

Metering for Exposure Stacks

O.Gallo¹, M. Tico², R. Manduchi³, N. Gelfand², and K. Pulli¹

¹ NVIDIA Research ² Nokia Research Center, Palo Alto ³ University of California, Santa Cruz



Figure 1: Uniformly sampling the space of exposure times until every pixel is correctly recorded at least once (i.e., it is not always clipped) can result in an unnecessarily large image stack with sub-optimal Signal-to-Noise ratio. For the scene shown in the tonemapped image on the left, this results in a 5-image stack. Our method determines that for this scene only 3 images suffice to capture the whole range while sampling the important intensity levels better. The insets for the 5-image stack (top row) and for our method (bottom row) are not tonemapped; rather they are linearly mapped to fit in 8 bits so as to preserve the noise characteristics. Notice the huge improvement in terms of noise with the smaller set of images selected by our method.

Abstract

When creating a High-Dynamic-Range (HDR) image from a sequence of differently exposed Low-Dynamic-Range (LDR) images, the set of LDR images is usually generated by sampling the space of exposure times with a geometric progression and without explicitly accounting for the distribution of irradiance values of the scene. We argue that this choice can produce sub-optimal results both in terms of the number of acquired pictures and the quality of the resulting HDR image.

This paper presents a method to estimate the full irradiance histogram of a scene, and a strategy to select the set of exposures that need to be acquired. Our selection usually requires a smaller or equal set of LDRs, yet produces higher quality HDR images.

Categories and Subject Descriptors (according to ACM CCS): HDR imaging, metering, exposure selection.

1. Introduction

Many real-world scenes cannot be captured in a single photograph as their dynamic range exceeds what the camera sensor can handle at once. While several hardware techniques have been developed for photographing high-dynamic-range (HDR) scenes [RHP*10], as of today, the most widespread approach still consists in taking a stack of LDR shots with different exposure time settings, thereby capturing different segments of the dynamic range of the scene [MP95, DM97].

Many of the HDR approaches first either determine the camera response function [MN99, GN03, RBS03] to linearize the input images into low-dynamic-range (LDR) ir-

radiance maps, or capture the data directly in linear RAW images, which are then combined into an HDR irradiance map [DM97, AR07, GAW*10], and finally tonemapped for display on regular monitors [DD02, FLW02].

Related work. The selection of exactly which LDR images should be captured to optimally sample the irradiance of a specific scene, has attracted little research attention. Taking many images can help to reduce the noise in the final irradiance estimate [BDH08], but requires more time. With a faster capture, the likelihood of artifacts in the presence of any scene or camera motion increases. Although various approaches [KAR06, GGC*09, JLW08] have been proposed to

deal with motion artifacts, the resulting images are often of lower quality, and the range they capture is often somewhat compromised.

The simplest metering approach is to use the built-in, single shot metering, and bracket the exposures up and down from that central exposure. A better solution is to estimate the scene irradiance range by finding its brightest and darkest points. Barakat et al. [BHD08] propose a capturing strategy that finds the minimum number of LDR images needed to capture an HDR scene assuming a linear camera response function, and uniform irradiance distribution. They find the minimum and maximum irradiance levels by iteratively changing the exposure time until no pixels are clipped, and subdivide the estimated dynamic range into shots. Assuming that the scene irradiance is uniformly distributed, they prove the optimality of their selection for a linear sensor. However, that assumption is often violated. Other methods take a similar route in that they simply estimate the extent of the range [PPL*08, BBKVv108]. Hasinoff et al. [HDF10] aim to minimize the acquisition time while maximizing the SNR, by favoring increasing the ISO sensitivity over the exposure time; however, they still do not explicitly account for the irradiance distribution of the scene. Piao and Xu [PX10] first find the median intensity of the scene and center the first shot there, then start working up and down until all pixels fall at least in one image in the interval [85, 170].

A particularly important reason to select the correct exposures, given a limited budget of many exposures, is the noise characteristics of the images. Both Chen and El Gamal [CEG02] and Granados et al. [GAW*10] adapt to the scene irradiance distribution, and propose an elegant framework to select the optimal exposures that maximize SNR of the final image captured with a linear sensor. However, the applicability of both approaches is somewhat limited by the assumption that the HDR histogram is given. Also, Chen and El Gamal propose to approximate it with a piece-wise uniform function—although no algorithm is provided to perform this approximation—and assume that the number of exposures is chosen *a priori*, which may lead to using more exposures than needed for a particular scene or not capturing all of its range.

The best existing metering approaches all assume a linear sensor, and are nontrivial to extend to a traditional non-linear sensor as they assume the quantization noise to be uniform over the captured range. Our first contribution is to provide an estimate of an HDR irradiance histogram, which can be used by the previous methods that assume such histogram be given. Our second contribution is a metering method that maximizes the peak signal-to-noise ratio (PSNR) of the captured image data for a non-linear camera.

The importance of non-linear cameras. Real cameras are non-linear for several reasons. The largest source of non-linearity is the processing done by the image signal processor (ISP), which mimics the behavior of the film and produces non-linear, gamma-corrected LDR images suitable for

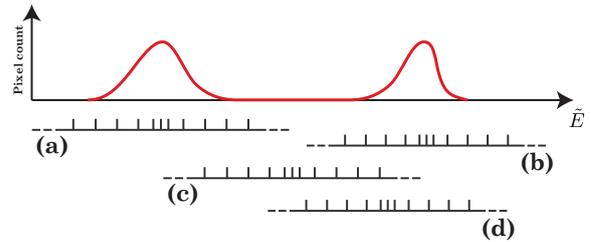


Figure 2: Different exposure times correspond, in the log-irradiance domain, to different shifts of the quantization bins of a picture, as shown underneath the plot. Covering every segment of the range might be unnecessary (image (c) does not capture any useful information). Image (b), which “wastes” a couple of bins, is arguably better than (d) because it uses the finer quantization intervals for values of log-irradiance that are more common in the scene. We claim that the optimal selection in this case should be (a) and (b).

viewing from traditional displays. However, even the values read directly from the sensor are not exactly linear. The optics is one source of nonlinearity, and the sensor’s material and electrical characteristics is another. The ISP algorithms and parameters are highly tuned to the particular sensor and optical characteristics of the system, and include corrections for non-uniform pixel response, lens shading, and dead pixels, and combined demosaicking and noise suppression. Short-circuiting this path and using the unprocessed RAW forfeits the benefits of this tuning.

