

Unbounded High Dynamic Range Photography using a Modulo Camera

Hang Zhao¹ Boxin Shi^{1,3} Christy Fernandez-Cull² Sai-Kit Yeung³ Ramesh Raskar¹

¹MIT Media Lab ²MIT Lincoln Lab ³Singapore University of Technology and Design
{hangzhao, shiboxin, raskar}@media.mit.edu, christy.cull@ll.mit.edu, saikit@sutd.edu.sg

Abstract

This paper presents a novel framework to extend the dynamic range of images called Unbounded High Dynamic Range (UHDR) photography with a modulo camera. A modulo camera could theoretically take unbounded radiance levels by keeping only the least significant bits. We show that with limited bit depth, very high radiance levels can be recovered from a single modulus image with our newly proposed unwrapping algorithm for natural images. We can also obtain an HDR image with details equally well preserved for all radiance levels by merging the least number of modulus images. Synthetic experiment and experiment with a real modulo camera show the effectiveness of the proposed approach.

1. Introduction

Real world scene radiance levels cover a huge dynamic range that is impossible to completely capture given limited quantization bits and well capacity on a traditional imaging sensor. Typically, all brightness and structural information are lost within the saturated regions. To increase the dynamic range that could be captured, high dynamic range (HDR) photography aims to address this problem by increasing camera bit depth of cameras through hardware modifications or using computational methods by merging multiple captures with varying exposure levels. With HDR, the dynamic range increases, but it is still bounded.

To design an imaging sensor that has infinite dynamic range is physically infeasible, since the sensor keeps on collecting photons but the storage and precision of analog-to-digital converter (ADC) cannot increase infinitely. A smart tradeoff in taking ultra high dynamic range data with a limited bit depth is to wrap the data in a periodical manner. This creates a sensor that never saturates: whenever the pixel value gets to its maximum capacity during photon collection, the saturated pixel counter is reset to zero at once, and following photons will cause another round of pixel value increase. This rollover in intensity is a close anal-

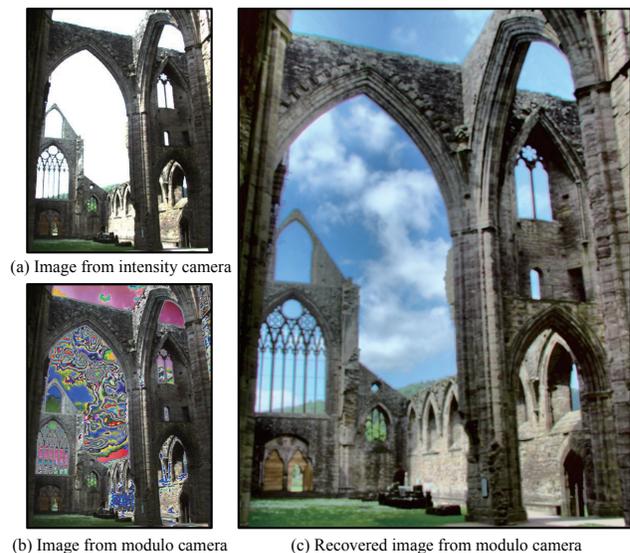


Figure 1. A modulo camera could well recover over-exposed regions: (a) an image taken by an 8-bit intensity camera, which has large areas saturated due to the bright sky; (b) an image taken by an 8-bit modulo camera, with the same exposure level as (a); (c) a recovered image from modulus (after tone-mapping), which retains the information in the saturated part. Radiance map is courtesy of Greg Ward.

ogy to phase wrapping in optics, so we borrow the words “(un)wrap” from optics to describe the similar process in the intensity domain. Based on this principle, a modulo camera could be designed to record modulus images that theoretically have an Unbounded High Dynamic Range (UHDR).

In this paper, we explore the use of a modulo camera in both single-shot and multi-shot scenarios to address the HDR problem from a brand new point of view. To extend the dynamic range with a single shot, we propose a graph-cuts-based unwrapping algorithm to recover information in the wrapped region (saturated in the intensity image). The unwrapped result from an 8-bit modulus image has much higher dynamic range than a single image captured

by a conventional 8-bit intensity camera ¹. Fig. 1(a) shows an image taken by an intensity camera with a significant over-exposed region due to bright sky background; in comparison, Fig. 1(b) is a modulus image with wrapped data; Fig. 1(c) shows the recovered image (tone-mapped with the method in [20]) taken by a modulo camera, which reveals high radiance details on the saturated sky. Apart from higher dynamic range by unwrapping, we further demonstrate that a modulo camera better preserves scene radiance details than an intensity camera under multiple image capturing scenarios with high fidelity. This results in getting accurate radiance images with the least number of captures. To our knowledge, there is no previous technology that could measure high radiance values with sufficient precision on the least significant bits as a modulo camera.

The key contributions of this paper are summarized as follows:

- A novel framework of unbounded high dynamic range (UHDR) photography using a modulo camera;
- A new formulation for single modulus image recovery using graph cuts and natural images properties for the energy minimization;
- A new view on merging multiple modulus images to preserve highly-detailed radiance characteristics in UHDR photography;
- The first trial of experiments with a real modulo sensor to validate our analyses.

In addition, we make simulations, compare our techniques with other approaches, and provide suggestions for future work.

2. Related Works

HDR photography. Increasing the dynamic range of an intensity camera could be realized by using either computational methods or novel pixel architecture. For computational methods, usually multiple images captured with different effective exposure levels are needed [6, 23]. Methods using multiple-image captures benefit from ease of implementation, but non-static scene capturing remains a challenging task. Therefore, some efforts are put in image registration and ghosting removal for multi-exposure HDR photography [9]. Subsequently, Hirakawa and Simon [17] showed that dynamic range could be extended within a single-shot based on color filtering of conventional cameras. Apart from these, many researches optimize exposure, noise and details recovery for HDR photography [13, 16, 14].

¹We will use “intensity camera” throughout the paper to refer to a conventional camera with linear response in comparison to our modulo camera.

