

Physical High Dynamic Range Imaging with Conventional Sensors

Holger Meuel, Hanno Ackermann, Bodo Rosenhahn, Jörn Ostermann
Institut für Informationsverarbeitung (TNT)
Gottfried Wilhelm Leibniz Universität Hannover
Hannover, Germany
{meuel, ackerman, rosenhahn, office}@tnt.uni-hannover

Abstract—This paper aims at simplified high dynamic range (HDR) image generation with non-modified, conventional camera sensors. One typical HDR approach is exposure bracketing, *e.g.* with varying shutter speeds. It requires to capture the same scene multiple times at different exposure times. These pictures are then merged into a single HDR picture which typically is converted back to an 8-bit image by using tone-mapping. Existing works on HDR imaging focus on image merging and tone mapping whereas we aim at simplified image acquisition. The proposed algorithm can be used in consumer-level cameras without hardware modifications at sensor level. Based on intermediate samplings of each sensor element during the total (pre-defined) exposure time, we extrapolate the luminance of sensor elements which are saturated after the total exposure time. Compared to existing HDR approaches which typically require three different images with carefully determined exposure times, we only take *one* image at the longest exposure time. The shortened total time between start and end of image acquisition can reduce ghosting artifacts. The experimental evaluation demonstrates the effectiveness of the algorithm.

I. INTRODUCTION

This paper is concerned with high dynamic range (HDR) image generation. This problem is important whenever the dynamic range of a scene exceeds the dynamic range of the image sensor capturing the image (Fig. 1). If the exposure time is set too short at the image acquisition, the photo cells corresponding to dark parts of the scene do not receive enough light, so the image there remains black or very dark at least (Fig. 1a). If, conversely, the exposure time is set too long, the cells corresponding to bright parts of the scene receive too much light, resulting in overexposed shots (Fig. 1b). The problem how to create images with sufficient contrast in both dark and bright parts of the image is not only of interest in industrial applications but also for consumers. For HDR image generation with commercially available cameras, multiple images of the scene need to be taken with different exposure times first. The exposure times have to be determined such that both bright and dark parts of the scene have sufficient contrast in at least one of the shots. This requires an experienced photographer or a sophisticated software. Nonetheless, this process is time-consuming, *i.e.* it requires a quasi static scene, or results in artifacts for motion in the scene (ghosting). After the acquisition of several low dynamic range (LDR) images, they need to be merged into a single HDR image. To cover the increased

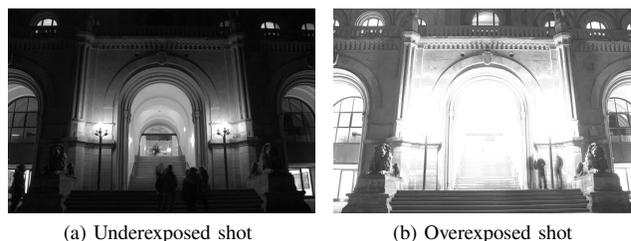


Figure 1: Example of a scene with high dynamic range which cannot be covered by the camera sensor: (a) if the exposure time is adjusted to accurately represent the lightest parts (lamps), details in the dark image parts are lost, or vice versa (b).

dynamic range, more bits are necessary for the HDR image (*e.g.* represented by high bit depths in OpenEXR format) [1], [2]. Typically, the HDR image is converted back to an 8-bit-per-color LDR image (*e.g.* JPEG) afterwards by using tone-mapping techniques [3], [4]. In this work we will not focus on tone mapping but aim at the physical image acquisition and thus, shortly review **related work**.

We distinguish three types of HDR image generation: firstly, the sensor design itself can be modified in order to cope with an increased dynamic range in the scene, *e.g.* by introducing anti-blooming drains [5], additional charge-to-voltage converters [6] or other hardware modifications [7], [8], [9], [10] not available in commercial camera sensors. As a variation of this approach, additional elements (like a controllable spatial light modulator) which are introduced in the light path in front of each sensor pixel have been proposed [11]. All approaches aiming at sensor modifications have in common that they are extremely expensive, highly complex and thus impractical for consumer camera sensors. Secondly, spatial exposure bracketing aims at the simultaneous generation of various images with different cameras and same settings. The amount of light reaching each sensor is controlled for instance by optical filters [12]. Obviously, by introducing additional elements, similar disadvantages apply as for the approaches mentioned above. Thirdly, time exposure bracketing aims at the generation of several LDR images in a temporal sequence with different exposure settings (*e.g.* [13]). Since it can be

performed with any camera, this is the most popular HDR generation approach. However, to generate shots representing exactly the same field of view, a fixed camera (*e. g.* mounted on a tripod) is necessary. Moreover, due to the relatively long time for the entire acquisition of all LDR images, those algorithms are prone to ghosting artifacts which are caused by motion in the scene. These artifacts become more prevalent for long acquisition times, thus it is in general desirable to reduce the entire acquisition time to a minimum.

