Robust Multi-Image HDR Reconstruction for the Modulo Camera

Florian Lang   Tobias Plötz   Stefan Roth

Department of Computer Science, TU Darmstadt

Abstract. Photographing scenes with high dynamic range (HDR) poses great challenges to consumer cameras with their limited sensor bit depth. To address this, Zhao *et al.* recently proposed a novel sensor concept – the modulo camera – which captures the least significant bits of the recorded scene instead of going into saturation. Similar to conventional pipelines, HDR images can be reconstructed from multiple exposures, but significantly fewer images are needed than with a typical saturating sensor. While the concept is appealing, we show that the original reconstruction approach assumes noise-free measurements and quickly breaks down otherwise. To address this, we propose a novel reconstruction algorithm that is robust to image noise and produces significantly fewer artifacts. We theoretically analyze correctness as well as limitations, and show that our approach significantly outperforms the baseline on real data.

1 Introduction

Real world scenes often exhibit a significant dynamic range [17]. The intricate interplay between brightness and darkness, shadowy and sunny areas is often highly desirable from a photographer's standpoint. However, consumer cameras with image sensors that saturate when certain brightness levels are exceeded can only measure a significantly smaller dynamic range, *e.g.* 12 bit. When taking only a single image, the photographer faces the dilemma of losing detail either in the bright or in the dark parts of the scene; the whole scene cannot be captured in full detail. While various special sensors for HDR imaging have been developed, these are expensive [1], or sacrifice spatial [18] or intensity resolution [14].

Hence, various approaches aim to retain detail in the entire scene by reconstructing an HDR image from multiple captures, each with a different exposure time [4, 9, 10, 17]. As conventional image sensors saturate at some brightness level, the bit depth in bright parts of the scene is necessarily limited and the reconstruction may lead to artifacts. The *modulo camera* concept of Zhao *et al.* [22] aims to mitigate this using a novel, practical sensor that, instead of saturating, resets pixels to zero as soon as their maximal value is reached during the exposure. Hence, the least significant bits of the signal can be measured independently of its overall magnitude. This is in contrast to conventional cameras, which measure the signal correctly only up to the saturation level. To recover an HDR image from just a single modulo camera exposure, phase unwrapping techniques well-known in radar interferometry [8] or MRI [2] can be applied.
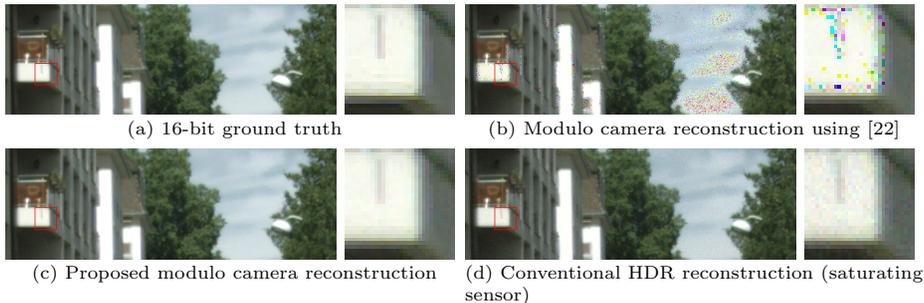
(a) 16-bit ground truth   (b) Modulo camera reconstruction using [22]

(c) Proposed modulo camera reconstruction   (d) Conventional HDR reconstruction (saturating sensor)

Fig. 1. Reconstruction of a 16-bit image from the Cityscapes dataset [3]. Multiple noisy 12-bit exposures of a modulo camera are simulated based on the ground truth in *(a)*. The approach of [22] produces visible outliers *(b)*. Our robust method *(c)* reconstructs the (noise-free) ground truth well. Reconstructing from the same number of exposures of a simulated conventional camera with saturation leads to much more noise *(d)*.

However, this requires the true image to be sufficiently smooth. For more complex and realistic scenes, [22] presents an approach for HDR reconstruction from multiple images. As in a conventional HDR pipeline, the exposure times of the captured images are chosen such that they measure different parts of the radiance range. The observed modulo values of each captured image are scaled by the exposure time and then iteratively combined into an estimate of the true HDR image. The benefit is that significantly fewer exposures are needed than with saturating sensors, making HDR imaging much more practical.