By modifying the image capturing process within the pixel architecture, single-shot HDR becomes feasible to implement. While a logarithmic intensity camera is a way to avoid image saturation, Tumblin *et al.* [24] proposed a gradient of logarithmic camera to cover most of the contrasts in natural scenes. Nayar *et al.* [22, 21] also provided some solutions including light modulation with spatially-varying exposure masking and adaptive pixel attenuation. For commercial products, a straightforward way is using a very high precision ADC ². Fujifilm has designed SuperCCD ³ that has paired pixels with different effective pixel areas, resulting in different effective exposures. Sony ⁴, on the other hand, proposed a per-pixel exposure camera by setting different exposure times for two groups of pixels. Most of these methods achieve higher dynamic range at the cost of spatial resolution.

Phase unwrapping. Phase unwrapping is a well-studied problem in imaging domains like optical metrology [7], magnetic resonance imaging (MRI) [5] and synthetic aperture radar (SAR) [10]. Famous solutions to phase unwrapping problems include solving Poisson’s equation with DFT/DCT, path-following method, iterative re-weighted L-p norm method, *etc.* A comprehensive study of these methods can be found in [11]. More recently, time-of-flight (ToF) cameras have also employed unwrapping techniques for depth estimation [18, 15].

With the similarity of modulo and wrapping operators, our problem is analogous to performing phase unwrapping for natural images, since we are interested in recovering an HDR image from its modulus counterpart. However, our task is much more challenging in two aspects: 1) In interferometric SAR and ToF, a complex-valued wave representation can be obtained, therefore “magnitude” information is available to guide a good “phase” reconstruction; 2) Natural images are more complicated and rich in content with high spatial frequencies (edges, peaks) and dynamic range as compared to interferometric optics, MRI and SAR.

3. Background

Wrapping and unwrapping. The proposed modulo camera resets the pixel values whenever the counters reach maximum. This behavior is similar to phase wrapping of electromagnetic waves. In Fig. 2, we use an over-simplified surface to illustrate the process of wrapping and unwrapping a 2-D intensity image. In the forward image formation, the 2-D data we get from a modulo camera is the same as the least

²http://www.viewplus.co.jp/product_category.html#!cat=xviii

³<http://www.fujifilmusa.com/shared/bin/4thGenSUPERCCDBrochure.pdf>

⁴<http://images3.freshpatents.com/pdf/US20140321766A1.pdf>

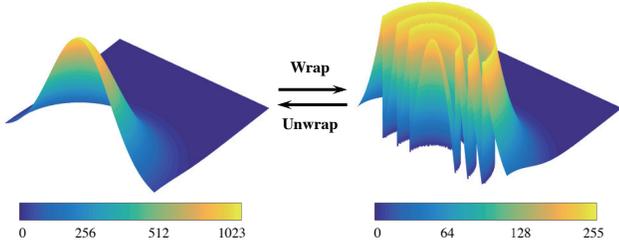


Figure 2. An over-simplified wrapping and unwrapping example: a surface with contrast ratio 1023 : 1 is wrapped to an 8-bit range of a contrast ratio 255 : 1; note the modulo fringes in the wrapped image.

significant N bits of an intensity camera with an infinite bit depth. However, restoration from the modulus of a single image (unwrapping) is an ill-posed problem, where the unknown is the number of rollovers $k(x, y)$ at each pixel. We have:

$$I_m(x, y) = \text{mod}(I(x, y), 2^N), \quad (1)$$

or

$$I(x, y) = I_m(x, y) + k(x, y) \cdot 2^N, \quad (2)$$

where I is the ground truth image captured by an intensity camera with an infinite bit depth, and I_m is the value taken by a modulo camera with N bits. In the case of Fig. 2, a 10-bit image with contrast ratio 1023 : 1 is wrapped to an 8-bit range (contrast ratio 255 : 1). Modulo fringes appear in the wrapped image, where each fringe represents a rollover around that region. The goal of UHDR is to recover the original surface from either single or multiple modulus images.

Modulo sensor. In conventional image acquisition pipeline, after the analog signal is obtained, the analog-to-digital converter (ADC) in an intensity camera quantizes the recorded signal to N bits. N is usually 8 for compact digital cameras and 12 or 14 for high-end DSLR cameras. Whenever the analog signal level is high enough to fill the well capacity, it maximizes digital output and leads to saturation. Recent designs of the Digital-pixel Focal Plane Array (DFPA) [25] feature the ability to do on-the-fly digital signal processing, as shown by the block diagram in Fig. 3. This novel pixel architecture can be recast as a modulo sensor. With the detector array mated to a silicon CMOS readout integrated circuit (ROIC), the sensor digitizes the signal using a small capacitor (minimum quantization level) that integrates photocurrent to a pre-defined threshold charge level. Once the threshold charge level is reached, the capacitor is automatically discharged and starts to accumulate charge again. A pulse generator

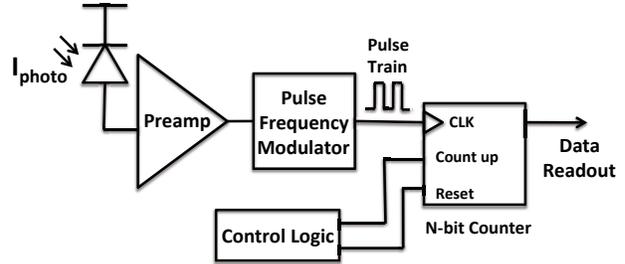


Figure 3. Pixel architecture of a modulo sensor: pulse frequency modulator determines the quantization level of the sensor. Whenever voltage passes the threshold, the counter will receive a pulse to count up. Digital signal processing control logic resets the counter to zero whenever the counter is full.

that is triggered on every discharge drives a digital counter contained within each pixel. The counter counts up by one in response to each pulse. When the N -bit counter reaches a maximum value, the digital control logic resets it to zero, thus forming an N -bit modulo sensor.

The real modulo camera sensor does not include a non-linear radiometric response function, therefore radiometric calibration is not included in our image formation model and all analysis in this paper.

