far greater than that of the sensor, yet our method is able to reconstruct HDR images with high quality.

Table 2 shows the comparison of graph cuts, UnModNet, and our method averaged over the test set captured with the SCAMP-5. We compare PSNR, Q score, SSIM, MS-SSIM, and LPIPS scores. Across all metrics, ours outperforms previous methods by a large margin.

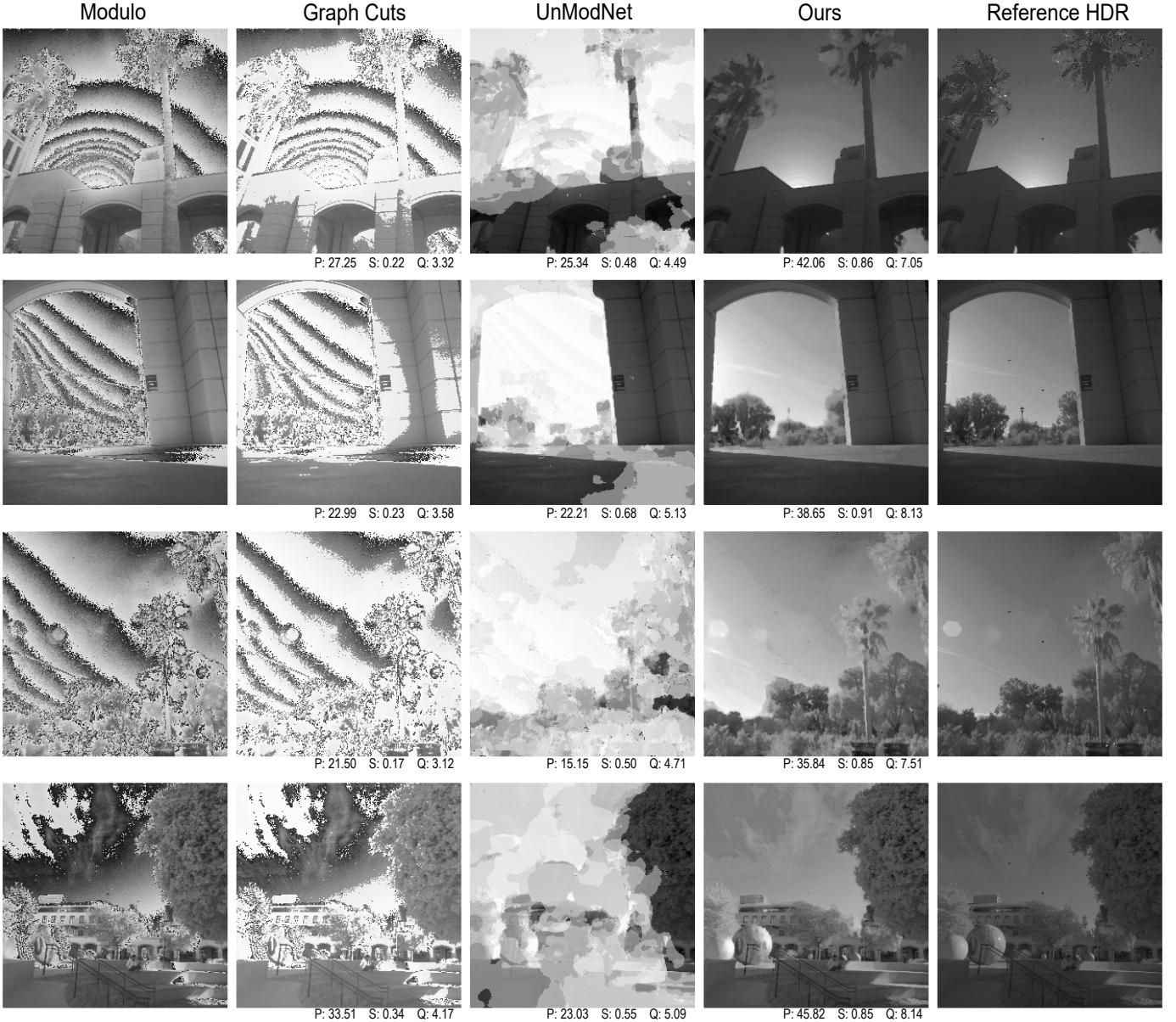


Fig. 8: Experimental results. Using a programmable sensor, SCAMP-5, we capture (noisy) modulo images (left) and process them using graph cuts, UnModNet, and our network applied to the captured modulo data. Tonemapped results using all these reconstruction methods as well as a reference HDR image are shown for several different scenes.