In this paper, we first consider the correctness of the reconstruction. As we will discuss, the original approach of Zhao *et al.* [22] provably works correctly, but only as long as there is no noise and exposure times can be set with perfect accuracy. Since the no-noise assumption is unrealistic, large areas of wrongly estimated pixels are produced when applied to images with typical noise levels. This is a significant impediment to the practical application of the modulo camera. To address this, we next introduce a novel, robust HDR reconstruction algorithm for modulo camera images. Analogous to the original approach, we calculate a simulated long exposure image and use its most significant bits. The crucial difference is that we additionally use the remainder of the simulated image and compare it to the actual values in the modulo capture. This way we can identify pixels for which the original algorithm would produce an incorrect reconstruction and correct these. Moreover, we make a number of theoretical contributions: First, we provide an analysis of correctness, showing that for a known noise distribution with certain properties, our algorithm reconstructs the true image even in the presence of noise. Moreover, we contribute explicit bounds on the optimal exposure times of the individual input images, which allow to assess the maximal measurable bit depth that can be recovered correctly with a chosen probability. Qualitative and quantitative experiments using realistic scenes show that our algorithm is robust against noise and consistently outperforms the original algorithm of [22]. Figure 1 shows a visual example.

2   Related Work

Capturing scenes with both bright and dark areas is challenging for consumer cameras, while the human visual system copes with such scenarios quite effortlessly. To address this gap, high dynamic range imaging aims to enable photographers to capture such scenes without loosing details. Significant research efforts have been dedicated to HDR imaging [17], resulting in techniques whose captured dynamic range even exceeds the capabilities of the human eye [12].

Multiple exposures. The perhaps most widely used family of HDR techniques is based on capturing several images with different exposure times (some under-, some overexposed) and combining these. Multishot methods can even be used with photographic film [4]. In general, they proceed by estimating the inverse of the camera response function and weighted averaging of the different images [19]. Since conventional cameras saturate at a certain level, long exposures do not add details to saturated regions. Consequently, many images with different exposure times are needed [9] to avoid quantization effects and artifacts [22]. Stumpfel *et al.* [20] report that seven conventional images are needed to capture the dynamic range of real life scenes, even with an elaborate selection of exposure times [10] and calibration. On the other hand, the modulo camera allows for an HDR reconstruction that exceeds the dynamic range of typical scenes from three modulo images alone [22] (assuming noise-free images). Other multishot HDR approaches avoid having to recover the response function [16]. All multishot methods are challenged by dynamic scenes as well as noise. Moving objects or camera ego-motion lead to ghosting artifacts in the reconstruction. Multiple approaches for ghost-free reconstruction have been proposed [7, 11, 13]. Here we focus on static scenes without camera movement, and leave the dynamic setting for future work.

HDR sensors. To increase the dynamic range without requiring multiple exposures, advanced sensors have been designed. They, for example, sacrifice spatial resolution for increased dynamic range using pixels of different size, effectively resulting in different exposures [18]. These have been realized in consumer cameras, *e.g.* the Fujifilm SuperCCD [6]. Using a high precision analog-digital converter can increase the dynamic range as well, while keeping the resolution. Another approach is to use thin-film on ASIC (TFA) sensors [1], which yield even more dynamic range. Adoption has been hindered by expensive production, and the sequential read-out for color images, requiring a three times longer exposure.

Special sensors. In contrast to conventional saturating sensors with increased bit depth, it is also possible to change the camera response function. Loose *et al.* [14] propose a sensor with a logarithmic response, which increases the dynamic range and allows for longer exposures without saturation. While the dynamic range increases, intensity resolution is reduced, especially for bright pixels. Another possibility is a sensor that measures gradients [21], which enables fine quantization and allows correcting saturated pixels. In contrast to these approaches, the modulo camera does not sacrifice spatial or intensity resolution. While the leading bits of the intensity values are lost, the details are preserved.