4. Single-shot UHDR

In this section, we show how to recover from a single modulus image to get an UHDR image through natural image unwrapping by formulating it as a graph-cut problem with specially designed cost functions.

4.1. Formulation of natural image unwrapping

An intensity-wrapped natural image contains many modulo fringes in the “saturated”⁵ region. For single-image UHDR problem, extra dynamic range is obtained from these fringes. Our key observation is that these modulo fringes are highly local: they form steps between two neighboring pixels. This can be easily observed from the simple surface example in Fig. 2. To provide a more intuitive example using a natural image, we render a synthesized scene with ground truth maximum contrast around 7800 : 1 (13 stops), from which its 8-bit modulus image is calculated and shown in Fig. 4(a). In this scene, the light source on the ceiling is so bright that the intensities of a large amount of pixels have been wrapped for many times. Based on this property, we formulate the restoration of I through minimizing the energy of the first-order Markov random field (MRF) with pairwise interactions as

$$C(k|I_m) = \sum_{(i,j) \in \mathbb{G}} V(|\hat{I}_i - \hat{I}_j|), \quad (3)$$

⁵“Saturated” here means the saturated pixels if an intensity sensor with limited bits were used.

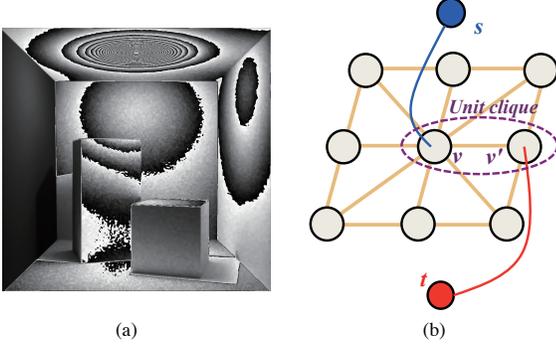


Figure 4. (a) Synthetic image of a scene taken by a modulo camera. Modulo fringes are brought by the bright light source on the ceiling. (b) Graph representation: each pixel forms cliques with its surrounding 8 pixels. Source s , sink t and an unit clique (v, v') form an elementary graph.

where $\hat{I} = I_m + k \cdot 2^N$ is the restored image ((x, y) omitted for simplicity), $V(\cdot)$ is the clique potential, \mathbb{G} represents the set of all pairwise cliques in the MRF and (i, j) are the two pixels in each clique. The energy function is designed to minimize neighboring pixel differences for two reasons: 1) to penalize steps around fringes brought by modulo operation, and 2) the well-studied gradient domain image statistics could be applied to solve our problem. Note that our problem is different from most MRF formulations in computer vision due to the lack of a data term in the energy function, which makes the problem difficult to solve. Our goal is to find the optimal two-dimensional map k that minimizes the energy cost function $C(k|I_m)$.

4.2. Energy minimization via graph cuts

The energy minimization given above is an integer optimization problem, therefore it can be decomposed into a series of binary minimizations that could be solved via graph cuts [3]. We iteratively seek 2-D binary sets $\delta \in \{0, 1\}$ that make $C(k + \delta|I_m) < C(k|I_m)$, and update $k = k + \delta$ until energy stops decreasing.

Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nonnegative edge weights, source s and sink t . An s - t cut $\mathcal{C} = \mathcal{S}, \mathcal{T}$ is a partition of the vertices \mathcal{V} into two disjoint sets \mathcal{S} and \mathcal{T} such that $s \in \mathcal{S}$ and $t \in \mathcal{T}$. According to the Class \mathcal{F}^2 Theorem proved by Kolmogorov and Zabih [19], the necessary and sufficient condition for our energy function to be graph-representable is

$$V(x + 2^N) + V(x - 2^N) \geq 2V(x). \quad (4)$$

The structure of an elementary graph with source s , sink t , and a clique pair is shown in Fig. 4(b). In each elementary graph, we assign weight $V(x + 2^N) + V(x - 2^N) - 2V(x)$ to the directed edge (v, v') , and $|V(x + 2^N) - V(x)|$ to

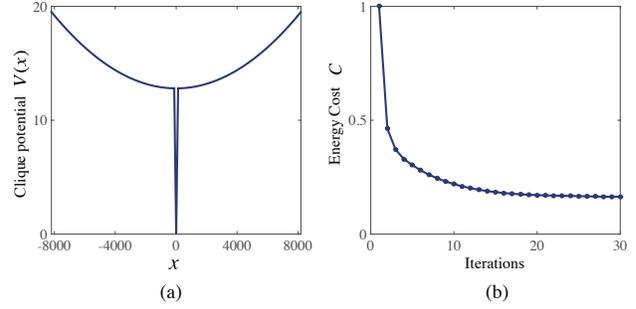


Figure 5. (a) The piecewise energy function to be minimized for natural HDR images. (b) Energy drop during each iteration for the case in Fig. 4(a).

(s, v) and (v', t) . Joining all elementary graphs together forms a global graph that can be segmented by existing max-flow/min-cut algorithms [2]. During each binary segmentation procedure, for all the pixels labeled with s , δ is set to 1, and for other pixels labeled with t , δ is set to 0.

The design of a good potential function $V(x)$ is crucial in the optimization. Quadratic functions are typically used in energy minimizations, but it is well-known that they tend to smooth the result while doing image restoration. Different from unwrapping in optics, where signals are mostly smooth, natural images usually contain high frequency features, such as edges. Based on the investigation of [12] on the statistical properties of HDR images under gradient domain representation, we design a piecewise potential function comprised of a linear and a quadratic part to estimate their gradient probability distribution:

$$V(x) = \begin{cases} a_1 x & |x| \leq x_0 \\ a_2 x^2 + b & |x| > x_0 \end{cases}, \quad (5)$$

where we empirically choose $a_1 = 0.1$, $a_2 = 10^{-5}$, and $x_0 = 2^{N-1}$ ($N = 8$), as plotted in Fig. 5(a). These parameters determine prior smoothness of the latent image, and they are fixed for all our experiments in this paper.