Another reason to use the non-linear LDR camera output is to avoid the circuitous route of first estimating the scene HDR irradiance and then having to tonemap the results for display [MKR07, ZC10, WDLL11, SCSB11]. The ISP processes each image in the exposure stack as well as it can, using all its image processing algorithms, producing an LDR image. The images are analyzed for qualities such as contrast and saturation, and the best parts of each image are fused into a composite LDR image, directly ready for display. With fewer processing steps, the results can be computed more quickly. This approach works *with* the system software and hardware of the digital cameras rather than *against* it.

Finally, many popular cameras do not provide access to the RAW images. An increasing number of images are captured on camera phones and immediately shared on services such as Flickr and Facebook. This suggests that post-processing of the images should happen directly on the device the images are captured with. However, hardly any smart phone SDK provides access to the RAW images. Additionally, working on 8-bit LDR data instead of 16-bit HDR data requires less RAM, a scarce resource on mobile phones; a 3-image stack of 8 MP RGB images requires 72 MB of RAM in LDR, twice that in HDR.

Our contributions. In this work, we aim to choose the exposure times that maximize the number of pixels captured

within the linear segment of the camera response function, while minimizing the effect of noise (both signal dependent and signal independent noises); in a situation such as that depicted in Figure 2, our method allows capturing fewer pictures than standard auto-exposure-bracketing methods by omitting exposures in the flat area between the two lobes. This yields to higher quality input data, faster capture, and faster processing time. Our method comprises two steps: we first estimate the full irradiance histogram of the scene, which we then use to find the optimal set of exposures. We implemented the proposed method in MATLAB to perform a thorough analysis of each step, and on a Nokia N900 smartphone to test the feasibility of running it on a mobile device. Also, we show that the framework we propose, while designed to address the non-linear case, can be easily adapted to the linear case.

2. Histogram Estimation

Cameras can compute histograms quickly, depending on the architecture even before the image is ready; moreover, histograms are robust to slight scene changes and, to some extent, to changes in image size. The latter consideration allows us to use a low-resolution image of the scene—such as a viewfinder frame—to compute the histograms, should they not be computed independently at sensor readout.

Instead of regular metering for a single image, when the shutter button is pressed half-way our algorithm starts gathering a small number of LDR histograms that cover the whole irradiance range. We first describe our image formation model, and then how the LDR histograms are combined to an HDR histogram.

2.1. Image Formation Model

The scene irradiance E , the power per unit area falling on the sensor, is integrated over the exposure time t and converted to a digital value Z via the reciprocity equation $Z = g(E \cdot t)$. Here g is the camera response function, which incorporates all the linear and non-linear operations performed by the camera [DM97]. In log-domain, we can write:

$$\tilde{E} = \ln(g^{-1}(Z)) - \ln(t), \quad (1)$$

where \tilde{E} is the log-irradiance. To simplify the notation we denote by f the camera transfer function in the log-domain, i.e., $f^{-1}(Z) = \ln(g^{-1}(Z))$.

The operation of taking a picture corresponds to quantizing the irradiance with 2^B levels, where B is the number of available bits. Equation 1 shows that changing t corresponds to shifting the position of the quantization bins, as depicted in Figure 2, where exposure (a) is longer than (b). Also note that the last bin captures all the log-irradiance values up to $+\infty$ and the first bin down to $-\infty$, hence the dotted lines. In the rest of the paper we will refer to the domain of Z , i.e., the set $[0, 255]$, as the “digital” or “compressed” domain in contrast with the irradiance domain.

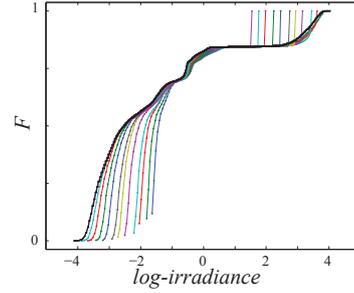


Figure 3: F^H , the HDR CDF for the scene (in black in this figure), can be estimated by taking the maximum of multiple LDR CDFs, F_j^L (the remaining plots in the figure). The noise strongly affects the lower segment of the range: ideally, all F_j^L should align perfectly in the areas where they are defined. Also, notice that the last bin of each of the F_j^L is affected by saturation and is therefore ignored.

2.2. From LDR Histograms to HDR Histogram

When trying to build the histogram of the irradiance values in the scene, we have two main objectives: we need it to reflect, as accurately as possible, the whole range of the irradiance values, which is in general larger than what a single picture can capture, and we need the estimation process to be reasonably fast. Camera APIs often allow access to the histograms that the sensors generate at the readout stage, relieving the application from computing them from the images themselves [ATP*10]. Since these histograms are virtually for free, a possible strategy is to capture them while changing exposure times, iterating until we reach both ends of the range, in other words, when the first bin of the histogram of the longest exposure, and the last bin of the shortest exposure are empty.