3   The Modulo Camera

Before introducing our approach, we formalize the properties of a conventional saturating sensor and review the modulo camera following [22]. For an ideal sensor with unbounded capacity, the observed image I arises from the scene radiance $R \in \mathbb{R}_0^+$ and the exposure time $\tau > 0$ as

$$I(\tau R) = \left\lfloor \lambda\big(\tau R + \epsilon(\tau R)\big) \right\rfloor, \tag{1}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation and λ subsumes all multiplicative factors involved in the photon-to-digit conversion, *e.g.* the quantum efficiency, the analog amplification, and the analog-to-digital conversion factor. Inevitably, the observed image will be corrupted by noise ϵ, which arises from the Poisson arrival process of photons and from the electronics involved in the imaging process. Following the literature [5], we model ϵ as intensity-dependent Gaussian noise

$$\epsilon(\tau R) \sim \mathcal{N}\big(0, \sigma^2(\tau R)\big) \qquad \text{with} \qquad \sigma^2(\tau R) = \beta_1 \tau R + \beta_2. \tag{2}$$

For simplicity, we assume *w. l. o. g.* that $0 \le R < 2^K$ and $\lambda = 1$. Hence, an exposure time $\tau = 1$ will result in a digital image with bit depth K. In practice, pixel elements have limited capacity. Hence, a saturating camera $S(\cdot)$ with bit depth $L < K$ will clip the recorded signal at a maximal value of $2^L - 1$:

$$S(\tau R) = \min(I(\tau R), 2^L - 1). \tag{3}$$

Due to the clipping all image structure in high-intensity areas will be lost. In contrast, a modulo camera $M(\cdot)$ [22] with bit depth L always retains the L least significant bits, as the sensing element is reset once it hits the maximal value:

$$M(\tau R) = I(\tau R) \bmod 2^L = I(\tau R) - k \cdot 2^L. \tag{4}$$

The true intensities can be reconstructed from the modulo image by estimating the number of rollovers k, *i.e.* how many times each pixel has been reset.

Multi-image reconstruction. Given a series of exposure times $0 < \tau_1 < \ldots < \tau_n = 1$, we observe the corresponding modulo images $M_i = M(\tau_i R)$ with unknown rollover maps k_i. We iteratively reconstruct radiance maps $\tilde{R}_i, i = 1, \ldots, n$ approximating R, as well as rollover maps $\tilde{k}_i$ approximating k_i. We assume that $k_1 = 0$ for all pixels in M_1, *i.e.* the first image has no rollovers. This is equivalent to requiring that the first exposure time is sufficiently short with $\tau_1 \le 2^{L-K}$. Hence, $\tilde{R}_1$ is obtained as the first modulo image M_1 divided by its exposure time

$$\tilde{R}_1 = \frac{M_1}{\tau_1}. \tag{5}$$

The original reconstruction algorithm of [22] now proceeds by recursively estimating the rollover map $\tilde{k}_i$ for the next exposure time and afterwards updating $\tilde{R}_i$. The rollover map for M_i can be estimated from $\tilde{R}_{i-1}$ by scaling with the new exposure time τ_i

$$\tilde{k}_i = \left\lfloor \frac{\tau_i \tilde{R}_{i-1}}{2^L} \right\rfloor. \tag{6}$$

Now we combine modulo image M_i and rollover map $\tilde{k}_i$ into a refined estimate

$$\tilde{R}_i = \frac{\tilde{k}_i 2^L + M_i}{\tau_i}. \tag{7}$$

Recalling that $\tau_n = 1$, a K-bit high dynamic range image is finally obtained as

$$\tilde{I} = \lfloor \tau_n \tilde{R}_n \rfloor = \tilde{R}_n. \tag{8}$$

Correctness. We show in the supplemental material that this reconstruction is provably exact, *i.e.* $\tilde{I} = I(\tau_n R)$, if the fraction $2^L \frac{\tau_{i-1}}{\tau_i}$ is a positive integer *and* if the recorded images are not corrupted by noise, *i.e.* $I(\tau_i R) = \lfloor \tau_i R \rfloor$, for all $i = 1, \ldots, n$. However, even miniscule amounts of noise can already cause the estimation of the rollovers $\tilde{k}_i$ from Eq. (6) to be incorrect, leading to visible artifacts, see Fig. 1b. A simple numerical example showing the limitations of the original approach is $R = 256$, $L = 8$ and $\tau_1 = 0.4$, $\tau_2 = 1$, for which it is easy to see that $\tilde{I} = 0 \neq I(\tau_2 R)$ even without noise.

