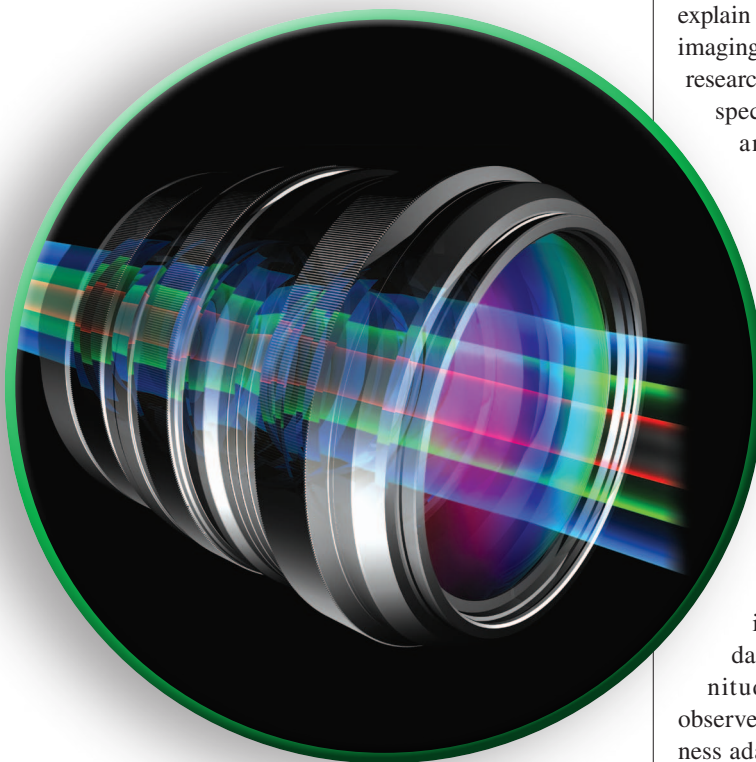


Practical High Dynamic Range Imaging of Everyday Scenes

Photographing the world as we see it with our own eyes



©ISTOCKPHOTO.COM/YAKOBCHUK

High dynamic range (HDR) imaging enables the capture of an extremely wide range of the illumination present in a scene and so produces images that more closely resemble what we see with our own eyes. In this article, we explain the problem of limited dynamic range in the standard imaging pipeline and then present a survey of state-of-the-art research in HDR imaging, including the technology's history, specialized cameras that capture HDR images directly, and algorithms for capturing HDR images using sequential stacks of differently exposed images. Because this last is among the most common methods for capturing HDR images using conventional digital cameras, we also discuss algorithms to address artifacts that occur when using with this method for dynamic scenes. Finally, we consider systems for the capture of HDR video and conclude by reviewing open problems and challenges in HDR imaging.

Overview of HDR imaging

The world around us is visually rich and complex. Some of this richness comes from the wide range of illumination present in daily scenes—the illumination intensity between the brightest and the darkest parts of a scene can vary by many orders magnitude. Fortunately, the human visual system can observe very wide ranges of luminosity by means of brightness adaptation, which allows us, for example, to easily see the bright scene outside a window as well as the darkened interior. A digital camera, on the other hand, has a sensor that responds linearly to illumination; coupled with the sensor pixels' limited capacity to store energy and the noise present in the acquisition process, this fundamentally limits the sensor's measurable dynamic range. The low dynamic range (LDR) of modern digital cameras is a major factor preventing them from capturing images as humans see them (Figure 1). For this reason, an entire research community, both in academia and industry, is engaged in



FIGURE 1. (a) Images captured by standard digital cameras cannot reproduce the wide range of illumination we see in everyday scenes, even after adjusting the exposure, as illustrated by these two images taken at different exposures. (b) HDR imaging allows for the capture of a wider range of illumination; here, a stack of images was captured at different exposures (left) and merged with the algorithm described in [1] to reduce motion artifacts and produce the result shown on the right.

developing HDR imaging algorithms and systems to allow better photographs to be captured.

In this article, we describe research within the computational photography community on HDR imaging that enables the capture of a wider range of illumination than is normally captured and produces images closer to what we see with our own eyes. In a way, HDR imaging represents the epitome of computational photography: many of the solutions involved require novel optics, new acquisition processes, and clever algorithms in the back end to produce better images. As such, this article will focus only on the acquisition of HDR images and will not discuss related topics that have been extensively studied such as HDR image representation (how to compress and store HDR images) or tone mapping (turning an HDR image into an LDR image suitable for standard display) [2]. Further, because of this tutorial's strict space limitations, we cannot cover in depth the large body of existing work on HDR imaging and refer interested readers instead to textbooks and papers that survey the subject [1]–[6].