Encoder	Modulo		
	Graph Cuts [36]	UnModNet [37]	Ours
PSNR (\uparrow)	20.3	15.2	33.7
Q Score (\uparrow)	3.81	4.55	7.32
SSIM (\uparrow)	0.27	0.52	0.85
MS-SSIM (\uparrow)	0.23	0.59	0.95
LPIPS (\downarrow)	0.14	0.12	0.09

TABLE 2: Quantitative evaluation of modulo in-pixel encoding combined with various reconstruction algorithms for experimentally captured data. Our algorithm processing the same modulo images as the others achieves significantly better results in all relevant metrics.

With our single-shot HDR image unwrapping method, we can also capture short HDR video clips, which would have been difficult to do with conventional HDR methods like bracketed exposures. In Fig. 9, we show a sequence of modulo-encoded frames that we captured while moving the camera. We also show tonemapped reconstructions using UnModNet and our network. Our method unwraps the modulo video sequence with high temporal consistency and good quality, while lots of flickering and poor image quality are observed for UnModNet. Video clips of these and other example scenes are included in the supplemental material.

5 DISCUSSION

Motivated by the emerging class of programmable sensors, we demonstrate new capabilities they could enable for the

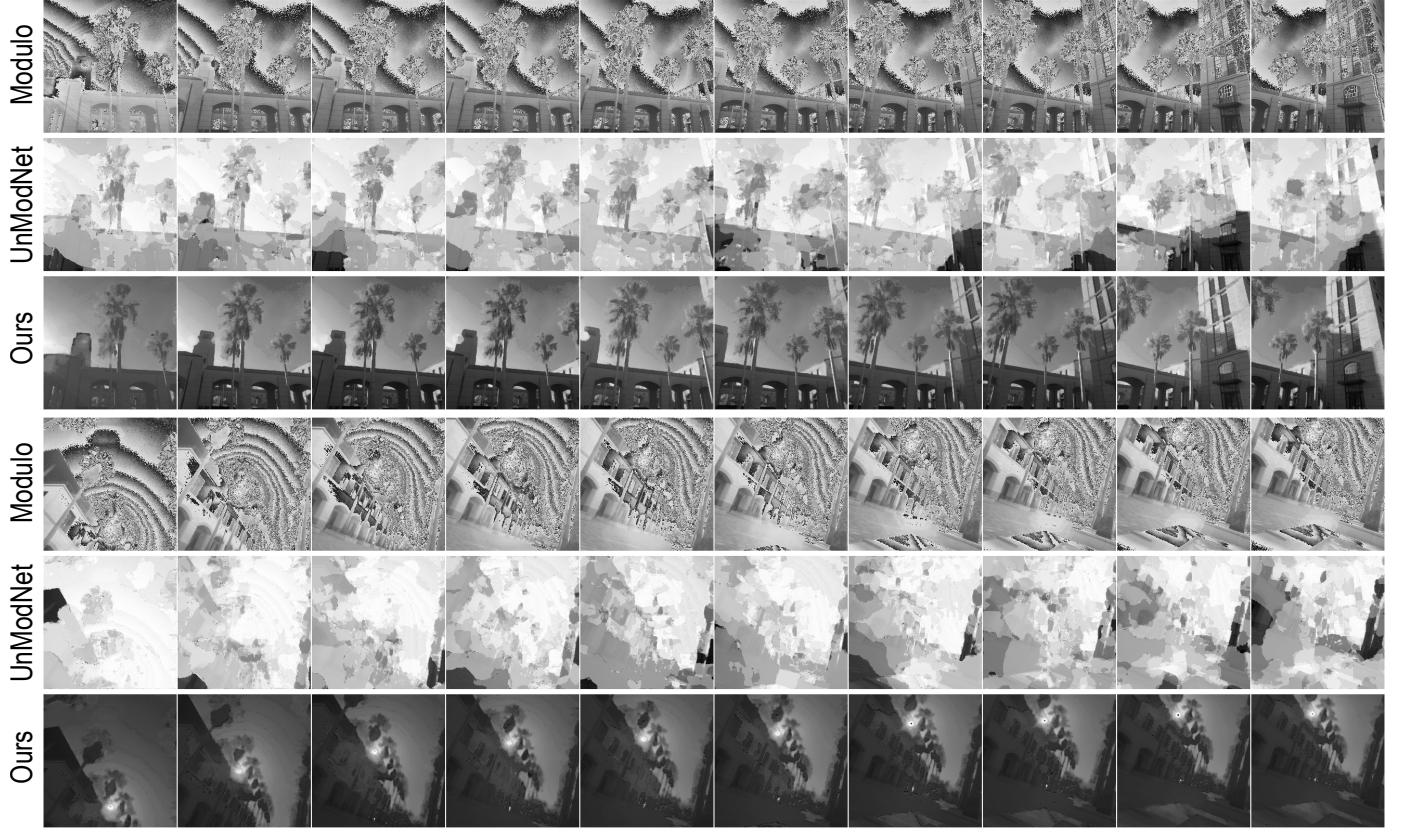


Fig. 9: HDR video experimental results. We show 10 frames of two captured modulo video sequences, UnModNet’s reconstruction, and our reconstruction. Our reconstruction shows temporal consistency and good image quality whereas UnModNet typically fails to estimate reasonable results.

long-standing challenge of snapshot HDR imaging. For this purpose, we develop a reconstruction algorithm for the modulo camera that is more robust and achieves better results than the current state of the art. Moreover, we introduce the mantissa encoding scheme that is inspired by the human visual system and achieves a favorable tradeoff between dynamic range, spatial frequency, and precision when encoding HDR scenes compared to the modulo camera. We evaluate our system in simulation but also show preliminary results captured with a prototype SCAMP-5 programmable sensor, demonstrating the effectiveness of our reconstruction algorithm on the modulo camera. The global shutter speed in our simulations and with the prototype are always set to capture the desired level of detail in the dark regions, relying on the encoder and reconstruction algorithm to recover the brightest parts of the scene.