Moreover, we note that image noise cannot be removed beforehand, since the modulo operation does not commute with image filtering in general. This is true even for linear filters with rational weights $\omega_i \in \mathbb{Q}$, *i.e.* there exist $x_i \in \mathbb{Z}$ such that

$$\left(\sum_i \omega_i x_i \right) \bmod n \neq \left(\sum_i \omega_i (x_i \bmod n) \right) \bmod n. \tag{9}$$

4 Robust HDR Reconstruction

Estimating the rollover image with the original reconstruction algorithm in Eq. (6) is susceptible to noise as $\tilde{k}_i$ may not match the true rollover map k_i that underlies the modulo observations M_i. We now present our algorithm that accounts for noise while reconstructing the scene radiance. In particular, we detect and correct for possible estimation errors Δ_i in $\tilde{k}_i$ such that

$$\hat{k}_i = \tilde{k}_i + \Delta_i, \qquad \Delta_i \in \mathbb{Z}. \tag{10}$$

Then we use $\hat{k}_i$ instead of $\tilde{k}_i$ for updating the radiance map as

$$\tilde{R}_i = \frac{\hat{k}_i 2^L + M_i}{\tau_i} = \frac{(\tilde{k}_i + \Delta_i) 2^L + M_i}{\tau_i}. \tag{11}$$

To find an optimal value for Δ_i, we want the simulated image $\tilde{I}_i = \lfloor \tau_i \tilde{R}_i \rfloor$ given the current radiance to be close to the simulated image $\bar{I}_i = \lfloor \tau_i \tilde{R}_{i-1} \rfloor$ based on the previous radiance estimate. Since the noise governing the non-rounded image intensities is assumed Gaussian, we formulate the following least-squares problem:

$$\min_{\Delta_i} \mathcal{L}(\Delta_i) \doteq \min_{\Delta_i} \| \lfloor \tau_i \tilde{R}_i \rfloor - \lfloor \tau_i \tilde{R}_{i-1} \rfloor \|^2 \tag{12}$$

$$= \min_{\Delta_i} \| \tau_i \tilde{R}_i - \lfloor \tau_i \tilde{R}_{i-1} \rfloor \|^2 \tag{13}$$

$$= \min_{\Delta_i} \| (\tilde{k}_i + \Delta_i) 2^L + M_i - \lfloor \tau_i \tilde{R}_{i-1} \rfloor \|^2, \tag{14}$$

which holds as $\tau_i \tilde{R}_i \in \mathbb{Z}$ and we plugged in Eq. (11). Using Eq. (6), we have

$$\mathcal{L}(\Delta_i) = \left\| \Delta_i 2^L + M_i - \underbrace{\left(\lfloor \tau_i \tilde{R}_{i-1} \rfloor - \left\lfloor \frac{\tau_i \tilde{R}_{i-1}}{2^L} \right\rfloor 2^L \right)}_{\doteq D_i} \right\|^2. \tag{15}$$

Since M_i and D_i are modulo values, it holds that

$$-2^L < M_i - D_i < 2^L, \tag{16}$$

which implies that $\mathcal{L}(\Delta_i) > \mathcal{L}(0)$ for $|\Delta_i| \geq 2$. Thus, the cost $\mathcal{L}$ is minimized for $\Delta_i \in \{-1, 0, 1\}$. We can now read off the optimal value as

$$\Delta_i = \arg\min_{\Delta} \|\Delta 2^L + M_i - D_i\|^2 = \begin{cases} +1, & M_i - D_i < -2^{L-1} \\ -1, & M_i - D_i > +2^{L-1} \\ 0, & \text{else.} \end{cases} \tag{17}$$

This ensures that the reconstructed $\tilde{I}_i$ lies in the interval $[\bar{I}_i - 2^{L-1}, \bar{I}_i + 2^{L-1}]$.

Optimality. We now establish optimality conditions and show that our novel algorithm is indeed more robust than the original approach of [22]. To simplify notation, we write $I_i = I(\tau_i R)$ for the ideal image taken with an unbounded sensor. We then consider the compound noise between two images

$$e_i \doteq I_i - \frac{\tau_i}{\tau_{i-1}} I_{i-1} \tag{18}$$

$$= \epsilon(\tau_i R) - r\big(\tau_i R + \epsilon(\tau_i R)\big) - \frac{\tau_i}{\tau_{i-1}} \epsilon(\tau_{i-1} R) + \frac{\tau_i}{\tau_{i-1}} r\big(\tau_{i-1} R + \epsilon(\tau_{i-1} R)\big), \tag{19}$$

where $r(x) \doteq x - \lfloor x \rfloor$ is the rounding error with $0 \leq r(x) < 1$.