Historical background

As early as the mid-1800s—soon after the invention of photography itself—early photography pioneers were already struggling with the limited dynamic range of film and began to develop techniques that provided the basis of what we now know as HDR imaging. The French photographer Hippolyte Bayard was the first to propose that two negatives, each one properly exposed for different content, could be combined to create a well-balanced photograph. His compatriot Gustave Le Gray captured many beautiful seascape photographs with his *ciel rapporté* technique, where one negative was used for the dark sea and the other for the bright sky. Others, such as Oscar Rejlander, combined many well-exposed negatives to produce photographs that emulated contemporary paintings in which everything was properly “exposed” (Figure 2).

This idea of combining images acquired with different exposures to produce an HDR result was reintroduced for

digital photography in the 1990s (almost 150 years later) by Madden [7] and Mann and Picard [8]. However, HDR imaging received relatively little attention until the seminal paper by Debevec and Malik [9] placed it at the forefront of the burgeoning computational photography community. Since then, there has been almost 20 years of research on HDR imaging. Before we delve into this research, however, we must first review the standard imaging pipeline and understand the reasons for its limited dynamic range. In addition, we need to formalize colloquial terms such as *brightness* by introducing the appropriate radiometric units that characterize light.

The standard imaging pipeline and its limited dynamic range

The standard imaging pipeline (Figure 3) starts with a set of rays leaving the scene in the direction of the camera, with each ray carrying some amount of radiant power called *radiance* (L ; units: W/m^2sr). The rays entering the lens aperture and striking the sensor at a point are integrated over the solid angle subtended by the aperture (thereby integrating away the steradian sr term), resulting in a radiant power density at the

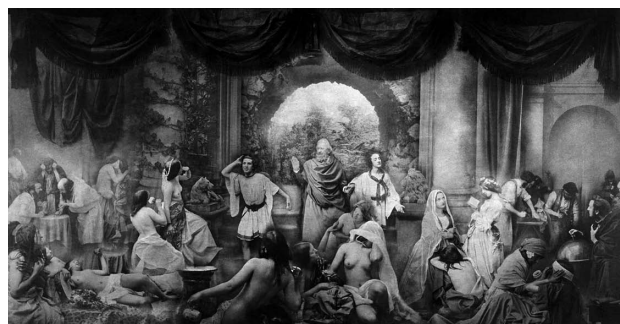


FIGURE 2. *Two Ways of Life*, Oscar Gustave Rejlander, 1857. This is one of the earliest examples of combination printing, in which differently exposed negatives are combined to extend the dynamic range of the final result. In this case, 32 negatives were combined to complete the final image. (Image in the public domain.)

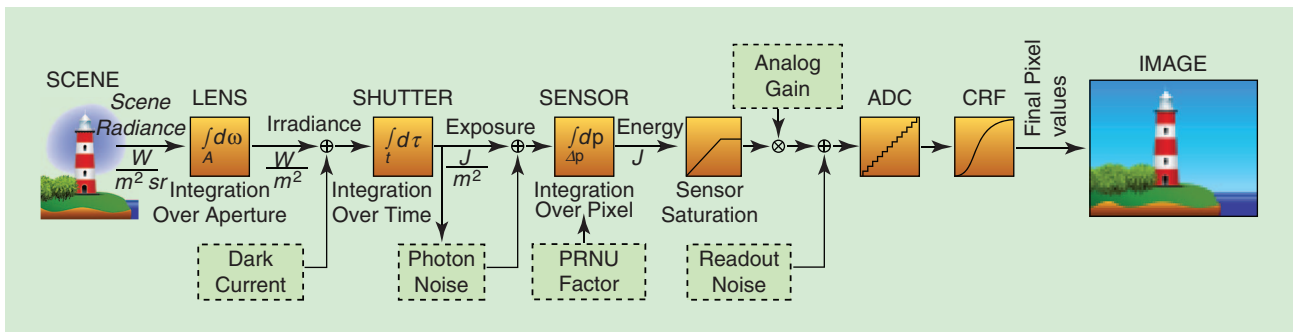


FIGURE 3. The standard imaging pipeline in modern digital cameras, inspired by diagrams in [9] and [10]. The radiance from scene rays captured by the camera are first integrated over the angle subtended by the lens aperture, over the time the shutter is open, and over the pixel’s footprint area. This energy can then be cut off by the saturation of the photon well at that pixel sensor, which limits the camera’s dynamic range. The result is then quantized by an ADC, and the CRF is applied to get the final digital pixel values. Different kinds of noise or error are injected at various stages in the pipeline, as described in the article text. (Lighthouse image designed by Freepik.com.)

sensor called *irradiance* (E ; units: W/m^2). This irradiance is then integrated over the time the shutter is open to produce an energy density, commonly referred to as *exposure* (X ; units: Jm^2). If the scene is static during this integration, the exposure can be written simply as $X(p) = E(p) \cdot t$, where p is the point on the sensor and t is the length of the exposure (integration time).