Limitations and Future Work. Although promising, the proposed system has several limitations. First, our reconstruction pipeline improves results over existing work by a large margin, yet it fails in some cases as shown in Figure 10. Thus, there is room for further improving the robustness of the algorithm. Other future directions include evaluating the usefulness of our unwrapping network architecture in other problems, such as in classical phase unwrapping. Second, our mantissa-based encoding scheme is intuitive and robust, but the base (α) of the logarithm could be optimized in the future. Furthermore, the question of what an optimal encoding scheme for HDR imaging or other applications remains.

Some prior work has studied end-to-end-optimized in-pixel irradiance encoding [30], which could be a fruitful direction for (un)wrapping-based HDR cameras, such as ours. Yet, optimizing periodic objective functions, such as modulo and mantissa-like functions, is not trivial and requires additional research. Third, the class of computational HDR cameras we discuss here seeks to improve the dynamic range of sensors for bright scene parts, but it does not necessarily improve the black level or performance in low-light conditions. It would be valuable to study how in-pixel intelligence offered by programmable sensors could help imaging in low-light scenarios, although this is beyond the scope of our work. Fourth, in our experiments we ignore the effect of the color filter array (CFA), primarily because our prototype is grayscale.

SCAMP-5 is a valuable tool and offers many unique capabilities and a software interface for researchers to do rapid-prototyping. However, it also has limitations including a high read noise level, low pixel fill factor, low resolution, and lack of color filters. Improving these aspects with newer processes, circuit design, 3D fabrication techniques, and improved firmware engineering could make this or related platforms better and more accessible to the computational photography community. But, there may still be a few challenges. As seen with logarithmic cameras today, uniformity per-pixel is difficult. This non-uniformity in the log-circuitry may need to be tackled for the MantissaCam to be widely adapted. But, with advances in fabrication tech-