The combination of the LDR histograms, however, is hardly straightforward. Recall that changing the exposure time corresponds to shifting the quantization boundaries $\{b_i\}_{i=0}^I$ in the log-domain:

$$b_{i,j}^{\tilde{E}} = f^{-1}(b_i^Z) - \ln(t_j), \quad (2)$$

where $b_{i,j}^{\tilde{E}}$ and b_i^Z are the i^{th} quantization boundaries in the log-irradiance and digital value domains, respectively, and t_j is the j^{th} exposure time. (Note that, in the linear irradiance domain, the same operation corresponds to stretching the bins as the exposure time changes.) This results in the quantization boundaries being, in general, misaligned for different exposure times, causing the inconvenient need for re-binning. If a bin in an exposure spans multiple bins in another, the samples falling in the former have to be redistributed among the multiple bins of the latter. The high-rate assumption, which states that the samples are uniformly distributed over a bin, is often violated, in which case, splitting the samples uniformly between the bins in the second exposure produces artifacts in the histogram.

On the other hand, the Cumulative Distribution Function (CDF) domain seems the natural choice for this combination. In this domain, the LDR histograms defined on the digital domain can be merged by simply computing, for each of the J exposures, the LDR CDFs; these can be then converted to irradiance CDFs, F_j^L , and the HDR CDF can be built by taking the maximum value at each location, see Figure 3. Contrary to the case of histograms, taking the envelope of the LDR CDFs requires no explicit re-binning, which will simply result from taking the derivative of the HDR CDF. Formally, using Eq. 2, we can compute the set of quantization boundaries which forms the support, or baseline, of each log-irradiance histogram; note that these are also the baselines for the corresponding F_j^L . Their union forms the baseline of the full cumulative distribution function, F^H :

$$\{B_k\} = \bigcup_j \{b_{i,j}^{\bar{E}}\}. \quad (3)$$

With this baseline we can use Algorithm 2.1, where K is the number of bins in F^H , J is the number of exposures used, and I is the number of bins in F_j^L .

Algorithm 2.1: COMPFULLCDF($\{B_k\}, \{b_{i,j}^{\bar{E}}\}, F_1^L, F_2^L, \dots, F_J^L$)

```

for  $k \leftarrow 0$  to  $K - 1$ 
  do  $F^H(B_k) \leftarrow 0$ 

for  $j \leftarrow 0$  to  $J - 1$ 
  do  $\left\{ \begin{array}{l} \text{for } i \leftarrow 0 \text{ to } I - 2 \\ \text{do } \left\{ \begin{array}{l} \text{for each } k : B_k \in (b_{i,j}^{\bar{E}}, b_{i+1,j}^{\bar{E}}] \\ \text{do } F^H(B_k) \leftarrow \max(F^H(B_k), F_j^L(b_{i,j}^{\bar{E}})) \end{array} \right. \end{array} \right.$ 

return ( $F^H$ )
  
```

As is apparent from Figure 3, the first bins of the CDFs provide an inaccurate reading of the irradiance values. Because of noise, some of the pixels that should fall in the first bins, i.e., whose digital value should be zero, may spread to the neighboring bins; for this reason, the first bins of the CDF in general start from a smaller value than they should, see Figure 3. To overcome this bias, instead of averaging the functions, we retain the envelope, that is, at each location B_k , we retain the maximum value of the CDFs that are defined in B_k . The same problem affects the last bins of F_j^L , however, the noise contribution in the upper segment of the range is drastically smaller, and the dominant factor becomes saturation; this is why Algorithm 2.1 disregards the last bin of each F_j^L . One last comment pertains to the number of LDR histograms that we need for our estimate: during the metering stage, the exposure times can be selected to ensure a minimal overlap of the bins of different histograms. Because of this, the number of collected histograms is limited.

It is worth noticing that, by construction, F^H is a *proper* cumulative distribution function; it is the envelope of monotonically non-decreasing functions so it is itself mono-

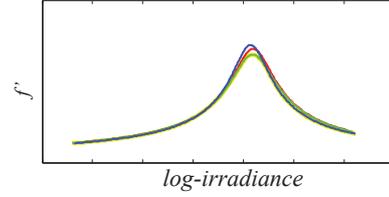


Figure 4: The derivative of the camera response function for the Canon XSi, here shown for $t = 1s$, for the three channels.

tonically non-decreasing. Also $\lim_{x \rightarrow -\infty} F^H(x) = 0$ and $\lim_{x \rightarrow \infty} F^H(x) = 1$. Additionally, it is right-continuous, again, by construction.

Finally, the HDR histogram of the full scene can be computed simply by differentiating F^H , using $\{B_k\}$ to account for the non-homogeneous bin size.

3. Exposure Selection

The knowledge of the full histogram of the scene irradiance allows us to make an educated choice for the number of pictures that are necessary and which exposure times should be used. On one hand we want to minimize the number of pictures that need to be taken, on the other we want to optimize the quality of the final result. Consider the simple case of a linear camera response function and an example such as that depicted in Figure 2; because the distribution of the irradiance is roughly bimodal, we might decide to avoid using an unnecessary extra picture that covers the middle of the range.

The problem is more complicated for non-linear camera response functions. The non-linearity of the camera response function f makes the size of the quantization bins vary across the range. Usually, f has a sigmoid-like shape which is roughly linear at the center of the range and which performs some compression towards the ends. This can be seen in terms of resolution with which the irradiance values are quantized, as shown in Figure 4. This means that two different exposure times, both of which cover the whole range, might not be equivalent in terms of quantization error: a method which seeks to simply “tile” exposures to cover the range may produce sub-optimal results in terms of SNR.

Quantization, however, is not the only source of noise involved in this process. To account for this we use the model proposed by Foi et al. [FTKE08] which comprises a Poissonian component, accounting for Photon Shot Noise (PSN), and a Gaussian component, capturing other disturbances, such as readout and thermal noise.

3.1. Maximizing the Peak SNR for a Single Picture

The quantized digital value Z recorded by a camera system at a given pixel is given by $Z = Q(g(X))$, where Q is the quantization operator, and X denotes the exposure value at the given pixel. The exposure $X = Et$ is a measure of the scene irradiance E integrated over the exposure time t .