For a highly accurate image matching, image alignment techniques are applied. To remove ghosting artifacts, *e. g.* patch match was proposed for images [14] as well as for HDR video [15]. It finds and merges patches from different exposed shots and extends them to the final image or video by combining optical flow techniques with patch-based reconstruction. A specifically adopted region-adaptive multiple-exposure fusion algorithm was proposed in [16], whereas in [17] a statistically optimal weighting function under the assumption of compound-Gaussian noise for a multiple-exposure sequence was proposed. Recently, a merging method that is robust to misalignment, based on a pairwise frequency-domain temporal filter operating was proposed [18]. One disadvantage, which all time exposure bracketing methods have in common, is that the alignment of different LDR images is challenging due to the long overall acquisition time for all exposures (especially for hand-held exposures), and it may not be possible to entirely remove or conceal ghosting artifacts.

To reduce the total acquisition time, the sensor sensitivity is increased (*i. e.* increase the amplification by using high ISO settings), either for the entire image or parts of it (*e. g.* each second scanline may be read with different ISO settings [19], [20]). Since noise is also further amplified in this case, special noise reduction has to be applied afterwards ([21]), which may impair the dynamic range and overall image quality again.

Here, we suggest the acquisition of a single image is indeed sufficient if the exposure time is set “large enough”. Although “large enough” typically may mean that the exposure time for our proposed approach is comparable to the largest one from a common time expose bracketing approach, the overall acquisition time is limited to this time span for our approach and consequently much shorter than the total acquisition time for taking several shots with different exposure times in the case of common time exposure bracketing. This procedure obviously causes over-exposure in some parts of the image. To nonetheless obtain all the relevant information about tone contrast, we propose to *sample* the image or parts thereof multiple times during the exposure. By sampling, we mean to read out the photo sensor yet to continue the exposure process.

This idea is motivated by the fact that the bottleneck between image sensor, image processing chip and memory is the transfer of data to the memory, not the transfer between image sensor and the chip. Commercially available (mid- and upper-class) cameras, *e. g.* digital single-lens reflex cameras (DSLRs), can take rapid sequences of pictures until the capacity of an internal cache in the image processing chip is reached. The limit how fast consecutive images can be taken is mainly determined by

mechanical properties, *i. e.* how fast the shutter can operate.

However, current consumer-level cameras do not allow to *read* the sensor information during the exposure process. This operation would only require changes to the firmware of the image processing chip which needs to allow this operation, not the image sensor itself. We therefore propose an algorithm that approximates the proposed idea to read out the image sensor multiple times during the exposure process.

The **contributions** can be summarized as follows:

- An algorithm to infer values of sensor elements outside the dynamic range of the image sensor is proposed.
- We propose to read-out intermediate images during the exposure process.
- The images used for estimating out-of-range sensor values can be taken by a sampling procedure.
- Only the image with longest exposure time need to be taken.
- Because the overall time required to take all the images is reduced, ghosting is less likely to occur.
- An analysis of the error caused by extrapolating quantized intermediate sampling values (“sampling error”) is presented.

The remaining paper is organized as follows: the basic model is introduced in Section II. It rests upon the idea that images can be sampled without interrupting the exposure process. In Section III, the impact of the sampling error introduced by our approach is analyzed. Experimental results are presented in Section IV before Section V concludes the paper.

II. SAMPLING DURING EXPOSURE

Let \mathcal{I} be the image and $\mathcal{I}(x, y, t)$ be the value of a pixel at position (x, y) at time t . Assume that the value of an pixel (x^*, y^*) measured after an exposure time t_E exceeds the maximal sensor limit v_{\max} . Thus, instead of the true value $\bar{v}(x^*, y^*, t_E)$ a measurement $\hat{v}(x^*, y^*, t_E) = v_{\max}$ is obtained with v_{\max} being the maximal measurable value of one pixel due to physical limits.