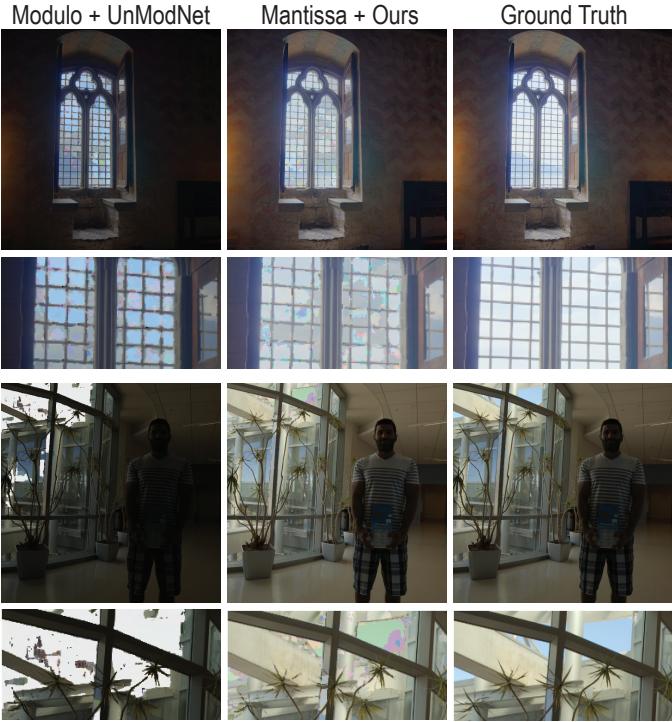


Fig. 10: Limitations. Challenging areas for unwrapping often include regions with high spatial detail and wrapping or dense edges where it may be difficult for the networks to differentiate between wrap and texture edges. While our method is able to better reconstruct some of these areas than a modulo camera with the UnModNet algorithm, some artifacts remain.

niques, along with 3D stacking to allow for more circuitry at the pixel level, eventually MantissaCam could become a promising way to capture huge dynamic ranges.

Conclusion. The emerging class of programmable sensors enables in-pixel intelligence, offering new imaging capabilities for computational photography systems. While our system demonstrates a new co-design of in-pixel irradiance encoding and decoding for snapshot HDR imaging, many other applications in computer vision, photography, and autonomous driving could be enabled by this platform. Our work takes first steps towards the vision of adaptive and domain-optimized computational cameras.

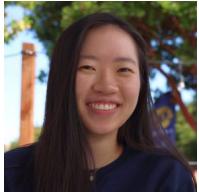
ACKNOWLEDGMENTS

This project was in part supported by NSF Award 1839974, the NSF Graduate Research Fellowship, and a PECASE by the ARL.

REFERENCES

- [1] P. M. Knoll, "Hdr vision for driver assistance," in *High-Dynamic-Range (HDR) Vision*. Springer, 2007, pp. 123–136.
- [2] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Trans. Graph. (SIGGRAPH)*, vol. 23, no. 3, pp. 760–768, 2004.
- [3] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.
- [4] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), 2011.
- [5] J. Ohta, *Smart CMOS image sensors and applications*. CRC press, 2020.
- [6] Mann, Picard, S. Mann, and R. W. Picard, "On being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proceedings of IS&T*, 1995, pp. 442–448.
- [7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.
- [8] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," in *Computer graphics forum (Eurographics)*, vol. 28, no. 1. Wiley Online Library, 2009, pp. 161–171.
- [9] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in *Proc. CVPR*, 2010, pp. 553–560.
- [10] M. Gupta, D. Iso, and S. K. Nayar, "Fibonacci exposure bracketing for high dynamic range imaging," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1473–1480.
- [11] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 192, 2016.
- [12] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph. (SIGGRAPH)*, vol. 22, no. 3, pp. 319–325, 2003.
- [13] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graph. (SIGGRAPH Asia)*, vol. 31, no. 6, pp. 203:1–203:11, 2012.
- [14] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen, "Patch-based High Dynamic Range Video," *ACM Trans. Graph. (SIGGRAPH Asia)*, vol. 32, no. 6, 2013.
- [15] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *International Journal of Computer Vision*, vol. 58, no. 1, pp. 7–17, 2004.
- [16] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar, "Optical splitting trees for high-precision monocular imaging," *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 32–42, 2007.
- [17] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," *ACM Trans. Graph. (SIGGRAPH)*, vol. 30, no. 4, pp. 41:1–41:10, 2011.
- [18] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," in *Computer Graphics Forum*, vol. 37, no. 2. Wiley Online Library, 2018, pp. 37–49.
- [19] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "Hdr image reconstruction from a single exposure using deep cnns," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, p. 178, 2017.
- [20] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 177:1–177:10, 2017.
- [21] S. Lee, G. Hwan An, and S.-J. Kang, "Deep recursive hdri: Inverse tone mapping using generative adversarial networks," in *Proc. ECCV*, 2018, pp. 596–611.
- [22] M. S. Santos, T. I. Ren, and N. K. Kalantari, "Single image hdr reconstruction using a cnn with masked features and perceptual loss," *ACM Trans. Graph. (SIGGRAPH)*, vol. 39, no. 4, 2020.
- [23] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2000, pp. 472–479.
- [24] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *IEEE Int. Conference on Computer Vision (ICCV)*, 2003, p. 1168.
- [25] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable imaging: Towards a flexible camera," *Int. Journal of Computer Vision*, vol. 70, no. 1, pp. 7–22, 2006.
- [26] G. Wetzstein, I. Ihrke, and W. Heidrich, "Sensor saturation in fourier multiplexed imaging," in *Proc. CVPR*. IEEE, 2010, pp. 545–552.