From Eq. (4), we would easily find that convex functions satisfy the graph-representable requirement, which means an energy drop is guaranteed at each step until convergence. While our designed potential function is able to preserve sharp changes in natural images, it violates the graph-representable requirement. Hence, we put additional constraints to ensure good cuts for each iteration and avoid being trapped in local minima. We adopt a similar approach as proved in [1]. Whenever Eq. (4) is not reached, we set the weight of the edge (v, v') to zero, meaning this elementary graph does not contribute in energy increase in the s - t cut; whenever energy is minimized, step size δ is extended to look for lower possible energy, namely we shift to $\delta \in \{0, \sigma\}$ for further iterations, where we empirically choose $\sigma = 2$.

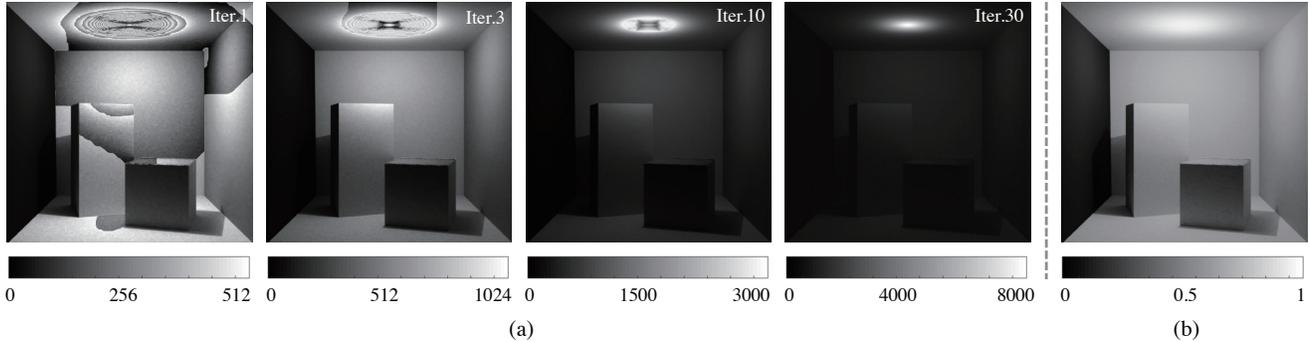


Figure 6. (a) Intermediate unwrapping solution at each iteration: from left to right $\{1, 3, 10, 30\}$. The images are linearly scaled in intensity, where gray bars below tell the corresponding intensity values for each image. (b) Tone-mapped image of the final result.

We show the intermediate and converged results of the unwrapping example in Fig. 6. The four linear images in Fig. 6(a) illustrate the progressive recovering capability of the graph cuts algorithm, and the disappearance of fringes due to the removal of rollovers could be obviously observed. The final unwrapping result is free of errors as compared to the ground truth in this test. Fig. 6(b) shows the tone-mapped image of the final result. The normalized energy drop at each iteration is plotted in Fig. 5(b) which clearly shows good convergence of the algorithm.

5. Multi-shot UHDR

Unwrapping a single modulus image can extend the dynamic range greatly, but this inverse problem is ill-posed. When multiple modulus images under different exposures are available, the problem becomes over-determined. We analyze in this section the advantages associated with merging multiple modulo images, namely both dynamic range recovery and details preservation.

5.1. Accurate scene radiance estimation

To record a scene with contrast $65,000 : 1$ (16 stops), theoretically capturing two images with an exposure ratio of $256 : 1$ using an 8-bit intensity camera is enough. Unfortunately, merely combining these two images will lose too many details since it is always challenging to capture both high dynamic range and detailed scene radiance using an intensity camera. Intuitively, the under-exposed image is “black” in most areas and the over-exposed image is almost “white” everywhere. That’s why people usually take more than two captures (*e.g.*, five images with exposure ratio of $4\times$, nine images with exposure ratio of $2\times$) for high quality HDR imaging. With a modulo camera, we can achieve the win-win in both very high dynamic range and very detailed scene radiance preservation using as few as two images.

As we focus on sensors with a linear response, observed intensity value is proportional to scene radiance. Here, we

borrow the expression from Grossberg and Nayar [13] to use a quantized response function $I = f(E)$ to represent the the quantized image intensity value I corresponding to the scene radiance E . The response function is a piecewise constant function that characterizes the quantization process of a camera. We start with a simple case of two shots by a 2-bit camera for an easy illustration. As shown in Fig. 7(a), due to saturation, an intensity camera does not provide any radiance discrimination for large radiance during long exposure image capture; for the short exposure image, on the other hand, large quantization steps “flatten” many radiance details.

A modulo camera makes the story completely different. As shown in Fig. 7(b), its response function of the long exposure capture periodically repeats with a much smaller quantization step. With this small uniform radiance measuring intervals, the complete dynamic range of the scene is sampled equivalently to using a 4-bit linear camera, providing much more measured scene radiance levels than an intensity camera. Although there exist wrapped values when only one long-exposure image is used, reconstruction is quite easy with the help of a short-exposure image, as we will introduce next.

5.2. Unwrapping multiple modulus images

Conventional linear image merging for HDR is done by weighting and averaging. Weights are assigned to all pixels in photos of different exposure levels. In the simplest case, a binary mask is generated for each image to exclude over-exposed and under-exposed pixels. An average is then taken for each pixel across different images to get a linear HDR map. In the masking process, about half of all pixels have zero weights.