Theorem 1. *If $|e_i| \leq 2^{L-1} - 1$ for $i = 2, \ldots, n$, the robust algorithm reconstructs I_i in every iteration correctly,* i.e. $\tilde{I}_i \doteq \tau_i \tilde{R}_i = I_i$. *Especially it holds that $\tilde{I}_n = \tilde{R}_n = I_n$ and hence the true radiance map is reconstructed as well as possible.*

Proof sketch. We proceed by induction. By assumption, the first modulo image has no rollovers, *i.e.* $I_1 = M_1$, and hence $I_1 = \tau_1 \tilde{R}_1$. The induction step proceeds by contradiction and relies on a case distinction on the possible values of Δ_i and whether the true k_i is over- or underestimated. In particular, we show that

$$\tau_i \tilde{R}_i \neq I_i \implies |e_i| > 2^{L-1} - 1. \tag{20}$$

We refer to the supplemental material for details, but note that it is somewhat intuitive that Eq. (17) leads to a correct reconstruction of the rollover maps. Consequently, the robust algorithm leads to the radiance being reconstructed as well as possible even for substantial amounts of noise.

This is in contrast to the original reconstruction approach of [22], for which the robustness depends on the image values itself. For some image values, it tolerates $|e_i| = 2^L - 1$, while for others $|e_i| = 1$ already produces wrong results. As we have seen in the simple example above, even rounding errors can already cause the original approach to fail. Even if rounding errors could be eliminated by carefully choosing exposure times, a small amount of noise leads to similar failures.

5 Robust Capture Protocol

Theorem 1 shows that our reconstruction algorithm recovers the true intensities as long as the compound noise of imaging $\tau_i R$ and $\tau_{i-1} R$ is small enough. Since the noise and rounding error of $\tau_{i-1} R$ enter the definition of e_i with a factor of $\tau_i / \tau_{i-1} > 1$ (Eq. 19), we can control the distribution of e_i by choosing the ratio of exposure times τ_i / τ_{i-1}. For practical applications it is highly desirable to find an explicit exposure time schedule such that the number of exposures is minimized while guaranteeing correctness with high probability. Let us consider that in iteration i we want to reconstruct a pixel correctly with some probability p. Then we aim to find the maximal τ_i given τ_{i-1} such that

$$\mathrm{P}\big[|e_i| \leq 2^{L-1} - 1\big] \geq p. \tag{21}$$

We can bound the probability on the left-hand side by the CDF of the absolute value of a Gaussian (for details see supplemental material)

$$\mathrm{P}\big[|e_i| \leq 2^{L-1} - 1\big] \geq \mathrm{P}\left[\big|\mathcal{N}(0, \sigma_i^2)\big| \leq 2^{L-1} - 1 - \frac{\tau_i}{\tau_{i-1}}\right] \tag{22}$$

with

$$\sigma_i^2 = \beta_1 \tau_i R \left(1 + \frac{\tau_i}{\tau_{i-1}}\right) + \beta_2 \left(1 + \frac{\tau_i^2}{\tau_{i-1}^2}\right). \tag{23}$$

This can be used to derive an upper bound on τ_i such that Eq. (21) is satisfied:

$$\tau_i \leq \tau_{i-1} \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{24}$$

with

$$a = \beta_1 \tau_{i-1} R + \beta_2 - \left(\Phi^{-1}\big(\tfrac{1}{2} + \tfrac{1}{2}p\big)\right)^{-2} \tag{25}$$

$$b = \beta_1 \tau_{i-1} R + \left(\Phi^{-1}\big(\tfrac{1}{2} + \tfrac{1}{2}p\big)\right)^{-2}(2^L - 2) \tag{26}$$

$$c = \beta_2 - \left(\Phi^{-1}\big(\tfrac{1}{2} + \tfrac{1}{2}p\big)\right)^{-2}(2^{L-1} - 1)^2, \tag{27}$$

where Φ^{-1} is the inverse CDF of a standard normal distribution. We again refer to the supplemental material for details. The noise parameters β_1, β_2 can be estimated for a specific camera, $e.g.$ with the technique of Foi $et\ al.$ [5]. However, the scene radiance R is needed in the calculation of τ_i. We make a conservative guess and use the maximal value, $i.e.$ 2^K. That way every pixel is reconstructed correctly with probability at least p, while pixels with a lower intensity will be reconstructed with higher probability.