- [27] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive dualiso hdr reconstruction," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 41, 2015.
- [28] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," in *Computer Graphics Forum*, vol. 35, no. 2, 2016, pp. 153–163.
- [29] M. M. Alghamdi, Q. Fu, A. K. Thabet, and W. Heidrich, "Reconfigurable snapshot hdr imaging using coded masks and inception network," in *Vision Modeling and Visualization*, 2019.
- [30] J. Martel, L. Müller, S. Carey, P. Dudek, and G. Wetzstein, "Neural Sensors: Learning Pixel Exposures for HDR Imaging and Video Compressive Sensing with Programmable Sensors," *Proc. IEEE ICCP*, 2020.
- [31] M. Rouf, R. Mantiuk, W. Heidrich, M. Trentacoste, and C. Lau, "Glare encoding of high dynamic range images," in *Proc. CVPR 2011*, 2011, pp. 289–296.
- [32] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1375–1385.
- [33] Q. Sun, E. Tseng, Q. Fu, W. Heidrich, and F. Heide, "Learning rank-1 diffractive optics for single-shot high dynamic range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1386–1396.
- [34] M. Loose, K. Meier, and J. Schemmel, "A self-calibrating single-chip cmos camera with logarithmic response," *IEEE Journal of Solid-state circuits*, vol. 36, no. 4, pp. 586–596, 2001.
- [35] D. Yang, A. E. Gamal, B. Fowler, and H. Tian, "A 640×512 cmos image sensor with ultra wide dynamic range floating-point pixel-level adc," 1999.
- [36] H. Zhao, B. Shi, C. Fernandez-Cull, S. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *Proc. ICCP*, 2015, pp. 1–10.
- [37] C. Zhou, H. Zhao, J. Han, C. Xu, C. Xu, T. Huang, and B. Shi, "Unmodnet: Learning to unwrap a modulo image for high dynamic range imaging," *Advances in Neural Information Processing Systems*, 2020.
- [38] IEEE Computer Society, "Ieee standard for floating-point arithmetic," IEEE STD 754-2019, pp. 1—84, 2019.
- [39] S. J. Carey, A. Lopich, D. R. Barr, B. Wang, and P. Dudek, "A 100,000 fps vision sensor with embedded 535gops/w 256× 256 simd processor array," in *2013 Symposium on VLSI Circuits*. IEEE, 2013, pp. C182–C183.
- [40] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, "The state of the art in hdr deghosting: a survey and evaluation," in *Computer Graphics Forum*, vol. 34, no. 2. Wiley Online Library, 2015, pp. 683–707.
- [41] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, 2006, pp. 349–356.
- [42] L. Meylan, S. Daly, and S. Süstrunk, "The reproduction of specular highlights on high dynamic range displays," *IS&T/SID 14th Color Imaging Conference (CIC)*, 2006.
- [43] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph. (SIGGRAPH)*, vol. 26, no. 3, 2007.
- [44] D. C. Ghiglia and M. D. Pritt, *Two-dimensional phase unwrapping: theory, algorithms, and software*. Wiley, 1998.
- [45] A. Bhandari, F. Krahmer, and R. Raskar, "On unlimited sampling and reconstruction," *IEEE Transactions on Signal Processing*, 2020.
- [46] F. Lang, T. Plötz, and S. Roth, "Robust multi-image hdr reconstruction for the modulo camera," in *German Conference on Pattern Recognition*. Springer, 2017, pp. 78–89.
- [47] V. Shah and C. Hegde, "Signal reconstruction from modulo observations," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2019, pp. 1–5.
- [48] V. Shah, M. Soltani, and C. Hegde, "Reconstruction from periodic nonlinearities, with applications to hdr imaging," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 863–867.
- [49] P. Acosta-Serafini, I. Masaki, and C. Sodini, "Predictive multiple sampling algorithm with overlapping integration intervals for linear wide dynamic range integrating image sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 1, pp. 33–41, 2004.