The modulus images avoid the annoying over-exposure problem and retain much more useful information. Since the images with longer exposures keep more radiance details but come with wrapped values, while shorter exposure

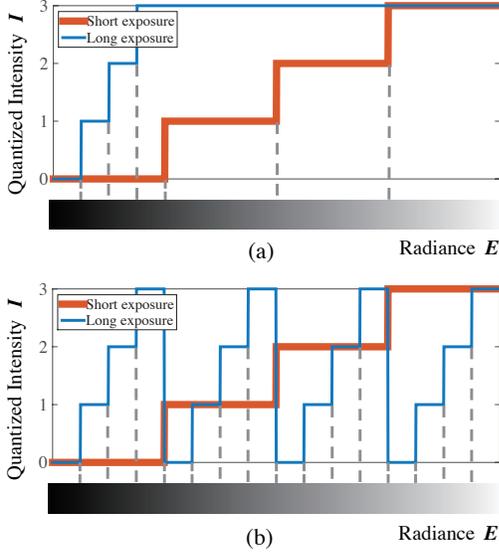


Figure 7. Illustration of response function $I = f(E)$ for short and long exposures. All possible (continuous) radiance levels are shown on the bar below: (a) an intensity camera; (b) a modulo camera.

images have less details without wrapped values, so we can simply use images with shorter exposures to help unwrap the longer ones.

Assuming we have m modulo images as input, we first sort them according to their exposure times in an ascending order, and the radiance map is initialized to be $E^{(0)} = 0$. We start from the shortest exposure image that does not have any rollovers and end up with the longest exposure. When processing the i -th image, the number of rollovers in the new image $k^{(i)}$ is calculated from the exposure time $\tau^{(i)}$ and the previous radiance map $E^{(i-1)}$, where the superscript indicates the image index:

$$k^{(i)} = \lfloor \frac{\tau^{(i)} \cdot E^{(i-1)}}{2^N} \rfloor. \quad (6)$$

And then a finer radiance map is updated by combining the rollovers $k^{(i)}$ and the modulus $I_m^{(i)}$:

$$E^{(i)} = \frac{k^{(i)} \cdot 2^N + I_m^{(i)}}{\tau^{(i)}}. \quad (7)$$

By this analogy, a final radiance map $E^{(m)}$ is achieved with the finest radiance levels. To guarantee that the radiance map can be obtained without wrapped values, the exposure ratio between neighboring images should be $\frac{\tau^{(i+1)}}{\tau^{(i)}} \leq 2^N$, $i = \{1, 2, 3, \dots, m-1\}$, which is usually satisfied. If rollovers exist in the shortest exposure image, the single image unwrapping method discussed in Sec. 4 can be applied first, but here we do not consider this case.

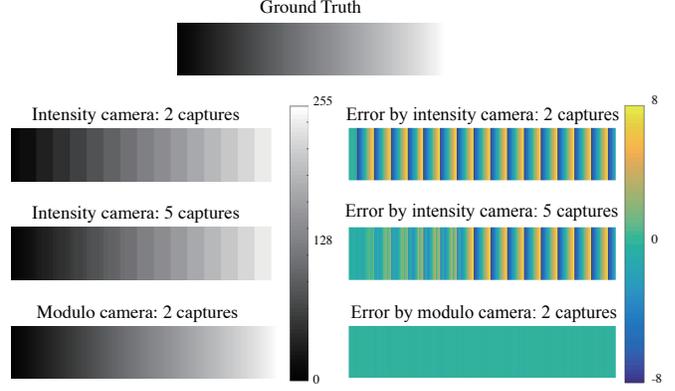


Figure 8. Comparison between multi-shot HDR using an intensity camera and UHDR using a modulo camera. Under multiple exposures, reconstructions with an intensity camera suffers from severe quantization effects in high radiance areas; a modulo camera is able to reconstruct radiance with low errors in all regions.

A more intuitive comparison between HDR with intensity images and UHDR with modulo images is shown in Fig. 8. A continuous band (1-D image) whose values linearly range from 0 to 255 is used to represent the complete dynamic range of a scene. We compare the radiance reconstruction by three methods with 4-bit cameras: an intensity camera with 2 captures, an intensity camera with 5 captures and a modulo camera with 2 captures. In the case of 2 captures, exposure difference of 16 times is used; and for the case of 5 captures, their exposures differ by twice for each capture. Left column of Fig. 8 shows the reconstructed bands with different methods, and right column is their difference maps with respect to ground truth. We can observe obvious quantization effects in both 2 captures and 5 captures cases for an intensity camera. Capturing more images is an effective strategy in reducing errors in low radiance regions, but it does not help for large radiance values, which is an inherent limitation. In contrast, multi-shot UHDR gives an almost perfect reconstruction that equally preserves details from low to high radiance levels. The analysis here applies to higher contrast images as well, but due to the difficulty in showing images higher than 8-bit in both printed papers and ordinary computer monitors we omit such an illustration here.

Sampling a large range of radiance values uniformly with a sufficiently small step (equivalent to using a higher bit depth linear camera) is meaningful to many computer vision applications like inverse rendering, photometric shape recovery, *etc.* However, to reach this goal with a linear intensity camera is not easy [13]. Highly irregular exposure ratios (like $\{1 : 1.003 : 2.985\}$ for three captures), which are extremely difficult to set in ordinary cameras, are required. What is worse, according to [13], taking

$\{2, 3, 4, 5, 6\}$ captures can only reach the dynamic range of $\{256 : 1, 761 : 1, 936 : 1, 1269 : 1, 1437 : 1\}$ under the constraint of uniform sampling. In contrast, to achieve the same result, a modulo camera only requires exposure ratios of $2^M (M \in \mathbb{Z})$, which are more commonly used setups in cameras. To maximize the dynamic range, a simple strategy is to set exposure ratio to maximum, which is 2^N . Visible radiance contrast in real life photography normally does not exceed 10,000,000:1 (within 24 stops), meaning three captures with a modulo camera is sufficient for capturing high quality UHDR images. Comparably, three captures with an intensity camera only gives an HDR with limited range, which is a known fact [13].

6. Experiments

6.1. Synthetic test

We simulate experiments for both single- and multi-shot UHDR. Ground truth images are either obtained from on-line datasets or taken with a 12-bit Sony NEX camera.

6.1.1 Single-shot UHDR

To validate our unwrapping framework, ground truth images are chosen to have maximum contrast ranging from 1024 : 1 (10 stops) to 4096 : 1 (12 stops), while actual contrasts differ as per example. The 8-bit intensity images are synthesized by truncating the brightness values over 255 to 255, and the 8-bit modulus images are easily simulated by dropping the higher bits (>8) of the ground truth. A tone mapping scheme by adaptive histogram equalization [20] is applied on our results for visualization.