The exposure can then be integrated over the pixel’s footprint (integrating away the m^2 term) to result in the total energy (units: J) accumulated in each pixel’s photon well. The measured energy is then read out by an analog-to-digital converter (ADC), often with an analog gain factor applied to amplify the energy before it is converted. For non-raw images, the digital value is then mapped through a nonlinear camera response function (CRF) to emulate the logarithmic response of the human eye and make the final image look better. This produces the final pixel values that are output in the image file.

Two aspects of the pipeline limit the sensor’s dynamic range of measurable light. First, the pixels’ photon wells are of finite size and will saturate if too much energy is accumulated, creating an upper limit for the amount of light energy that can be measured at each pixel. Second, the minimum amount of detectable light is limited by the sources of noise in the imaging pipeline. The first is dark current, which is caused by thermal generation and induces a signal even if no photons arrive at the sensor (i.e., it is dark). Next is photon shot noise, which is caused by the discrete nature of light and is the variance of the number of photons arriving at the sensor during exposure time t . Like many arrival processes, this count is modeled by a Poisson random variable, the expected value (as well as the variance) of which is based on the true irradiance $E(p)$. The spatial nonuniformity of the sensor also causes different pixels to respond differently to the same amount of incident photons, which is modeled by the photo-response nonuniformity (PRNU) factor. Finally, there is readout noise caused by thermal generation of electrons when the signal is being read from the sensor.

Given all of these noise sources (excepting dark current), the actual measured exposure value $\hat{X}(p)$ for well-exposed

regions can be modeled as a Gaussian random variable with mean and variance [4]

$$\begin{aligned} \mu_{\hat{X}(p)} &= ga(p)E(p) \cdot t + \mu_R \\ \sigma_{\hat{X}(p)}^2 &= g^2 a(p)E(p) \cdot t + \sigma_R^2, \end{aligned} \quad (1)$$

where g is the camera gain, $a(p)$ is the PRNU factor for the pixel, and μ_R and σ_R^2 are the readout mean and variance, respectively. The Poisson nature of the photon shot noise is responsible for the pixel variance’s dependence on the irradiance. Without loss of generality, we can think of this measured exposure $\hat{X}(p)$ at each point p in the sensor as being mapped to a final digital pixel value $Z(p)$ with a function f that effectively combines the CRF with the quantization and saturation steps: $Z(p) = f(\hat{X}(p))$.

The challenge of HDR imaging, therefore, is to recover the original HDR irradiance $E(p)$ from noisy LDR images such as $Z(p)$. To do this, two main approaches have been proposed: 1) specialized HDR camera systems that measure a larger dynamic range directly and 2) capturing a stack of differently exposed LDR images that are merged together to produce an HDR result, as described in the following two sections, respectively.

Specialized HDR camera systems

Previous work on specialized HDR camera systems can be divided into two main categories: 1) those that modify the measurement properties of a single sensor to capture a larger dynamic range and 2) those that use prisms, beamsplitters, or mirrors in the optical path to image a number of sensors at different exposures simultaneously.

In the first category, researchers have proposed HDR sensors that measure light in alternate ways, such as measuring the pixel saturation time [11], counting the number of times each pixel reaches a threshold charge level [12], or incorporating a logarithmic response like that of the human eye [13]. Others, such as Nayar and Mitsunaga [14], have proposed to fit different neutral-density filters over individual pixels in the sensor to vary the amount of light absorbed at each pixel.

The main advantage of this spatially varying pixel exposures (SVE) approach is that it allows HDR imaging from a single exposure, thus avoiding the need for alignment and motion estimation. Later, Nayar et al. [15] proposed using a digital micromirror device in front of the sensor for modulating the amount of light that arrives at each pixel to acquire HDR images. Hirakawa and Simon [16] proposed another SVE system that exploits the different sensitivities already present in a regular Bayer pattern, while Schöberl et al. [17] improved this idea further, introducing a nonregular filter pattern to avoid aliasing problems. In addition, a patch-based approach to single-image HDR with SVE acquisition [18] uses a piecewise linear estimation strategy to reconstruct an irradiance image by simultaneously estimating over- and underexposed pixels as well as denoising the well-exposed ones. Finally, there has been related work that uses a spatial light modulator displaying a random mask pattern to modulate the light before it arrives at the sensor and then uses compressed sensing or sparse reconstruction to recover the HDR image [19].