- [50] P. M. Acosta-Serafini, I. Masaki, and C. G. Sodini, "A 1/3" vga linear wide dynamic range cmos image sensor implementing a predictive multiple sampling algorithm with overlapping integration intervals," in *Proceedings of the IEEE 2003 Custom Integrated Circuits Conference*, 2003., 2003, pp. 485–488.
- [51] J. Rhee, D. Park, and Y. Joo, "Analysis and design of a robust floating point cmos image sensor," *IEEE Sensors Journal*, vol. 9, no. 5, pp. 578–585, 2009.
- [52] M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori, and M. Furuta, "A wide dynamic range cmos image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic a/d converters," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2787–2795, 2005.
- [53] P. Dudek, "Adaptive sensing and image processing with a general-purpose pixel-parallel sensor/processor array integrated circuit," in *Int. Workshop on Computer Architecture for Machine Perception and Sensing*. IEEE, 2006, pp. 1–6.
- [54] R. Wagner, Á. Zarándy, and T. Roska, "High dynamic range perception with spatially variant exposure," in *IEEE Int. Workshop*, 2004.
- [55] J. N. Martel, L. K. Müller, S. J. Carey, and P. Dudek, "Parallel hdr tone mapping and auto-focus on a cellular processor array vision chip," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2016, pp. 1430–1433.
- [56] B. Tyrrell, K. Anderson, J. Baker, R. Berger, M. Brown, C. Colonero, J. Costa, B. Holford, M. Kelly, E. Ringdahl, K. Schultz, and J. Wey, "Time delay integration and in-pixel spatiotemporal filtering using a nanoscale digital cmos focal plane readout," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2516–2523, 2009.
- [57] M. G. Brown, J. Baker, C. Colonero, J. Costa, T. Gardner, M. Kelly, K. Schultz, B. Tyrrell, and J. Wey, "Digital-pixel focal plane array development," in *Quantum Sensing and Nanophotonic Devices VII*, M. Razeghi, R. Sudharsanan, and G. J. Brown, Eds., vol. 7608, International Society for Optics and Photonics. SPIE, 2010, pp. 765 – 774. [Online]. Available: <https://doi.org/10.1117/12.838314>
- [58] E. R. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *Sensors*, vol. 16, no. 8, 2016.
- [59] A. Ingle, A. Velten, and M. Gupta, "High flux passive imaging with single-photon sensors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6760–6769.
- [60] Á. Zarándy, *Focal-plane sensor-processor chips*. Springer Science & Business Media, 2011.
- [61] W. Miao, Q. Lin, W. Zhang, and N.-J. Wu, "A programmable simd vision chip for real-time vision applications," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 6, pp. 1470–1479, 2008.
- [62] A. Lopich and P. Dudek, "An 80× 80 general-purpose digital vision chip in 0.18 μm cmos technology," in *IEEE Int. Symposium on Circuits and Systems*, 2010, pp. 4257–4260.
- [63] A. Rodríguez-Vázquez, R. Domínguez-Castro, F. Jiménez-Garrido, S. Morillas, A. García, C. Utrera, M. D. Pardo, J. Listan, and R. Romay, "A cmos vision system on-chip with multi-core, cellular sensory-processing front-end," in *Cellular nanoscale sensory wave computing*. Springer, 2010, pp. 129–146.
- [64] W. Zhang, Q. Fu, and N.-J. Wu, "A programmable vision chip based on multiple levels of parallel processors," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 9, pp. 2132–2147, 2011.
- [65] J. Fernández-Berni, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "Flip-q: A qcif resolution focal-plane array for low-power image processing," in *Low-Power Smart Imagers for Vision-Enabled Sensor Networks*. Springer, 2012, pp. 67–109.
- [66] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," 2018.
- [67] A. Bhandari, F. Krahmer, and R. Raskar, "On unlimited sampling," in *2017 International Conference on Sampling Theory and Applications (SampTA)*. IEEE, 2017, pp. 31–35.
- [68] G. T. Fechner, "Elements of psychophysics, 1860." 1948.
- [69] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph. (SIGGRAPH)*, vol. 30, no. 4, pp. 40:1–40:14, 2011.