In Fig. 9, ground truth, images taken by an intensity camera, images taken by a modulo camera and unwrapped results are shown from left to right. Wrapped regions in the modulus images correspond to saturated regions in the intensity images. Our unwrapping algorithm recovers these wrapped values back to their true values. As shown, images are restored successfully even with large areas of “saturation”.

6.1.2 Multi-shot UHDR

Now, we compare the multi-shot HDR recovery of an intensity camera and a modulo camera. 8-bit ground truth images are chosen to make visualizable evaluation, and 4-bit images are synthesized as outputs of both cameras.

Fig. 10 shows several results from our experiment comparing: ground truth, reconstruction by an intensity camera and a modulo camera with different number of captures. Exposure ratio is $\{1 : 16\}$ for the case of 2 captures, and $\{1 : 2 : 4 : 8 : 16\}$ for 5 captures. Close-up views highlight the visible details in the dashed squares. Quantization

noise is visible in images from an intensity camera, but not in those from a modulo camera.

6.2. Real experiments

Finally our work leverages an existing modulo camera (DFPA) to show the effectiveness of UHDR. Details associated with DFPA sensor design and camera prototype are introduced in [25, 4, 8]. We acquired 8-bit modulus images at a resolution of 256×256 .

The single-shot modulus image shown in Fig. 11 is captured with an exposure time spanning 2000 *ms*. Our proposed unwrapping algorithm is applied to the modulus image for recovery.

Multi-shot merging experiment is shown in Fig. 12. The exposure time for short and long exposure images are 100 *ms* and 4000 *ms*, respectively. The short exposure image suffers from a strong quantization effects. After fusing with a long exposure image, we can visualize the reduction in quantization noise, best seen in the close-ups.

7. Discussion

In summary, our experiments demonstrate that our proposed unwrapping algorithms significantly extend the dynamic range of the modulo camera. Theoretically, the dynamic range of recovery is not bounded in a modulus image. However, single-image unwrapping is ill-posed as recoveries are done solely based on local properties of pixel values, making the formulation only come with a smoothness constraint. Therefore, the success of unwrapping is partially dependent on the input data as well. In general, extremely large local contrast in original intensities makes the unwrapping solution more prone to errors, *i.e.*, the unwrapping operation is ideal for continuous, smooth HDR reconstruction, but it will fail in cases like impulse functions. Natural images are cases just sitting between them. Under the representation of modulus images, when fringes are dense enough to cause spatial aliasing, they are hard to be solved correctly. Fig. 13 is an example where unwrapping fails in the region of a super bright sun, as marked by the blue dashed lines. In this case, two captures are required to help recover the radiance map. The quantitative limit of unwrapping algorithm is an interesting open problem.

In multi-shot UHDR, if we only care about capturing the finest radiance levels, but not uniform sampling like a linear camera, choosing integer exposure ratios is not optimal due to overlaps of quantization levels. Therefore, non-integer exposure ratios are necessary to achieve even finer radiance recovery, but seeking the best non-integer exposure values is left as future work.

In the real experiment, as our camera prototype is still at its initial stage, it suffers from limitations like low sensor filling factor and low quantum efficiency, which result

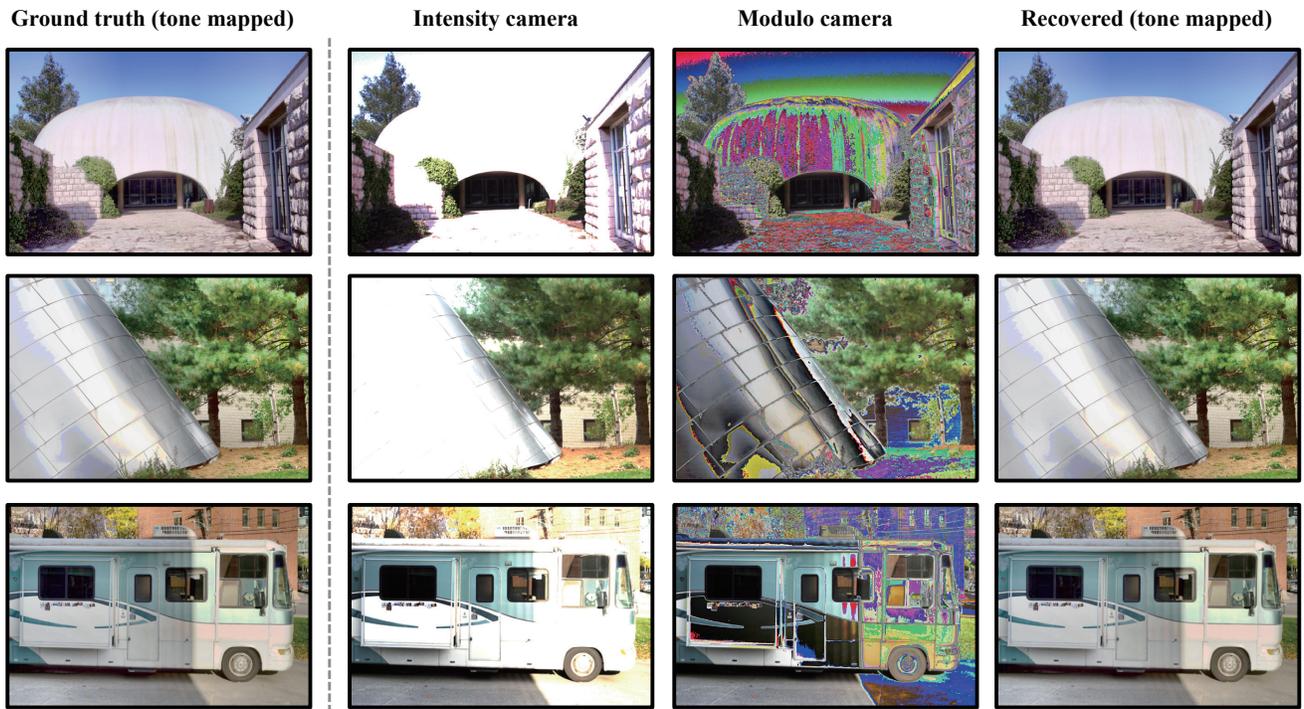


Figure 9. Single-shot UHDR. Regions in the modulus images with wrapped values correspond to saturated regions in the intensity images. Radiance data of the first row is courtesy of Raanan Fattal.