In the second category, approaches include those that do not use a single sensor but rather split the light onto a set of sensors with different absorptive filters to produce simultaneous images with varying exposures. These exposures can then be merged to form the final HDR result using the stack-based approaches described in the following section. Some systems use pyramid-shaped mirrors, refracting prisms, or beamsplitters to do this [21], although each such approach suffers from parallax errors (because each “looks” through the camera lens from a slightly different angle) as well as wasted light (because of the absorptive filters in front of the sensors). Tocci et al. [20] addressed these problems with a novel beamsplitter design that efficiently reflects the light onto three different sensors to produce high-quality HDR images (Figure 4).

However, despite promising results, all of these specialized HDR systems require the manufacture of new camera hardware, and so they are not widely available today. Nevertheless, this could change as HDR imaging becomes more mainstream.

HDR imaging using image stacks

With conventional cameras, the most practical approach for HDR imaging is to capture a sequence of LDR images at different exposures and combine them into a final HDR result [7]–[9]. Specifically, if we acquire a stack of N different exposures Z_1, \dots, Z_N , we can merge them and estimate the irradiance map \tilde{E} using a simple weighting scheme that takes into account the measured irradiance $\hat{E}_i = \hat{X}_i(p)/t_i$ from each image:

$$\tilde{E}(p) = \frac{\sum_{i=1}^N w_i(p) \cdot \hat{X}_i(p)/t_i}{\sum_{i=1}^N w_i(p)}. \quad (2)$$

Here, the measured exposure \hat{X}_i can be recovered from well-exposed pixel values using the inverse of the camera response function: $\hat{X}_i(p) = f^{-1}(Z_i(p))$. Of course, this requires the CRF to be known, but methods have been proposed to estimate it from the image stack [9], even for highly dynamic scenes [22].

Because poorly exposed pixels do not have a good estimate for the irradiance map, the weight $w_i(p)$ should be adjusted at each pixel based on how well-exposed it is. For example, Debevec and Malik [9] proposed a simple triangle function for this weight that gives priority to pixels in the middle of the pixel range and reduces the influence of poorly exposed pixels: $w_i(p) = \min(Z_i(p), 255 - Z_i(p))$, where we assume the pixel values range from 0 to 255. Once the stack of images has been merged in this way, the resulting irradiance map \tilde{E} is output as the final HDR result. This method is commonly implemented on modern smartphones to extend their camera’s dynamic range (i.e., “HDR mode”).

Fundamental limits on irradiance estimation performance

It is interesting to understand the fundamental limits of irradiance estimation performance for stack-based algorithms such as these. To study this, the problem of irradiance estimation from an image stack can be posed as a parameter estimation

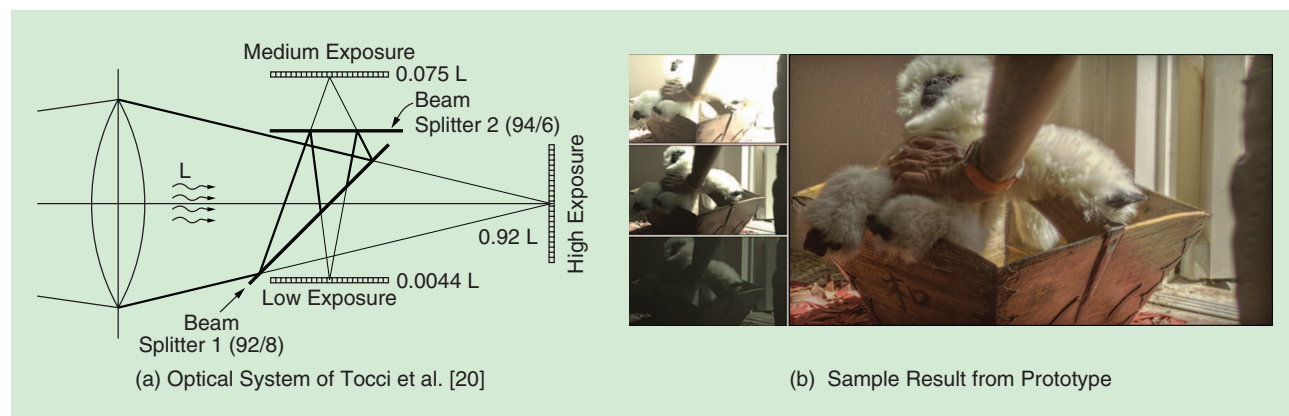


FIGURE 4. In the optical system of Tocci et al. [20], (a) two beamsplitters reflect the light so that the three sensors capture images with 92%, 7.52%, and 0.44% of the total light gathered by the camera lens (increasing the dynamic range by a factor of over 200 ×), and only 0.04% of it is wasted. (b) shows the sample HDR result captured by the camera (the three captured LDR images are on left); note that the detail in both the white fur and dark regions is captured faithfully, even though it does not appear simultaneously in any of the input images. (Figure courtesy of [20].)