Let $t_s = t_E/s$ indicate some shorter time span with s being the number of intermediate read-outs, *i. e.* is $s > 1$. Under the assumption that the light source is not too bright, we can expect that the measurement $\hat{v}(x^*, y^*, t_s) < v_{\max}$ if s is sufficiently large. This assumption is the basis of all HDR image composition algorithms. If it is possible to *read* the photo cell at (x^*, y^*, t_s) at time t_s yet not to reset it, then we may estimate $\hat{v}(x^*, y^*, t_E)$. To reconstruct the non-measurable value $\hat{v}(x^*, y^*, t_E) > v_{\max}$ of sensor cell (x^*, y^*) at time t_E , all that is necessary to do is to determine a time $t_s < t_E$ at which $\hat{v}(x^*, y^*, t_s) \leq v_{\max}$. In the absence of any noise and assuming a constant illumination during the exposure time t_E (the latter is usually satisfied for all but very long exposure times), $\hat{v}(x^*, y^*, t_E)$ can be approximated by

$$\hat{v}(x^*, y^*, t_E) \approx s \cdot \hat{v}(x^*, y^*, t_s) \quad (1)$$

since $t_s = t_E/s$.

If we assume temporal noise, the measured value $\hat{v}(x^*, y^*, t_s)$ of pixel (x^*, y^*) at time t_s is not exact, thus the extrapolation $s \cdot \hat{v}(x^*, y^*, t_s)$ is perturbed by noise. In fact, for larger values of s (i.e. shorter times t_s), the error introduced by the extrapolation increases (see Section III). Denote by $E\{\hat{v}(x^*, y^*)\}$ the expectation value of the reconstructed pixel value $\hat{v}(x^*, y^*)$ at time t_E , let k indicate multiples of the shortest exposure time t_E/s , $k = 1, \dots, s$ and let k_{\max} indicate the last sampling step for which $\hat{v}(x^*, y^*, t_k) \leq v_{\max}$. Then, by using Eq. (1), we can define

$$\begin{aligned} \hat{v}(x^*, y^*, t_E) &\approx E\{\hat{v}(x^*, y^*)\} \\ &= \frac{1}{k_{\max}} \sum_{j=1}^{k_{\max}} \frac{s}{j} \hat{v}(x^*, y^*, \frac{j}{s} t_E). \end{aligned} \quad (2)$$

Additional weighting may be applied as investigated below (see next section and Eq. (6)).

One appealing property of this procedure is that it requires no expensive changes to the photo sensor. Only the image processor and the firmware are required to permit read-outs during the exposure. This operation should also be implementable without expensive changes of the hardware itself.

III. ERROR ANALYSIS

The rounding error of each sampling step is $|e| \leq \frac{1}{2}$. Therefore, for the k -th out of s sampling steps, the extrapolation error, i.e. the extrapolation of the rounding error, is limited to $|e_k| \leq \frac{s}{2k}$. The maximum error $|e_k| \leq \frac{1}{2}s$ can occur for $k = 1$ and decreases for larger values of k .

Let $[\cdot]$ denote the rounding function to the nearest integer value. The particular sampling error e_k^* corresponding to pixel (x^*, y^*) is calculated as the difference between the true, non-measurable value $\bar{v}(x^*, y^*, t_E)$ and the extrapolated value $\hat{v}(x^*, y^*, t_E) \approx \frac{s}{k} \cdot \hat{v}(x^*, y^*, t_k) = \frac{s}{k} \cdot \left[\frac{k}{s} \cdot \bar{v}(x^*, y^*, t_E) \right]$. In the noise-free case (see Eq. (1)), the error e_k^* can thus be modeled as

$$e_k^* = \left| \bar{v}(x^*, y^*, t_E) - g_k \right| \quad (3)$$

where $g_k = \frac{s}{k} \cdot \left[\frac{k}{s} \cdot \bar{v}(x^*, y^*, t_E) \right]$. Assuming zero-mean white noise $\eta_k = \mathcal{N}(0, \sigma)$, we obtain

$$e_k^* \leq \left| \bar{v}(x^*, y^*, t_E) - \left(g_k + \frac{s}{k} [\eta_k] \right) \right| + \nu_k. \quad (4)$$