The range $[X_{min}, X_{max}]$ of exposures a sensor can reliably capture is limited on one side by the dark noise level, and on the other side by the full well capacity of the pixels. Within this range the pixels are affected by noise that includes signal-dependent components (photon-shot noise), and signal-independent sources (readout noise, thermal noise). Following the noise model by Foi et al. [FTKE08], we can write the camera noise variance before gamma correction as

$$PG(X) = a \frac{X - X_{min}}{X_{max} - X_{min}} + b, \quad (4)$$

where the parameters a and b are camera-dependent parameters that can be estimated from one or more images captured with the same camera. The two parameters depend of the analog gain; in our experiments we used ISO100.

In addition, in the non-linear case, the shape of the camera response function determines different quantization errors in different regions of the exposure range $[X_{min}, X_{max}]$. These errors depend on the first derivative of the camera response function and, in general, they are smaller in the middle of the range and larger towards the ends.

To simplify the derivations, we formulate the proof of the optimal exposure selection in the logarithmic domain. We denote the variables in the logarithmic domain with a tilde above the variable name, e.g., $\tilde{X} = \ln X$.

The digital level detected at a given pixel can be reformulated as

$$Z = Q(f(\tilde{X})), \quad (5)$$

where $f(\tilde{X}) = g(\exp(\tilde{X}))$ stands for the camera response function in the logarithmic domain. The noise model (4) in the logarithmic domain is

$$PG(\tilde{X}) = a' \exp(\tilde{X}) + b', \quad (6)$$

where $a' = a/(X_{max} - X_{min})$ and $b' = b - a'X_{min}$. To address the case of non-linear camera response functions, we have to include an explicit description of the non-linear quantization noise. Recall from Equation 5 that Z is quantized uniformly, and let us call Δ the size of the quantization steps. Because f induces a non-linear relationship between Z and \tilde{X} , said quantization steps correspond to non-uniform steps in the domain of \tilde{X} . The size of the i^{th} quantization interval, centered at $\tilde{X}_i = f^{-1}(Z_i)$, can be approximated assuming the camera response function locally linear, by

$$\Delta_i = \Delta / f'(\tilde{X}_i), \quad (7)$$

where f' is the first derivative of the camera response function in the log domain. For a signal normalized between 0 and 1, $\Delta = 2^{-B}$, where B is the number bits used to represent the digital value Z .

The quantization error is a linear function inside each quantization interval, i.e., $e(\tilde{X}) = \tilde{X} - \tilde{X}_i$ for any $\tilde{X} \in [\tilde{X}_i - \Delta_i/2, \tilde{X}_i + \Delta_i/2]$, and assuming that it is uniformly dis-

tributed inside the interval we have the quantization noise variance inside the i^{th} quantization interval [Say00]

$$\sigma_Q^2(\tilde{X}_i) = \frac{1}{\Delta_i} \int_{\tilde{X}_i - \Delta_i/2}^{\tilde{X}_i + \Delta_i/2} e^2(\tilde{X}) de = \frac{\Delta_i^2}{12} = \frac{1}{12} \left[\frac{\Delta}{f'(\tilde{X}_i)} \right]^2 \quad (8)$$

where we used the result in Equation 7. Variances of independent variables can simply be added up, even if they arise from different distributions. Therefore, inside the i^{th} quantization interval the noise variance can be formulated as

$$\sigma_i^2 = PG(\tilde{X}_i) + \sigma_Q^2(\tilde{X}_i). \quad (9)$$

The peak signal to noise ratio (PSNR) inside the i^{th} quantization bin is thereby

$$PSNR(\tilde{X}_i) = \frac{1}{a' \exp(\tilde{X}_i) + b' + \frac{1}{12} \left[\frac{\Delta}{f'(\tilde{X}_i)} \right]^2} \quad (10)$$

$$= \frac{[f'(\tilde{X}_i)]^2}{[f'(\tilde{X}_i)]^2 [a' e^{\tilde{X}_i} + b'] + \frac{\Delta^2}{12}}. \quad (11)$$

Denoting the i^{th} bin of the empirical distribution (histogram) of the scene log-irradiance \tilde{E} by $P_{\tilde{E}}(\tilde{E}_i)$, we can formulate the total PSNR corresponding to a given exposure time t as

$$PSNR(\tilde{t}) = \sum_i P_{\tilde{E}}(\tilde{E}_i) PSNR(\tilde{E}_i + \tilde{t}) \quad (12)$$

$$= \sum_i \frac{P_{\tilde{E}}(\tilde{E}_i) [f'(\tilde{E}_i + \tilde{t})]^2}{[f'(\tilde{E}_i + \tilde{t})]^2 [a' e^{\tilde{E}_i + \tilde{t}} + b'] + \frac{\Delta^2}{12}}.$$

Because we formulated the problem in the log-domain, where changing \tilde{t} corresponds to shifting Equation 10, Equation 12 is a simple correlation, and the optimal exposure time \tilde{t}_0 corresponds to its maximum:

$$\tilde{t}_0 = \underset{\tilde{t}}{\operatorname{argmax}} PSNR(\tilde{t}). \quad (13)$$

3.2. Selection of the Set of Optimal Exposures

When optimizing for the set of exposures, we want to get a reasonable computational load so the method can run quickly on a mobile device. A greedy method would produce that, but it might get stuck to a local minimum. We define the quality of a set of N exposures as:

$$O(\{\tilde{t}_i\}_{i=1}^N) = \sum_{i=1}^N PSNR(\tilde{t}_i). \quad (14)$$

Recall from Equation 12 that, for each exposure, we want to maximize the correlation between a mask and the irradiance histogram of the scene. This correlation needs to be evaluated only for the exposure times t that a specific camera offers. For example, the camera we used for this paper, the Canon XSi, only allows for 52 settings, 1/3 of a stop apart from each other. Additionally, the number of pictures required for a specific scene is limited—we had a hard time finding a scene requiring more than 3 LDR images, aside