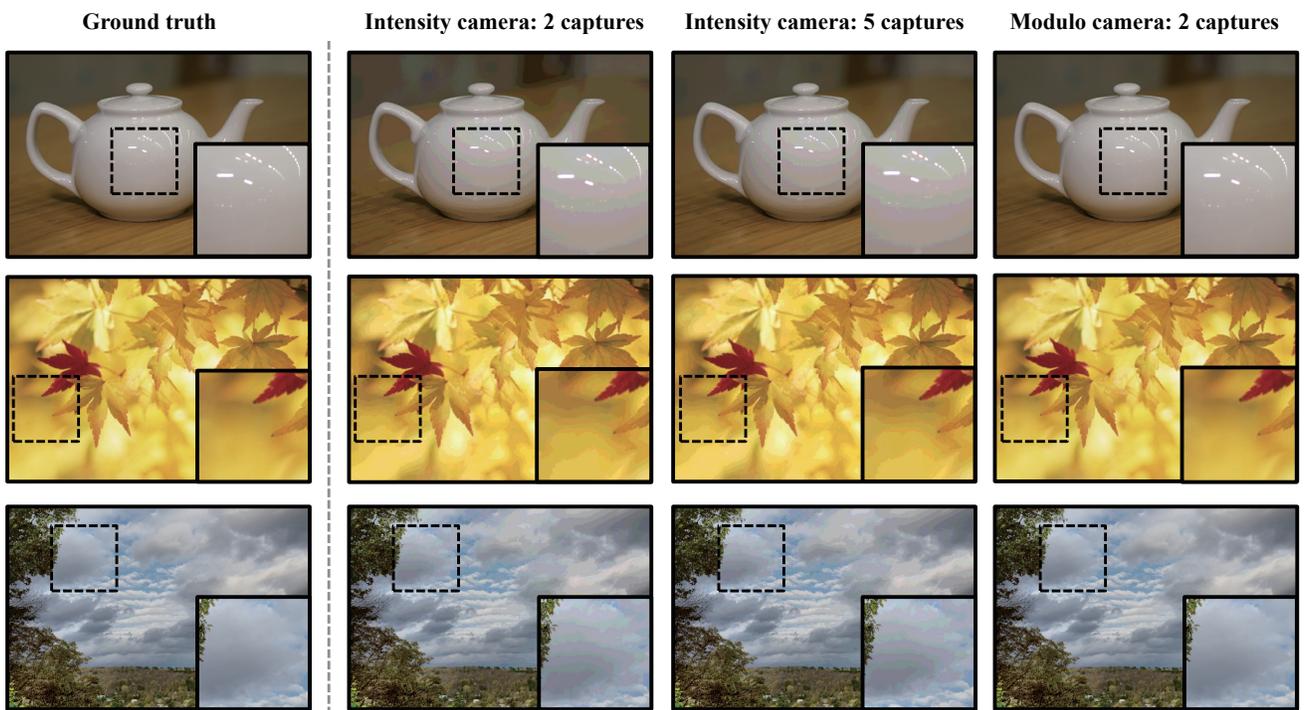


Figure 10. Multi-shot UHDR. Close-up views highlight the visible details in dashed squares. Quantization artifacts are visible in images from an intensity camera, but not in those from a modulo camera. Best viewed electronically.

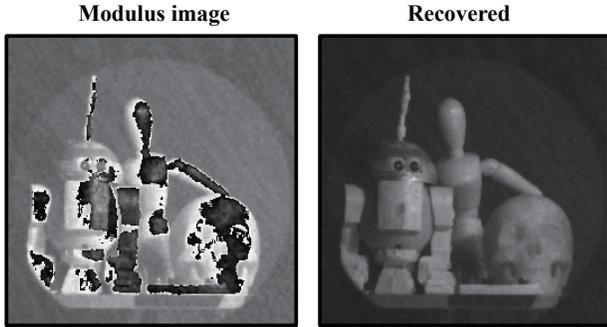


Figure 11. Real experiment: single shot. The modulus image is unwrapped to become an HDR image.

in relatively long exposure time and strong noise. As noise is more obvious in the least significant bits, it will degrade the recovery performance. To visualize the possible failures caused by noise, a longer exposure of 8000 *ms* modulo image is intentionally captured, as shown in Fig. 14. The number of wraps and noise level are both much higher than the same scene in Fig. 11. Artifacts in the unwrapped result are caused by both factors.

8. Conclusion

In conclusion, the framework of UHDR photography using a modulo camera is proposed in this paper. Saturation could be avoided with a modulo sensor so that the camera is not bounded by its bit depth. Compared with conventional single-shot photography, our framework recovers much larger dynamic range images. While taking multiple captures, apart from higher dynamic range, the proposed scheme provides finer scene radiance values that greatly improve the recovered image quality. The study presented here opens a door to addressing HDR photography problems from a novel view.

Acknowledgment We would like to acknowledge Brian Tyrrell, Thomas Gardner, Julio Estrada and Xin Xie for their valuable insight and help.

This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under air Force Contract #FA8721-05-C-0002. Opinions interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government. Boxin Shi is supported by SUTD-MIT joint post-doctoral fellowship. Sai-Kit Yeung is supported by SUTD-MIT International Design Center Grant IDG31300106 and Singapore MOE Academic Research Fund MOE2013-T2-1-159.