In general, if we always take the upper bound in Eq. (24) the resulting series will converge to a limiting exposure time τ^*. We state an explicit formula for τ^* in the supplementary. Note, that the limiting exposure time can be greater than 1 and that it effectively allows us to characterize the $maximum\ bit\ depth$ that can be resolved using a modulo camera with bit depth L that, in contrast to [22], assumes realistic noise with parameters β_1, β_2. If β_2 is too large, τ^* will become complex-valued. In contrast, if β_1 is too large, we get $\tau^* < 1$, $i.e.$ we cannot reconstruct R with the required certainty.

6   Evaluation

We now analyze our robust reconstruction algorithm on real HDR data. We use the Cityscapes dataset [3], which provides 16-bit linear HDR images of road scenes. Since the images have already been debayered, we apply a color filter array to remove the interpolated pixels. Given the 16-bit images, we simulate the imaging process of an L-bit modulo camera as well as a conventional L-bit camera that goes into saturation. Note that we need to rely on simulation here, as modulo sensors have not been manufactured yet apart from early prototypes [22]. We choose the base parameters of the noise distribution as $10^{-8} \leq \beta_1 \leq 10^{-2}$ and $\beta_2 = 0.01\beta_1$, which are reasonable for consumer grade cameras [15]. In practice a certain camera has a fixed bit depth. In order to nevertheless compare the results across different bit depths for the same camera, we normalize the noise parameters such that an L-bit camera has $\beta_1^{(L)} = \alpha\beta_1$ and $\beta_2^{(L)} = \alpha^2\beta_2$, where $\alpha = 2^L - 1$. This implies that an L-bit camera collects 2^{K-L} times more photons to increment the intensity by one than the K-bit ground truth camera, thus improving the signal-to-noise ratio. This assumption is reasonable since cameras with a smaller bit depth can afford to integrate bigger capacitors instead.

Maximal bit depth. We first analyze the maximum bit depth that can be reconstructed and compare the proposed approach against the original reconstruction from [22]. We choose the exposure times according to Eq. (24) such that a pixel is reconstructed correctly with $p > 0.99$ in each iteration. While the effect of having a maximum of 1% of pixels potentially incorrect could be drastic when many iterations need to be made, this is a worst case estimate. In practice, many fewer errors occur, since the image is not bright everywhere. Moreover, we observe that choosing $p > 0.999$ or 0.9999 did not yield significantly better results in practice, but required more images. Figure 2 shows how many bits can be reconstructed accurately from 2 and 5 exposures, respectively, as well as in the hypothetical case of infinitely many exposures, where the last exposure time is given by τ^*. Depending on the bit depth of the camera and the noise characteristics, we achieve very high maximal exposure times and therefore a large dynamic range that can be recovered. For example, roughly 18 bits can be reconstructed with only 2 images, if $\beta_1 = 10^{-5}, \beta_2 = 10^{-7}$ and $L = 12$. Moreover, with just 5 exposures we already achieve a bit depth that is close to the theoretical maximum, meaning that the exposure time schedule given by Eq. (24) quickly converges to its limit. It is important to note that in the presence of realistic amounts of sensor noise, and even with our robust reconstruction algorithm, the modulo camera does not enable HDR images with an unbounded dynamic range, unlike what is claimed in [22]. Nevertheless, our derivation shows that even with realistic camera noise, significant gains in dynamic range can be achieved. Further improvements could potentially be made by considering the modulo images jointly, *c.f.* [8], instead of sequentially. We leave that for future work.