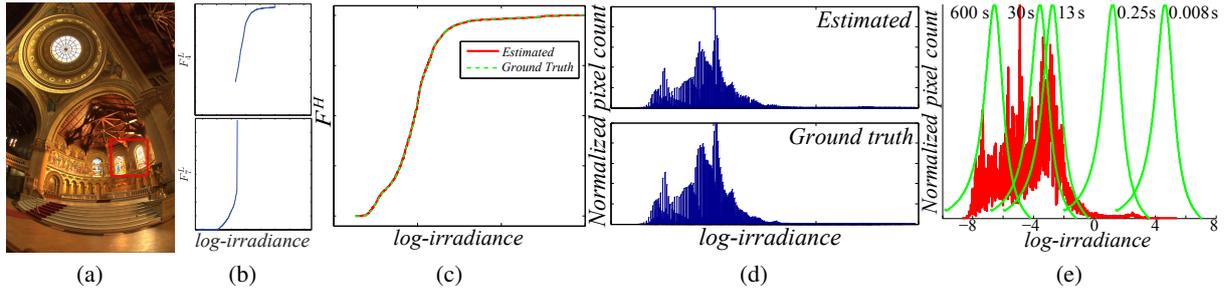


Figure 5: (a) The original HDR image, (b) the cumulative distribution functions of two of the synthetic LDRs generated from the HDR image, (c) the HDR cumulative distribution function, (d) comparison of the HDR histograms, and (e) our algorithm's selection of the exposures. The red square in (a) indicates the location of the details shown in Figure 7.

from the example in Figure 5. These observations suggest that it is feasible to perform an exhaustive search.

For this purpose, we generate the list of all possible combinations of N exposures off-line. When the histogram becomes available, we compute its correlation with the mask in Equation 10, which, as we noticed, consists of only a few dozen points. To further reduce the computational load, we make two observations; a satisfactory HDR image cannot have saturated or under-exposed pixels, and it is useless to take an exposure that has a minimal (or no) overlap with the histogram. Therefore we only compute the quality measure of Equation 14 for the combinations of exposures that correctly capture with at least one exposure the first and last bin of the histogram and for which all the exposures overlap with the scene's range.

While we do not want to prevent the algorithm from choosing very similar exposures (which might be at times beneficial to better attenuate the noise) we do want to encourage a choice of exposures that “covers” as much of the range as possible. For this reason, if exposure i captures a certain part of the range, we attenuate the corresponding samples of the correlation function by an amount that is roughly proportional to how well they were captured. Since we do not set the value of these samples to zero, we allow the algorithm to choose two exposures that are close to each other (as in the example in Figure 5) but, since these samples are indeed attenuated, other parts of the range will provide a stronger contribution to the quality measure in Equation 14, and will therefore be encouraged.

For a given N , the optimal set $\{\tilde{t}_i\}_{i=1}^N$ is the one that maximizes Equation 14. To select the number of exposures, we start with $N = 1$, and we stop when the derivative of $\max(O^N)$ as a function of N drops, the rationale being that the cost associated with taking an additional picture is not justified by the limited increase of the predicted quality of the stack.

For the sake of simplicity, until now we have assumed to have an estimate of P_E , while really we have $[P_{E_r}, P_{E_g}, P_{E_b}]$ and $[f_r, f_g, f_b]$. Obviously, we cannot run our algorithm on the three channels separately as that might result in different

optimal exposure times for the different channels. Therefore, we define:

$$P_E = \text{mean}(P_{E_r}, P_{E_g}, P_{E_b}) \text{ and } f' = \min(f'_r, f'_g, f'_b). \quad (15)$$

We average the histograms to avoid the exposure being selected mostly based on the more prominent channel in a given part of the range, and also because this gives a cheap approximation of the luminance histogram. For the derivative of the camera response function, which can be considered as the sampling resolution, we take the minimum of the three channels at each point because we want to perform the choice based on the resolution of the worst channel.

4. Results

In this section we present our results. All the HDR images shown in this paper are generated using the weights suggested by Debevec and Malik [DM97]. Since tonemapping algorithms might attenuate visual advantages and disadvantages of the different exposure selection strategies, we linearly map the irradiance to fit different segments of the range in 8 bits. However, to allow the reader to get an idea of the scene in question, we also present the HDR image generated with all the exposures available and tonemapped with the method proposed by Fattal et al. [FLW02], unless noted otherwise.

4.1. A Case Study: Synthetic Data

In order to perform a thorough analysis of the proposed method we need the ground truth of the scene irradiance. For this purpose we use the HDR image of the Stanford Memorial Church from Debevec and Malik [DM97]. From it, and using a generic camera response function, we generate 13 LDR (8-bit) histograms simulating exposure times that would cover the whole range, and add random noise generated with the model proposed by Foi et al. [FTKE08]. We then compute the histograms from these synthetic images. Also note that Debevec and Malik recovered the irradiance map up to a scale factor; while this does not affect our computation, it can yield seemingly unrealistic exposure

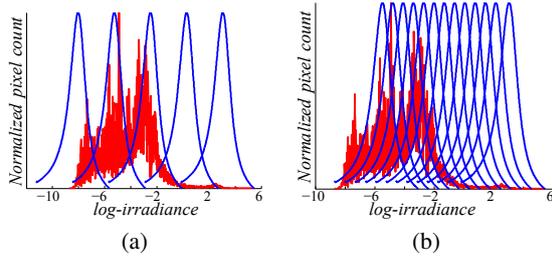


Figure 6: The distribution of the exposure times for the methods AEB_3 (a) and “cover the range” (b) for the example shown in Figure 5.

times. All the exposure times indicated in this section share the same scale factor.