problem from a set of noisy samples. In the case of static scenes, N independent samples $\hat{X}_1(p), \dots, \hat{X}_N(p)$ following the random model in (1) are given per pixel, corresponding to exposure times t_1, \dots, t_N . Assuming the camera parameters are known from a calibration stage, the only unknown parameter in (1) is the irradiance $E(p)$ reaching each pixel p .

In this statistical framework, the Cramér–Rao lower bound (CRLB) gives a lower bound on the variance of any unbiased estimator of $E(p)$ computed from those samples. Aguerrebere et al. [4] introduced the CRLB for this problem and showed that, because the bound cannot be attained, no efficient estimator exists for $E(p)$ under the considered hypotheses. Nevertheless, it was shown experimentally that the approximation of the maximum-likelihood estimator (MLE) proposed by Granados et al. [23] not only outperforms the other evaluated estimators but also has nearly optimal behavior. Theoretically, the MLE is efficient for a large number of samples (asymptotically efficient), which is not the case in HDR imaging, where very few samples are usually available (normally $N = 2$ to 4 exposures). Therefore, it is remarkable that, under the considered hypotheses, the MLE is still experimentally the best possible estimator for the pixel-wise irradiance estimation for static scenes. Improvements, however, may be possible by combining information from different pixel positions with similar irradiance values, such as in recent patch-based denoising approaches [24], or even by considering information from saturated samples [4].

Handling dynamic scenes

The stack-based HDR capture algorithms described in the previous section work very well when the scene is static and the camera is tripod-mounted. However, when the scenes are dynamic or the camera moves while the different pictures are being captured, the images in the stack will not line up properly with one another. This misalignment results in ghost-like artifacts in the final HDR image, which are often more objectionable than the limited dynamic range that is being compensated for (see Figure 5). Because this is the most common scenario in imaging, there has been almost 20 years of research into HDR deghosting algorithms that seek to eliminate these artifacts from motion. Specifically, three different

kinds of methods have been proposed to deal with motion, each of which we discuss in the three sections that follow, using a taxonomy similar to those in two previous publications by the first author [1], [10]. Because of space limitations, we limit the discussion here to a couple of key algorithms in each category.

Algorithms that align the different exposures

The first kind are algorithms that attempt to deghost the HDR reconstruction by warping the individual images in the stack to match a reference image and so eliminate misalignment artifacts. Unlike the rejection methods discussed in the “Algorithms That Reject Misaligned Information” and “Patched-Based Optimization Algorithms” sections, these algorithms can actually move content around in each image and can, therefore, potentially handle dynamic HDR objects.

The simplest methods in this category assume the images can be aligned with rigid transformations. For example, a common method is to compute scale-invariant feature transform (commonly called *SIFT*) features in the image and use them to estimate a homography that warps the images to match [25]. Of course, these simple rigid-alignment algorithms cannot handle artifacts caused by parallax due to camera translation or from significant motion in the scene, although they can serve as a preprocess for more complex algorithms, such as those described later in the article.

One of the first algorithms of this kind was proposed by Bogoni [26]. This method first uses an affine motion estimation step to globally align the images and then estimates motion using optical flow to further align the images. To make the optical flow more robust, some have proposed acquisition schemes to make the different exposures more similar. The Fibonacci exposure bracketing work of Gupta et al. [27], for example, cleverly adjusts the exposure times in the sequence so that the longer exposure times are equal to the sum of the shorter exposure times. Because of this, optical flow can be computed between a longer exposure and the sum of the shorter exposures, thereby ensuring that the two images will have similar exposure times and, therefore, comparable motion blur.