HDR reconstruction. We now compare the results from the original algorithm, our robust approach, and a simple HDR reconstruction method based on images from a saturating sensor, all with the same number of exposures. For the latter,
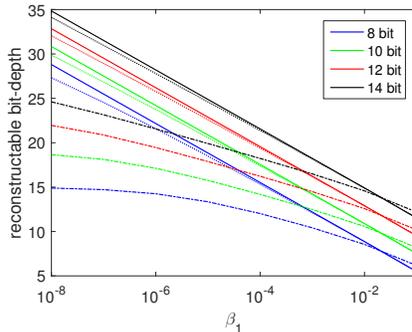
Fig. 2. Maximal bit depth that can be reconstructed from modulo images under realistic image noise using 2 (dot-dashed), 5 (dotted) and infinitely many (solid) images for varying bit depth and noise parameters. We set $\beta_2 = 0.01\beta_1$.

we average over all images in which a pixel is neither under- nor oversaturated. We measure the quality of the reconstruction using the peak signal-to-noise ratio between the debayered and tone-mapped images: $\mathrm{PSNR} \doteq 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right)$, where $\mathrm{MAX}_I = 2^{16} - 1$ is the maximum intensity of the reconstructed image and MSE is the mean squared error. We determined the number of required exposures using the bound from Section 5. In the case of $\tau^* < 1$ we padded the first two exposure times to reach $\tau_n = 1$, for $\tau^* \ll 1$, *i.e.* extreme noise, we used 6 predefined exposures. Other schedules did not significantly affect the findings.

As we can see in Fig. 3 for low and moderate noise, the result of our robust algorithm is very close to the noisy ground truth. For strong noise the results become slightly suboptimal, since the calculated exposure times are too small, *i.e.* $\tau^* < 1$. The original modulo reconstruction approach of [22] and a simple HDR pipeline with a saturating sensor do much worse, since we are using as few exposures as possible. Especially for weak noise we often only need 2 images for our approach. Note, that the HDR pipeline with a saturating sensor combines a long-exposure image with a short-exposure image by appropriately upscaling the short-exposure image and subsequently averaging both images. In the upscaled image this effectively leads to an amplification of the noise parameter β_2 that is otherwise independent of the exposure time. This does not occur when reconstruction an HDR image using our approach as in every step only information from a single image is used to add detail to the reconstruction. Figure 4 shows an image from another dataset. We experience the same effects as in Fig. 1.

7   Conclusion

Modulo cameras have promised to be an interesting, practical alternative to conventional sensors that enable HDR reconstruction from a small number of exposures. However, as we have shown in this paper, the original reconstruction algorithm fails even in the noise-free case. Realistic image noise severely limits
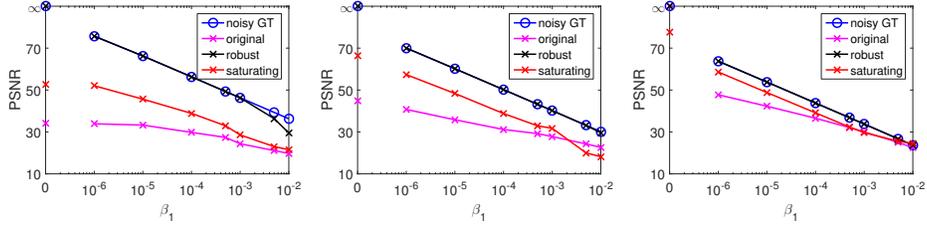
Fig. 3. PSNR of HDR reconstruction on the Cityscapes dataset with various methods for different noise levels ($\beta_2 = 0.01\beta_1$). *(from left to right)* 10, 12, 14 bit cameras. The overall noise increases with the bit depth since fewer photons are needed to increment the intensity value in a pixel site.

Fig. 4. HDR reconstruction of a scene in the Stanford Memorial Church. We use the HDR reconstruction of [4] from 18 exposures as ground truth for simulation. *(from left to right)* ground truth, conventional reconstruction using simulated saturating images, reconstruction from modulo images using the original algorithm, and using the proposed robust reconstruction approach. We simulated 12-bit cameras and added moderate amounts of noise ($\beta_1 = 10^{-3}, \beta_2 = 10^{-5}$).

the attainable image quality. We proposed a novel, robust HDR reconstruction algorithm for images from a modulo camera and established clear criteria for its correctness. It can deal with significant amounts of noise and allows to assert stringent bounds on the probability of a successful reconstruction. We derived an optimal exposure time schedule for images with realistic noise and empirically showed that our robust algorithm performs very close to the ground truth, clearly outperforming the original modulo camera algorithm as well as standard HDR pipelines. Future work should consider extending the multi-image approach to dynamic scenes. Furthermore, an evaluation of a physical modulo camera in terms of the attainable noise level as well as production costs would be useful to determine the practicality of modulo sensors in consumer cameras.