Histogram Estimation. Figure 5 (b) shows two of the LDR CDFs, while Figure 5 (c) offers a visual comparison between estimated and computed HDR CDFs. The two distributions are extremely close. To better quantify how close, we also perform the Kolmogorov-Smirnov test (K-S test), which is a non-parametric statistical test that compares two empirical CDFs to decide whether the underlying samples can be considered different realizations of the same random variable. The K-S test is sensitive to both the local slope and to the location of the CDFs and is therefore a good measure for our purposes. We compare our estimate of F^H and the ground truth, i.e., the CDF computed directly from the HDR image: all the three channels pass the test with the standard significance level $\alpha = 0.05$.

Exposure Selection. Figure 5 (e) shows the exposures that the algorithm described in Section 3 selects, superimposed on the estimate of the histogram (times in seconds):

$$\{t_j\} = \{ 0.008, 0.25, 13, 30, 600 \}. \quad (16)$$

We compare our selection, both visually and in terms of SNR, with two different approaches. First, assuming no knowledge of the irradiance of the scene, one can follow the auto-exposure-bracketing (AEB) approach: meter as if a single picture is to be taken and then take a number of pictures a few stops apart. In this case we can choose 5 pictures, as many as our method selects. To compensate for the intrinsic disadvantage of this method, we test 3 different progressions, all starting from the middle of the range (10 seconds) but with increasing distance between the shots: 2 stops (AEB_1), 3 stops (AEB_2), and 4 stops (AEB_3). A second, more educated approach consists in finding the shortest and longest exposures required to avoid saturation, and then padding the range in between with uniformly spaced exposure times, say 1 stop apart from each other. This results in choosing (times in seconds):

$$\{t_j\} = \{ 200, 100, 50, 25, 12.5, 6.25, \\ 3.1250, 1.5625, 0.7813, \\ 0.3906, 0.1953, 0.0977, 0.04 \}. \quad (17)$$

Ours (5)	“Cover the range” (13)	AEB (5)
10.941 dB	3.144 dB	0.306 dB [AEB_1] 1.139 dB [AEB_2] 3.882 dB [AEB_3]

Table 1: Comparison of the SNR of the three approaches described in Section 4.1 (the number of images used by each method are in parentheses). Note that our choice provides a huge improvement using as many or fewer images.



Figure 7: Details from the Stanford Memorial Church example. From left to right: our selection of the exposures, the results of covering the entire range, and AEB_3 . With the exposure times selected using the three methods, we generated three different HDR images. The insets are mapped linearly to 8 bits. Note the artifacts for the standard approaches.

Note that the number of exposures in Eq. 17 is more than twice as large as our choice, and the exposures span the whole range (see Figure 6 (b)). Table 1 shows the SNR of the HDR image generated with the three methods (compared with the original HDR image) as computed by

$$\text{SNR(dB)} = 10 \cdot \log_{10} \left(\frac{\mu^2}{e^2} \right), \quad (18)$$

where μ^2 is average of the squared value of the ground truth (in this case the original irradiance map) over all the pixels, and e^2 is the mean squared error, the average of the squared difference between the original irradiance map and the reconstructed HDR image over all the pixels. Our method achieves an SNR at least an order of magnitude greater than the two alternative methods with at most as many images. This improvement can be understood with the example in Figure 2.

A visual comparison of a particularly interesting region for the three approaches is shown in Figure 7. Notice how the two alternative approaches are affected by quantization error. The artifacts visible in both Figure 7 and in the detail of the monitor of Figure 8 arise from the fact that the three channels are calibrated independently: two pixels that fall in the same bin in one channel might fall in different bins in another.

An additional observation about the above comparisons is in order; when the extreme exposures do not extend *beyond* the range of the scene’s irradiance, as mentioned before, the SNR of the final image might be impacted by the coarse resolution of the CRF in its top and bottom segments. For AEB_3

from Figure 6, for example, the longest exposure extends beyond the range, but the shortest exposure barely covers the brightest parts of the image; this is part of the reason for the poor performance of the progression. When pushing the shortest exposure to the right, for instance enforcing that its brightest pixel be at most 127, this combination achieves a higher SNR (6.144 dB) which is, however, still lower than our method. Similar considerations apply for the other examples. Moreover choosing a particular number, such as 127, is a somewhat arbitrary heuristic, which might be an overkill in some cases, requiring extremely long and short exposures (even beyond the limits of the camera capabilities, for some examples in this paper). On the contrary, part of the contribution of our method is precisely to decide how far to push the shortest and longest exposures for a specific irradiance distribution.

4.2. Real Data

To evaluate our results on real data, we gathered three stacks modifying only the exposure time; we took a picture for each of the 52 settings available on a Canon XSi in fully manual mode (excluding "bulb") and performed the computation off-line. Note that we used only as few LDR histograms as are needed to cover the entire range to generate the HDR histogram on a real camera (in these examples, we generated the LDR histograms from the images themselves).

The results of different exposure selection strategies are shown for visual inspection and comparison. For the examples in Figures 8 and 10 we also present a numerical evaluation based on the SNR from Equation 18, and the HDR image generated from all the images in the stack as the ground truth. An evaluation based on the SNR is only valid for static scenes such as these. Motion during the capturing process, such as the few branches blown by the wind during the capture of Figure 1 have a significant effect on the SNR, which is by nature sensitive to strong outliers. However, our method for choosing the best exposures to capture is robust against small scene changes as long as the overall histogram does not change drastically.

We claim that the strategy of evenly sampling the exposures between the detected minimum and maximum scene irradiance can result in unnecessarily many captured images. The estimated histogram for the scene in Figure 8 is shown below the image; in the histogram, some of the log-irradiance values within the full range are missing from the scene (flat segment between the two lobes). This is reflected by the choice of the necessary exposures performed by our algorithm: the three exposures are not uniformly spaced. The SNR achieved with this selection is 25.09 dB. A 5-image stack which covers the detected range uniformly (see Figure 8) achieves an SNR of only 14.93 dB; notice the dramatic amount of noise on the football for 5-image stack (middle column) compared to our 3-image selection (right column). Figure 9 shows the error of the two approaches for comparison.