The state-of-the-art HDR alignment algorithm is perhaps the work of Zimmer et al. [28], which aligns the images using



FIGURE 5. Ghosting artifacts can occur when stack-based HDR algorithms are applied to dynamic scenes. (a) Stack of input LDR images. Note, how some images capture the details in the dark sweater, while others capture the detail in the bright exterior. (b) HDR results from the standard HDR merging algorithm produces ghosting artifacts because of the motion. (c) HDR results from the patch-based optimization algorithm of Sen et al. [1] contains detail in all regions of the image without artifacts.

an energy-based optical flow optimization robust to changes in exposure. Specifically, their energy function has a data term that encourages the image to align to the reference and a regularizer that enforces smooth flow wherever the reference is poorly exposed. However, these alignment algorithms all suffer from the problem of finding good correspondences, which is extremely difficult, in particular for highly dynamic scenes with deformable motion (e.g., a person moving). Furthermore, scenes with occlusion and/or parallax do not even have valid correspondences between the images in these regions, making it impossible to align the images in the stack correctly. Therefore, the HDR results from alignment algorithms often still contain objectionable ghosting artifacts for scenes with complex motion.

Algorithms that reject misaligned information

A second set of algorithms for HDR reconstruction assume that the camera is static (or that the images have been preregistered using a rigid alignment process, such as those described in the “Algorithms That Align the Different Exposures” section) and that the scene motion is localized, meaning that the majority of pixels contain no motion artifacts. The basic goal of these methods is to identify those pixels that are affected by motion and those that are not. The pixels that do not contain motion artifacts can be merged using the standard HDR merging algorithms described in the “HDR Imaging Using Image Stacks” section. For the pixels that are affected by motion, however, only a subset of the images deemed to be static at these pixels will be merged to suppress artifacts from moving objects.

To accomplish this, two different kinds of rejection methods are possible: 1) those in which a reference image is specified by the user and 2) those that do not use a reference image. For algorithms in the first category, the user first selects an image from the stack as the reference. These algorithms then simply revert back to this reference for any pixels where motion is detected so that the main difference between them is in how they detect motion. For example, the method of Grosch [29] assumes two images in the stack and predicts values in the second image by multiplying the values in the reference by the ratio of the exposure times, taking into account the nonlinear camera response curves. With this approach, a pixel is deemed to be affected by motion if the actual color is beyond a given threshold from the predicted value. In such cases, the algorithm simply reverts back to using the values in the reference image for these pixels.

Gallo et al. [30] improved on this work by using the log-irradiance domain to do the threshold comparisons. Further, for robustness they compare patches instead of individual pixels, so that a patch from an image in the stack would be merged with the corresponding patch from the reference only if a certain number of pixels meet the threshold constraint. To

reduce visible seams between different patches, the authors apply Poisson blending to the final results.

In the second category are rejection algorithms without a reference image, which must select a “static” subset of images at every pixel to merge to produce HDR values. These methods have a fundamental advantage over those that utilize a single reference image because motion may occur in areas where the reference might be poorly exposed. At these pixels, an HDR value cannot be properly computed solely from the reference image. However, rejection algorithms that do not use a reference must ensure that subsets are selected for neighboring pixels in a way that does not introduce artifacts.

Reinhard et al. [3] proposed one of the earliest methods in this category. For every pixel that is deemed to be affected by motion, the authors try to use the longest exposure that is not saturated (effectively, a single-image subset). To determine which pixels are affected by motion, they first compute the variance of the irradiance values at each pixel p , weighted to exclude poorly exposed pixels. This estimated variance is then thresholded, and the result is smeared out with a 3×3 kernel to reduce edge and noise effects. Adjacent regions are then joined together to form the “ghosted” regions for

which a single image from the stack will be used. To select which image they will use for each region, the authors find the biggest irradiance value in the region that is not in the top 2% (deemed to be outliers). They then select the longest exposure that includes this value within its valid range to fill in this ghosted region, because the longest exposure will contain least noise. To further suppress artifacts, Reinhard et al. linearly interpolate this exposure with the original HDR result, using the per-pixel variance as a blending parameter.

An alternative approach is proposed by

Khan et al. [31]; here, instead of detecting and handling differently the pixels affected by motion, the authors propose to iteratively weight the contribution of each pixel depending on the probability of its being static (i.e., belonging to the background of the scene). To do this, they assume that most of the pixels are of the static background and so determine the probability of a pixel being static by measuring its similarity to the neighborhood around it.

Finally, some recent methods cleverly use rank minimization to deghost HDR images [32], [33]. These methods are based on the observation that if the scene is static, the different exposure images $X(p)$ would simply be linear scalings of one another. Therefore, they use the different exposure images to construct a matrix and essentially minimize its rank to solve for the motion-free image.

The biggest problem with these and other rejection algorithms is that they cannot handle dynamic HDR content because they do not move information between pixels but rather only merge information from corresponding pixels across the image stack. Therefore, if different parts of a moving HDR

The biggest problem with rejection algorithms is that they cannot handle dynamic HDR content because they do not move information between pixels but rather only merge information from corresponding pixels across the image stack.

object are well exposed in disjoint regions of the different images, these parts cannot be brought together to produce an acceptable result.