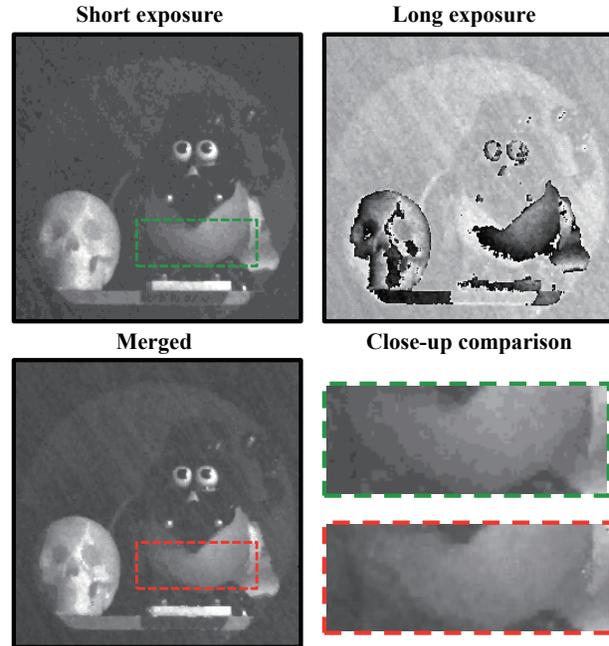


Figure 12. Real experiment: multiple shots. Brightness is linearly adjusted for visualization. After merging the short and long exposure, the details of radiance values are better preserved.

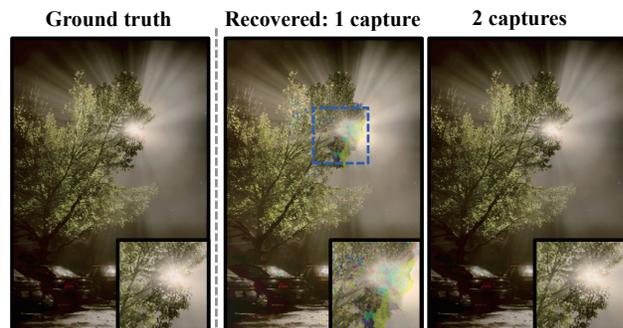


Figure 13. Failure case – synthetic data: when the local dynamic range of ground truth is extremely huge (the region in the sun), our single modulo image unwrapping method may fail, multi-shot recovery is required. Radiance map is courtesy of Jack Tumblin.

References

- [1] J. M. Bioucas-Dias and G. Valadão. Phase unwrapping via graph cuts. *IEEE Trans. on Image Processing*, 16(3):698–709, 2007.
- [2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate en-

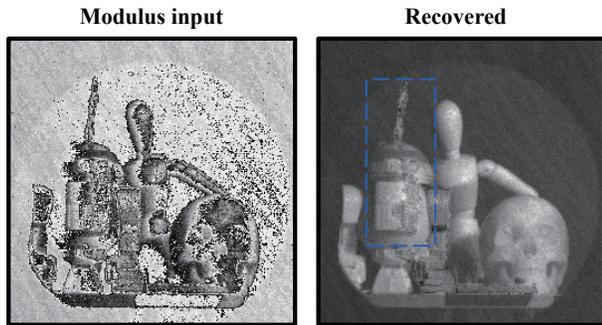


Figure 14. Failure case – real data: strong noise in this case is created by extending the exposure time.

ergy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

- [4] 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. International Society for Optics and Photonics, 2010.
- [5] S. Chavez, Q.-S. Xiang, and L. An. Understanding phase maps in MRI: a new outline phase unwrapping method. *IEEE Trans. on Medical Imaging*, 21(8):966–977, 2002.
- [6] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH classes*, page 31, 2008.
- [7] J. C. Estrada, M. Servin, and J. Vargas. 2D simultaneous phase unwrapping and filtering: A review and comparison. *Optics and Lasers in Engineering*, 50(8):1026–1029, 2012.
- [8] C. Fernandez-Cull, R. H. Shepard, B. M. Tyrrell, and R. D’Onofrio. Simultaneous dynamic pupil coding with on-chip coded aperture temporal imaging. In *Signal Recovery and Synthesis*. Optical Society of America, 2014.
- [9] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *Proc. of International Conference on Computational Photography (ICCP)*, 2009.
- [10] R. Gens. Two-dimensional phase unwrapping for radar interferometry: developments and new challenges. *International Journal of Remote Sensing*, 24(4):703–710, 2003.
- [11] D. C. Ghiglia and M. D. Pritt. *Two-dimensional phase unwrapping: theory, algorithms, and software*. Wiley New York, 1998.
- [12] D. J. Graham and D. J. Field. Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, 21(1-2):149–164, 2008.
- [13] M. D. Grossberg and S. K. Nayar. High dynamic range from multiple images: Which exposures to combine. In *Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPCV)*, 2003.
- [14] M. Gupta, D. Iso, and S. K. Nayar. Fibonacci exposure bracketing for high dynamic range imaging. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 1473–1480, 2013.
- [15] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *Technical Report #CUCS-019-14*, Department of Computer Science, Columbia University, 2014.
- [16] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [17] K. Hirakawa and P. M. Simon. Single-shot high dynamic range imaging with conventional camera hardware. In *Proc. of International Conference on Computer Vision (ICCV)*, 2011.
- [18] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. on Graphics*, 32(6):167, 2013.
- [19] V. Kolmogorov and R. Zabini. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.
- [20] G. W. Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Computer Graphics*, 3(4):291–306, 1997.
- [21] S. K. Nayar and V. Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proc. of International Conference on Computer Vision (ICCV)*, 2003.
- [22] S. K. Nayar and T. Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.
- [23] 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.
- [24] J. Tumblin, A. Agrawal, and R. Raskar. Why I want a gradient camera. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [25] B. Tyrrell, K. Anderson, J. Baker, R. Berger, M. Brown, C. Colonero, J. Costa, B. Holford, M. Kelly, E. Ringdahl, et al. Time delay integration and in-pixel spatiotemporal filtering using a nanoscale digital CMOS focal plane readout. *IEEE Trans. on Electron Devices*, 56(11):2516–2523, 2009.