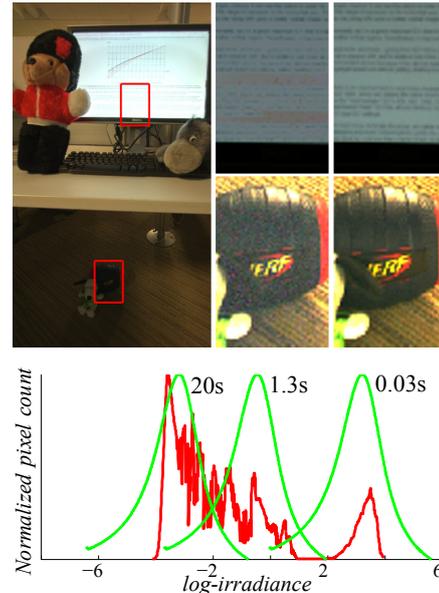


Figure 8: For this scene, the 5 images that span the whole range uniformly produce the results in the middle column. Our method selects only the 3 exposures shown in the histogram below, and produces the results on the right. Notice the increased quality of our result.



Figure 9: Distribution of the reconstruction error, divided by the value of the ground truth, for the scene in Figure 8. On the left the 5-image, uniformly-distributed selection that spans the whole range, on the right our result.

Figure 10 shows a scene for which our algorithm selects 3 exposures, achieving an SNR of 15.83 dB; a 5-image stack, required to uniformly span the range, achieves an SNR of 13.1 dB. Another result demonstrating the better quantization achieved by our method is shown in Figure 1.

4.3. The linear case

Unlike methods that only work in the linear domain, our approach can be used for extended-range methods and allows to exploit the camera-specific denoising, demosaicing, and compression algorithms. However, it can be applied to linear cameras as well, by just replacing the camera response

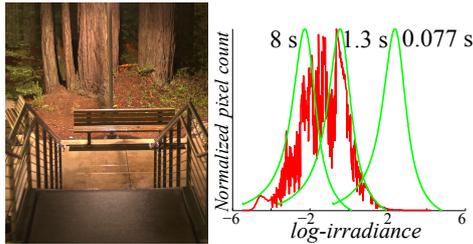


Figure 10: For the scene on the left, our algorithm selects three exposures, shown superimposed on the estimated histogram. The SNR achieved by our selection is 15.83dB, almost 3dB larger than that of a uniform selection which covers the whole range requiring five images.

function in Equation 10 accordingly. To validate this claim, we acquired a stack of linear pictures for the scene in Figure 11, and compared our algorithm with that of Granados et al. [GAW*10]. On this example, our algorithm selects two exposure times and achieves an SNR of 27 dB. We compare this result against the HDR generated with the first two exposures selected by the method by Granados et al.[†] Their code, run as-is, selects exposures that yield to an SNR of 8.21 dB; a careful investigation on the reason behind such a low SNR reveals that the color channels have very different ranges, while Granados et al. perform their selection only using the blue channel. (Results get even worse using the luminance instead). Since their paper does not mention the reason behind using the blue channel, we hand-picked the channel with the widest range, which, for this example, results in a selection that achieves essentially the same results as our method (27.3 dB).

A few remarks are in order. Because Granados et al. do not propose a method to estimate the histogram, we fed their code the irradiance map computed using all the images in the stack; our method, on the contrary, estimates the histogram as well. Moreover, selecting the channel with the widest range, while reasonable, is not part of their algorithm; we defined such strategy only to perform the most fair comparison. Finally, the comparison should be weighted by the fact that, contrary to Granados et al., the framework we propose works both in the linear and non-linear domain.

5. Implementation Considerations

We implemented the proposed method on a Nokia N900 smartphone using the FrankenCamera API [ATP*10]. This platform allows direct access to the histograms computed in hardware by the Image Signal Processor (ISP); these are 64-bin histograms computed directly in the linear domain.

[†] To solve for the scale ambiguity caused by working in the irradiance domain, we scaled the exposure times selected by Granados et al. using the fact that the first exposure selected is always *expose-to-the-right*.



Figure 11: For this scene, and performing the exposure selection in the linear domain, our algorithm selects the images on the left, which, combined in the image on the right, achieve an SNR of 27dB.

In our N900 implementation, starting from the exposure time of the current viewfinder frame, we keep collecting histograms until the whole range is covered, i.e., until every pixel is captured correctly in at least one exposure. To minimize the time required to gather the LDR histograms, we require that these be minimally overlapping, so as to minimize their number, and we encourage boosting the ISO—even to extreme values—as opposed to the exposure time (note that we do this *only* for the collection of the LDR histograms, not for capturing the actual pictures).

Regarding the actual selection of the optimal exposures, it is worth noticing that the total number of combinations to test for a specific scene is drastically reduced by the considerations in Section 3.2 (recall that we only test combinations containing sensible exposures). Moreover we found that most scenes required only two or three exposures, thus the algorithm never had to evaluate the combinations of four or five exposures. To get a feel for the order of magnitude of the execution time, we can report that the autofocus routine, which our application performs in parallel with the HDR metering, takes longer than our metering procedure. While the computational load of the whole algorithm is somewhat dependent on the radiance range of the scene, using the strategy described above we found that the only cases in which the metering caused a perceivable delay were scenes with regions requiring long exposure times (for the metering stage, as well as for the actual capture of the stack) even after boosting the ISO to its maximum value.

Figure 12 shows the images selected and captured by our method, and their combination performed with the method by Mertens et al. [MKR07]; the capturing time was below one second from the moment the shutter was pressed to the time the last picture was captured.