Patch-based optimization algorithms

Recently, Sen et al. [1] proposed a new alternative for HDR deghosting that uses patch-based optimization, which addresses the problems of both the rejection and alignment methods. Specifically, a formulated equation codifies the objective of most reference-based HDR reconstruction algorithms: 1) to produce an HDR result that resembles the reference image in the parts where the reference is well exposed and 2) to leverage well-exposed information from other images in the stack wherever the reference is poorly exposed. This HDR synthesis equation can be written as

$$\text{Energy}(E) = \sum_{p \in \text{pixels}} [\alpha_{\text{ref}}(p) \cdot (f^{-1}(Z_{\text{ref}}(p))/t_{\text{ref}} - E(p))^2 + (1 - \alpha_{\text{ref}}(p)) \cdot E_{\text{BDS}}(E | Z_1, \dots, Z_N)]. \quad (3)$$

The first term states that the desired HDR image E should be close in an L_2 sense to the LDR reference Z_{ref} mapped to the linear irradiance domain by applying the inverse camera response function f^{-1} and dividing by the exposure time t_{ref} . This is only to be done for the pixels where the reference is

properly exposed, as given by the α_{ref} term, which is a trapezoidal function in the pixel value domain [similar to the weighting function in (2)] that favors intensities near the middle of the pixel value range.

In the regions where the reference image Z_{ref} is poorly exposed (indicated by $1 - \alpha_{\text{ref}}$), the algorithm draws information from the other images in the stack using a bidirectional similarity metric, given by the E_{BDS} term. This energy term enforces that for every pixel patch in the image stack (given by Z_1, \dots, Z_N), there must be a similar patch in the final result E , and vice versa. The first similarity ensures that as much well-exposed content from the image stack is included in the final HDR result, while the second ensures that the final result does not contain objectionable artifacts, as these artifacts would not be found anywhere in the stack. This energy equation is optimized with an iterative method that solves for the aligned LDR images and the HDR image simultaneously, producing high-quality results (Figure 6).

Patch-based optimization algorithms like this are fundamentally different from those discussed in the “Algorithms that Align the Different Exposures” section, which warp the images to match based on correspondences. As was pointed out earlier, alignment methods fail in cases of occlusion or parallax (which happen commonly in dynamic scenes) because



FIGURE 6. (a) and (b) show sample HDR results (right) from the input LDR images (left) using the patch-based optimization of Sen et al. [1].

they do not have valid correspondences in these regions and so the images cannot be aligned in these parts. Patch-based HDR reconstruction, on the other hand, is related to patch-based image synthesis methods (e.g., for single-image hole filling) because they both use a patch-based similarity optimization to resynthesize content in the final reconstruction without an underlying correspondence. Because of this advantage, these methods have proved to be the most successful HDR deghosting algorithms proposed to date.

For example, a recent state-of-the-art report by Tursun et al. [6] testing many deghosting algorithms found that the algorithm of Sen et al. [1] and the later, related method of Hu et al. [34] ranked first and second over other deghosting techniques by a fairly large margin. The success of patch-based optimization for HDR reconstruction has led others to explore ways to further improve the quality of these approaches. For example, Aguerreberre et al. [24] focused on reducing the noise of the estimated irradiance. First, this method synthesizes a “reference” containing well-exposed, de-ghosted information in all parts of the image using Poisson image editing (although the method in Sen et al. [1] could also be used). Noise is then reduced through a patch-based denoising method that finds all patches in the image stack within a threshold to each patch in the reference, where the L_2 distance between patches is normalized by the variance from (1). The MLE of the patch-centers at each pixel is then computed to significantly reduce the noise in the final result.

HDR video

Up to now, we have focused exclusively on the HDR acquisition of still images. However, the problem of capturing HDR video sequences is of considerable interest as well. For example, filmmaking companies incur a significant cost to light sets, a cost that would be largely eliminated by high-quality, HDR video systems. For this reason, professional movie camera system suppliers such as RED have been pushing the dynamic range of standard sensors. Moreover, specialized HDR camera systems such as that of Tocci et al. [20] have been proved capable of capturing high-quality HDR video, although they are not yet widely available.

For conventional digital cameras, the only way to capture HDR video is to alternate exposures through the entire sequence. This problem was first tackled by Kang et al. [35], who use gradient-based optical flow to compute a bidirectional flow from the current frame to neighboring frames and unidirectional flows from neighboring frames to the current frame (four flows total). Once computed, the flows can be used to produce four warped images by deforming each of the two neighboring frames. The resulting images can be merged with the reference to produce an HDR image at every frame of the

sequence, while rejecting the pixels that are still misaligned, to avoid artifacts.