Denote by η , g and ν the random variables with realizations η_k , g_k and $\nu_k = \frac{s}{k} \cdot 1$, respectively. The correction term ν_k of Eq. (4) is due to rounding g_k and η_k separately. The true value of ν_k typically is much smaller, consequently ν_k represents the upper bound of the error. Taking the expectation value as in Eq. (2) to reduce the influence of temporal noise, we may decompose $\hat{v}(x^*, y^*)$ into two parts: the expectation over k of the second term in Eq. (4) and the expectation of the Gaussian noise terms η_k . With the discrete random variable ν , we therefore have

$$E\{e^*\} \leq \left| \bar{v}(x^*, y^*, t_E) - E\{g\} - E\{[\eta]\} \right| + E\{\nu\}. \quad (5)$$

with $\nu_k = \frac{s}{k}$.

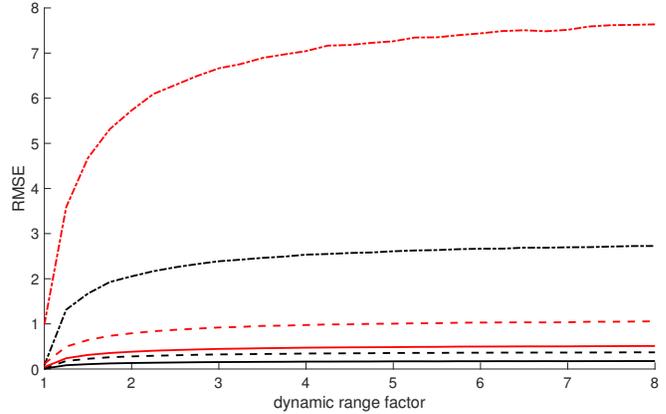


Figure 2: *Figure in color.* Results using simulated images. horizontal axis: dynamic range factor d_R which specifies how much the maximal sensor value is exceeded; vertical axis: RMSE. Black lines: results of noise-free experiment with $\sigma = 0$; red lines: experiment with noise $\sigma = 3$. Dash-dotted lines: naive extrapolation by Eq. (1); dashed lines: reconstruction by Eq. (5) and uniform probability distributions; solid lines: non-uniform probability distributions using Eqs. (5) and (6).

The noise η_k follows a normal distribution, $\eta_k \sim \mathcal{N}(0, \sigma)$, hence (with Equation (4)) we have $E\{\eta\} \sim \mathcal{N}(0, \frac{s}{k_{\max}}(1 + \frac{1}{2} + \dots + \frac{1}{k_{\max}})\sigma)$.

Motivated by the higher signal-to-noise ratio (SNR) for longer exposure times [22], [23], we can reduce the impact of sampling errors from earlier sampling steps by assuming a non-uniform probability distribution function w_k for the expectations in Eq. (5),

$$w_k = \frac{k}{\sum_{j=1}^s j}. \quad (6)$$

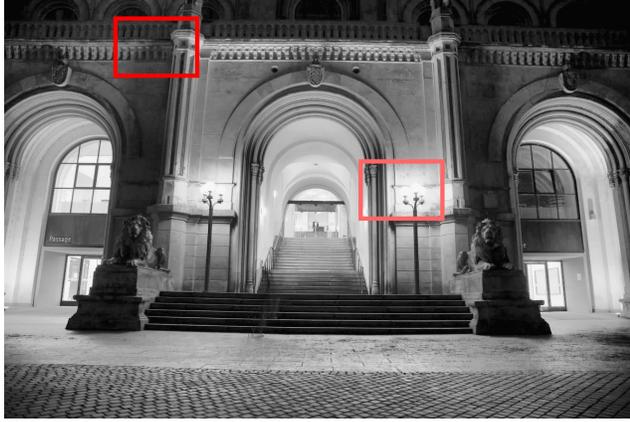
We would like to emphasize that the modeled sampling errors are not particular to our algorithm but apply for all HDR algorithms relying on several images independently of how the single images were acquired.

IV. EXPERIMENTAL RESULTS

We present simulated results for synthetic images in Section IV-A and present real-world photos in Section IV-B.

A. Simulation

In this section, we show the precision with which the proposed models reconstruct pixel values outside the valid dynamic range. Three different models are compared: First, a naïve extrapolation by Eq. (1) only using a single under-exposed image. This model is most similar to what photographers do without using special HDR tools. Obviously, any noise present in the used image will be extrapolated as well. Second, the model based upon the expectation using a uniform distribution in Eq. (5). It is intended to reduce noise not caused by the quantization. Third, the model based upon expectations using non-uniform distributions in Eqs. (5) and (6).