References

1. Böhm, M., Blecher, F., Eckhardt, A., Schneider, B., Benthien, S., Keller, H., Lulé, T., Rieve, P., Sommer, M., Lind, R.C., Humm, L., Daniels, M., Wu, N., Yen, H., Efron, U.: High dynamic range image sensors in thin film on ASIC technology for automotive applications. In: Ricken, D.E., Gessner, W. (eds.) Advanced Microsystems for Automotive Applications, pp. 157–172. Springer (Jan 1998)
2. Chavez, S., Xiang, Q.S., An, L.: Understanding phase maps in MRI: A new cutline phase unwrapping method. IEEE Trans Med Imaging 21, 966–977 (Sep 2002)
3. Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The Cityscapes dataset for semantic urban scene understanding. In: IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR) (Jun 2016)
4. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. pp. 369–378
5. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE Trans. Image Process. 17(10), 1737–1754 (Oct 2008)
6. Fuji Photo Film Co., Ltd.: 4th-generation Super CCD (2003), http://www.fujifilmusa.com/shared/bin/4thGenSUPERCCDBrochure.pdf
7. Gallo, O., Gelfand, N., Chen, W., Tico, M., Pulli, K.: Artifact-free high dynamic range imaging. IEEE Int. Conf. on Comp. Photog. (ICCP) (Apr 2009)
8. Goldstein, R.M., Zebker, H.A., Werner, C.L.: Satellite radar interferometry: Two-dimensional phase unwrapping. Radio Science 23(4), 713–720 (July 1988)
9. Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H.P., Lensch, H.P.A.: Optimal HDR reconstruction with linear digital cameras. In: IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR). pp. 215–222 (Jun 2010)
10. Grossberg, M.D., Nayar, S.K.: High dynamic range from multiple images: Which exposures to combine? In: ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV) (Oct 2003)
11. Heo, Y.S., Lee, K.M., Lee, S.U., Moon, Y., Cha, J.: Ghost-free high dynamic range imaging. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) Asian Conf. on Comp. Vis. (ACCV). pp. 486–500. Springer (2010)
12. Hoefflinger, B. (ed.): High-Dynamic-Range (HDR) Vision. Springer (2007)
13. Jacobs, K., Loscos, C., Ward, G.: Automatic high-dynamic range image generation for dynamic scenes. IEEE Computer Graphics and Applications 28(2), 84–93 (Mar 2008)
14. Loose, M., Meier, K., Schemmel, J.: A self-calibrating single-chip CMOS camera with logarithmic response. IEEE Journal of Solid-State Circuits 36(4), 586–596 (Apr 2001)
15. Mäkitalo, M., Foi, A.: Noise parameter mismatch in variance stabilization, with an application to Poisson–Gaussian noise estimation. IEEE Trans. Image Process. 23(12), 5348–5359 (Oct 2014)
16. Mann, S., Picard, R.W.: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In: Proc. of Imaging Science and Technology. pp. 442–448 (May 1995)
17. Mantiuk, R.K., Myszkowski, K., Seidel, H.P.: High dynamic range imaging. In: Wiley Encyclopedia of Electrical and Electronics Engineering. John Wiley & Sons, Inc. (2015)

18. Nayar, S.K., Mitsunaga, T.: High dynamic range imaging: Spatially varying pixel exposures. In: IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR) (Jun 2000)
19. Robertson, M.A., Borman, S., Stevenson, R.L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. Electronic Imaging 12(2), 219–228 (2003)
20. Stumpfel, J., Tchou, C., Jones, A., Hawkins, T., Wenger, A., Debevec, P.: Direct HDR capture of the sun and sky. In: Proc. of the 3rd Int. Conf. on Comp. Graph., Virt. Real., Visualis. and Interact. in Africa. pp. 145–149. AFRIGRAPH '04 (Nov 2004)
21. Tumblin, J., Agrawal, A., Raskar, R.: Why I want a gradient camera. In: IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR). pp. 103–110 (Jun 2014)
22. Zhao, H., Shi, B., Fernandez-Cull, C., Yeung, S.K., Raskar, R.: Unbounded high dynamic range photography using a modulo camera. In: Int. Conf. on Comp. Photog. (ICCP) (Apr 2015)