Haley M. So is a PhD Candidate in the Stanford Computational Imaging Lab. She is interested in creating new imaging algorithms enabled by in-pixel processing. She received her B.S. in Electrical Engineering from Columbia University, NY in 2016.



Gordon Wetzstein (Senior Member, IEEE) received the graduation (with Hons.) degree from the Bauhaus-Universität Weimar, Weimar, Germany and the Ph.D. degree in computer science from the University of British Columbia, BC, Canada, in 2011. He is currently an Associate Professor of Electrical Engineering and, by courtesy, of Computer Science, with Stanford University, Stanford, CA, USA. He is the Leader of Stanford Computational Imaging Lab and a Faculty Co-Director of the Stanford Center for Image Systems Engineering. At the intersection of computer graphics and vision, computational optics, and applied vision science, his research has a wide range of applications in next-generation imaging, display, wearable computing, and microscopy systems. He is the recipient of an NSF CAREER Award, an Alfred P. Sloan Fellowship, an ACM SIGGRAPH Significant New Researcher Award, a Presidential Early Career Award for Scientists and Engineers (PECASE), an SPIE Early Career Achievement Award, a Terman Fellowship, an Okawa Research Grant, the Electronic Imaging Scientist of the Year 2017 Award, an Alain Fournier Ph.D. Dissertation Award, Laval Virtual Award, and the Best Paper and Demo Awards at ICCP 2011, 2014, and 2016 and at ICIP 2016.



Julien N.P. Martel obtained his Ph.D in 2019 at ETH Zurich in the Institute of Neuroinformatics. He was a postdoctoral research scholar at Stanford University from 2019 to 2022 in the Computational Imaging Lab. His broad interests are in vision and graphics but specifically in computational imaging and rendering systems and algorithms. Many of his projects have revolved around exploiting sensors that embed little processors in their pixels, allowing them to change the way light is captured.



Piotr Dudek is a Professor of Circuits and Systems in the School of Electrical and Electronic Engineering, The University of Manchester, leading the Microelectronics Design Lab. He received his mgr inz degree from the Technical University of Gdańsk, Poland, and the MSc and PhD degrees from the University of Manchester Institute of Science and Technology (UMIST) in 1996 and 2000 respectively. He worked as a Research Associate, and since 2002 as a Lecturer, at UMIST/The University of Manchester. He is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) and a member of IEEE Circuits and Systems Society Technical Committees on Neural Systems, and on Cellular Neural Networks and Array Processing. He was the Chair of the IEEE CAS Technical Committee on Sensory Systems (2015-17). He has been a member of Scientific Committees, Track Chair or Session Chair at many international conferences. He is an Associate Editor for IEEE Transactions on Circuits and Systems II and Review Editor for Frontiers in Neuroscience and an active reviewer for many journals and funding agencies. His work received several Best Paper and Best Demo awards at international conferences (ISCAS, IJCNN, CNNA, ICDSC), and he has given numerous keynote and invited talks, and examined PhD theses worldwide.