(a) HDR with *MatLab HDRs toolbox*



(b) Proposed algorithm



(c)



(d)



(e)



(f)

Figure 3: (Figure best viewed in color.) Processed images: (a) Result of HDR processing using the *HDR toolbox* of *MatLab 2016b* and magnifications in (c) and (e); (b),(d),(f) proposed algorithm with temporal averaging and weighting (Eqs. (2) and (6)). Higher-frequency details are lost by the HDR composition of *MatLab*, whereas they are preserved by the proposed approach both in dark and bright areas.

Artificial images are created by randomly choosing a number in $[0, 1]$, scaling it to $255 \cdot d_R$, before rounding to the nearest integer number. Here, $d_R > 1$ denotes the factor with which the image is supposed to exceed the dynamic range of the sensor. Thus, setting $d_R = 2, 4, 8, \dots$ increases the sensor resolution by 1, 2, 3, ... bits. All pixels beyond the physical limit of $v_{\max} = 255$ were estimated using the proposed models. For the first model we used $s = 10$ and $k = 1$, for the other two $s = 10$ and $k = 1, \dots, s - 1$. Normally distributed noise is added to each of the intermediate images with mean zero and standard deviations σ_k in percent of the maximal pixel value, $\sigma_k = 0.01 \cdot 255 \cdot d_R \frac{1}{s} \sigma$. The estimation error is given by the root mean square error (RMSE) between the estimated and the true values. The experiment was repeated 1000 times with random

images and increasing dynamic range $d_R = \{ 1.0, 1.25, \dots, 8 \}$.

The plot in Fig. 2 shows the results (*figure in color*). The horizontal axis corresponds to the dynamic range factor d_R , the vertical axis indicates the RMSE. The black lines indicate the results of a noise-free experiment with $\sigma = 0$, while the red lines indicate an experiment with noise $\sigma = 3$. The dash-dotted lines indicate the naïve extrapolation by Eq. (1), the dashed lines the reconstruction by Eq. (5) and uniform probability distributions, and the solid lines by non-uniform probability distributions using Eqs. (5) and (6).

It can be seen that the errors grow while the dynamic range factor increases for all models. This is not surprising since increasing the extrapolation also enhances both quantization errors and noise. The two models based upon the expectation

perform much better than the naïve approach. The model using non-uniform probability distributions consistently outperforms the others. For conservative values of d_R , the reconstructions of the latter two models improve the RMSE by more than a factor of 2 compared to the first model.

B. Real World Proof-of-Concept

We captured photos of the main entrance of a historical building at night-time with a Canon EOS 5D MARK III DSLR and simulated intermediate read-outs by taking several photos from a tripod with manually set exposure times. Figure 3a shows the HDR processed image using the *HDR toolbox* of *MatLab 2016b* (functions “makehdr” and “tonemap”). We used MatLab’s automatic processing using default settings. The result of the proposed algorithm using Eqs. (5) and (6) is shown in Figure 3b. The colored boxes in Figs. 3c–3f show magnifications of the areas in Figs. 3a and 3b, respectively.

It can be seen that high-frequency information is lost if the baseline algorithm is used. In areas with large luminance, for instance near the lamps, artifacts occur (cf. lower image in Fig. 3a) whereas the proposed algorithms is not susceptible to this error. Furthermore, by limiting the total timespan for image acquisition, i. e. only one image has to be taken with exposure time t_E for our approach in contrast to taking several images, our method is less prone to ghosting artifacts.

V. CONCLUSIONS

This paper is concerned with the acquisition of high dynamic range images using a conventional, low dynamic range camera sensor. Here, we claim that taking a single image is sufficient for HDR image creation if it is the image with longest exposure time and the camera allows to read intermediate pixel values *during* exposure. Thus, carefully determining multiple exposure times is not necessary anymore. This not only eliminates a possible cause of errors but also reduces the overall time of exposure, thereby reducing the requirement for the scene to be static. An error model is proposed to assess the effects of sampling error and noise. Since most or all HDR algorithms are based upon some form of extrapolation, these two types of errors are not a feature of the proposed algorithm but common to other algorithms as well.