6. Conclusions

This paper proposes a novel strategy to determine the exposure times required to generate an HDR image; it is the first approach that explicitly and fully adapts this choice to the *specific* scene that is being captured and that works for



Figure 12: We implemented our method on a Nokia N900 smartphone. For this scene our algorithm selected two pictures approximately 1.4 stops apart (left). We also show the result of Exposure Fusion [MKR07] (right), demonstrating that all the range was correctly captured.

non-linear cameras. Our method comprises two steps. First, we define a strategy to estimate the HDR histogram of the scene. Using the histogram, we select exposure times that optimize the peak SNR of the resulting image. We show that our method outperforms the standard heuristics in terms of SNR and provides visual evidence of the attenuation of the noise across the range. To prove that our method is suitable for low-end cameras we also deployed our algorithm to a Nokia N900; the metering runs in parallel with the autofocus. While this paper focuses on the non-linear case, we show that our method compares favorably with the state-of-the-art for linear cameras as well.

References

- [AR07] AKYÜZ A. O., REINHARD E.: Noise reduction in high dynamic range imaging. *Journal of Visual Communication and Image Representation* 18, 5 (2007), 366–376. 1
- [ATP*10] ADAMS A., TALVALA E.-V., PARK S. H., JACOBS D. E., AJDIN B., GELFAND N., DOLSON J., VAQUERO D., BAEK J., TICO M., LENSCH H. P. A., MATUSIK W., PULLI K., HOROWITZ M., LEVOY M.: The Frankencamera: An experimental platform for computational photography. *ACM Trans. Graph.* 29 (July 2010), 29:1–29:12. 3, 9
- [BBKV108] BILCU R. C., BURIAN A., KNUUTILA A., VEHVÄ LÄINEN M.: High Dynamic Range Imaging on Mobile Devices. In *IEEE Int. Conf. Electronics, Circuits, and Systems* (2008), pp. 1312–1315. 2
- [BDH08] BARAKAT N., DARCIE T. E., HONE A. N.: The trade-off between SNR and exposure-set size in HDR imaging. In *IEEE Int. Conf. Image Processing* (2008), pp. 1848–1851. 1
- [BHD08] BARAKAT N., HONE A., DARCIE T.: Minimal-bracketing sets for high-dynamic-range image capture. *IEEE Trans. Image Processing* 17, 10 (2008), 1864–1875. 2
- [CEG02] CHEN T., EL GAMAL A.: Optimal scheduling of capture times in a multiple-capture imaging system. In *Proc. SPIE* (2002), vol. 4669, pp. 288–296. 2
- [DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.* 21 (July 2002), 257–266. 1
- [DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH* (1997), pp. 369–378. 1, 3, 6
- [FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Trans. Graph.* 21 (July 2002), 249–256. 1, 6
- [FTKE08] FOI A., TRIMECHE M., KATKOVNIK V., EGIAZARIAN K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans. Image Processing* 17 (2008), 1737–1754. 4, 5, 6
- [GAW*10] GRANADOS M., AJDIN B., WAND M., THEOBALT C., SEIDEL H., LENSCH H.: Optimal HDR reconstruction with linear digital cameras. In *CVPR* (2010), pp. 215–223. 1, 2, 9
- [GGC*09] GALLO O., GELFAND N., CHEN W., TICO M., PULLI K.: Artifact-free high dynamic range imaging. In *IEEE Int. Conf. Computational Photography* (2009), pp. 1–7. 1
- [GN03] GROSSBERG M., NAYAR S.: What is the Space of Camera Response Functions? In *CVPR* (2003), pp. 602–609. 1
- [HDF10] HASINOFF S. W., DURAND F., FREEMAN W. T.: Noise-optimal capture for high dynamic range photography. In *CVPR* (2010), pp. 553–560. 2
- [JLW08] JACOBS K., LOSCOS C., WARD G.: Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications* 28, 2 (2008), 84–93. 1
- [KAR06] KHAN E., AKYÜZ A., REINHARD E.: Ghost removal in high dynamic range images. In *IEEE Int. Conf. Image Processing* (Oct. 2006), pp. 2005–2008. 1
- [MKR07] MERTENS T., KAUTZ J., REETH F. V.: Exposure fusion. In *Pacific Conf. on Computer Graphics and Applications* (Jan 2007), pp. 382–390. 2, 9, 10
- [MN99] MITSUNAGA T., NAYAR S.: Radiometric Self Calibration. In *CVPR* (1999), vol. 1, pp. 374–380. 1
- [MP95] MANN S., PICARD R.: Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures. In *IS&T 46th Annual Conference* (1995), pp. 422–428. 1
- [PPL*08] PARK D., PARK K., LEE T., CHOI M., HA Y.: Efficient hdr image acquisition using estimation of scenic dynamic range in camera images with different exposures. In *Proc. SPIE* (2008), vol. 6807, pp. 1–12. 2
- [PX10] PIAO Y., XU W.: Method of auto multi-exposure for high dynamic range imaging. In *Int. Conf. Computer, Mechatronics, Control and Electronic Engineering* (2010), pp. 93–97. 2
- [RBS03] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging* 12, 2 (2003), 219–228. 1
- [RHP*10] REINHARD E., HEIDRICH W., PATTANAİK S., DEBEVEC P., WARD G., MYSZKOWSKI K.: *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, 2010. 1
- [Say00] SAYOOD K.: *Introduction to Data Compression*. Morgan Kaufmann, 2000. 5
- [SCSB11] SHEN R., CHENG I., SHI J., BASU A.: Generalized random walks for fusion of multi-exposure images. *IEEE Trans. Image Processing*, 99 (2011), 3634–3646. 2
- [WDLL11] WANG J., DE X., LANG C., LI B.: Exposure fusion based on shift-invariant discrete wavelet transform. *Journal of Information Science and Engineering* 27 (2011), 197–211. 2
- [ZC10] ZHANG W., CHAM W.-K.: Gradient-directed composition of multi-exposure images. In *CVPR* (2010), pp. 530–536. 2