Since current consumer cameras do not permit read-outs during exposure without reset, we introduced an approximation to intermediate read-outs. A quantitative evaluation shows that the proposed solutions indeed reconstruct HDR images while the errors are relatively low. By a qualitative comparison with the *HDR toolbox* from *Matlab 2016b* as a representative of state-of-the-art algorithms, it was demonstrated that the created HDR image has better quality and that higher-frequency details are preserved.

REFERENCES

[1] Greg Ward (Anywhere Software), “High Dynamic Range Image Encodings,” 2016.

[2] ITU-R, *Recommendation ITU-R BT.2100-1: Image parameter values for high dynamic range television for use in production and international programme exchange*, International Telecommunication Union, Geneva, Switzerland, June 2017.

[3] Karol Myszkowski, Rafal Mantiuk, and Grzegorz Krawczyk, *High Dynamic Range Video*, Morgan & Claypool, 2008.

[4] Erik Reinhard, Guiseppe Valenzise, and Frederic Dufaux, “Tutorial on *High Dynamic Range Video*,” in *23rd IEEE Int. Conf. on Image Processing (ICIP)*, Sept. 2016.

[5] P.A. Levine and R. Zhu, “High dynamic range CMOS pixel and method of operating same,” July 1 2014, US Patent 8,766,157.

[6] A. Fenigstein, R. Reshef, S. Alfassi, and G. Yehudian, “Single-exposure high dynamic range CMOS image sensor pixel with internal charge amplifier,” Aug. 11 2015, US Patent 9,106,851.

[7] J.N. Border and E.O. Morales, “Producing an extended dynamic range digital image,” July 13 2010, US Patent 7,756,330.

[8] S. Shen, “Method and apparatus for verifying CGA signature,” Nov. 19 2013, US Patent 8,589,692.

[9] M. Farrier, “CMOS active pixel sensor with improved dynamic range and method of operation,” Apr. 14 2009, US Patent 7,518,645.

[10] 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, Dec 2005.

[11] S. K. Nayar and V. Branzoi, “Adaptive dynamic range imaging: optical control of pixel exposures over space and time,” in *Proc. of 9th IEEE Int. Conf. on Computer Vision (ICCV)*, Oct 2003, pp. 1168–1175 vol.2.

[12] Michael D. Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen, “A Versatile HDR Video Production System,” *ACM Trans. Graph.*, vol. 30, no. 4, pp. 41:1–41:10, July 2011.

[13] M. A. Robertson, S. Borman, and R. L. Stevenson, “Estimation-theoretic approach to dynamic range enhancement using multiple exposures,” *Journal of Electronic Imaging*, vol. 12, pp. 219–228, Apr. 2003.

[14] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman, “Robust Patch-Based HDR Reconstruction of Dynamic Scenes,” *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203, 2012.

[15] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen, “Patch-based High Dynamic Range Video,” *ACM Trans. on Graphics (TOG) (Proc. of SIGGRAPH Asia 2013)*, vol. 32, no. 6, 2013.

[16] T. Jinno and M. Okuda, “Multiple exposure fusion for high dynamic range image acquisition,” *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 358–365, Jan 2012.

[17] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch, “Optimal HDR reconstruction with linear digital cameras,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 215–222.

[18] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy, “Burst photography for high dynamic range and low-light imaging on mobile cameras,” *ACM Trans. Graph.*, vol. 35, no. 6, pp. 192:1–192:12, Nov. 2016.

[19] Ana Serrano, Felix Heide, Diego Gutierrez, Gordon Wetzstein, and Belen Masia, “Convolutional sparse coding for high dynamic range imaging,” in *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics*, Goslar Germany, Germany, 2016, EG ’16, pp. 153–163, Eurographics Association.

[20] I. Choi, S. H. Baek, and M. H. Kim, “Reconstructing interlaced high-dynamic-range video using joint learning,” *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5353–5366, Nov 2017.

[21] Ahmet Oğuz Akyüz and Erik Reinhard, “Noise reduction in high dynamic range imaging,” *Journal of Visual Communication and Image Representation*, vol. 18, no. 5, pp. 366 – 376, 2007, Special issue on High Dynamic Range Imaging.

[22] Rastislav Lukac, *Single-Sensor Imaging: Methods and Applications for Digital Cameras*, CRC Press, Inc., Boca Raton, FL, USA, 1 edition, 2008.

[23] European Machine Vision Association (EMVA), *EMVA Standard 1288 Standard for Characterization of Image Sensors and Cameras*, 3.0 edition, Nov 2010.