The state of the art in HDR video reconstruction is the work of Kalantari et al. [5], which extended the patch-based optimization work of Sen et al. [1] to produce coherent HDR video streams. Specifically, they modify the HDR image synthesis equation (3) to enforce temporal coherence by performing a bidirectional similarity between adjacent frames. In addition, they use optical flow during the optimization to constrain the patch-based search, which produces a stream of high-quality HDR frames.

Open problems and challenges

Despite the tremendous progress of the computational photography community on HDR imaging over the last 20 years, many challenges remain. For example, the capture of high-quality HDR images of highly dynamic scenes with conventional digital cameras is still a challenging problem. Although state-of-the-art deghosting algorithms like the patch-based optimization of Sen et al. [1] can suppress many of the ghosting artifacts

that would normally occur in these scenes, these methods cannot recover scene content that is poorly exposed in the reference image and is not visible in any of the other images in the stack. Moreover, the patch-based optimization in these algorithms is computationally expensive and can take several minutes to compute an image. This limits the applicability of these methods to long video sequences or for real-time, on-board computation in current smart phones, for example.

It is entirely possible that new sensor technologies, such as Fuji Film’s recent Super CCD EXR sensor, will bypass the problems inherent in stack-based methods by capturing a single image with extended dynamic range. However, even these new technologies will likely raise interesting questions, such as how users will employ and interact with HDR images. Furthermore, as HDR imaging becomes more mainstream, we expect that new applications for HDR imaging (such as for medical imaging or manufacturing) will be proposed and explored.

Conclusions

In this article, we first summarized the main aspects of HDR imaging, starting with an overview of the problem of limited dynamic range in standard digital cameras and the physical constraints responsible for this limitation. We then surveyed state-of-the-art approaches developed to tackle the HDR imaging problem, focusing on both specialized HDR camera systems and stack-based approaches captured with standard cameras. For the latter, we discussed algorithms to address ghosting artifacts that can occur when capturing dynamic scenes. Finally, we discussed algorithms for capturing HDR video and concluded with a review of open problems in HDR imaging. We hope that

Patch-based HDR reconstruction is related to patch-based image synthesis methods (e.g., for single-image hole filling) because they both use a patch-based similarity optimization to resynthesize content in the final reconstruction without an underlying correspondence.

this article encourages researchers from areas such as signal processing, solid-state devices, and image processing to continue to pursue this interesting set of problems.

Acknowledgments

We would each like to thank our previous coauthors of articles on this topic. P. Sen was funded by National Science Foundation grants IIS-1342931 and IIS-1321168, and C. Aguerrebere was funded by the U.S. Department of Defense.

Authors

Pradeep Sen (psen@ece.ucsb.edu) received his B.S. degree from Purdue University, West Lafayette, Indiana, and his M.S. and Ph.D. degrees from Stanford University, California. He is currently an associate professor with the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. He has coauthored more than 30 technical publications, including eight in *ACM Transactions on Graphics*. His research interests include algorithms for image synthesis, computational image processing, and computational photography. He received the 2009 National Science Foundation CAREER Award and more than US\$2.2 million in funding.

Cecilia Aguerrebere (cecilia.aguerrebere@duke.edu) received her B.S., M.S., and Ph.D. degrees in electrical engineering from the Universidad de la República, Montevideo, Uruguay, in 2006, 2011, and 2014, respectively; an M.S. degree in applied mathematics from the École normale supérieure de Cachan, France, in 2011; and a Ph.D. degree in signal and image processing from Télécom ParisTech, France, in 2014 (as part of a joint Ph.D. program with Universidad de la República). Since August 2015, she has been with the Department of Electrical and Computer Engineering at Duke University, Durham, North Carolina, where she holds a post-doctoral research associate position.

References

- [1] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203:1–203:11, 2012.
- [2] R. K. Mantiuk, K. Myszkowski, and H.-P. Seidel, *High Dynamic Range Imaging*. Hoboken, NJ: Wiley, 2015.
- [3] E. Reinhard, G. Ward, S. N. Pattanaik, and P. E. Debevec, *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting*. San Mateo, CA: Morgan Kaufmann, 2005.
- [4] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Best algorithms for HDR image generation. A study of performance bounds," *SIAM J. Imaging Sci.*, vol. 7, no. 1, pp. 1–34, 2014.
- [5] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen, "Patch-based high dynamic range video," *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, vol. 32, no. 6, pp. 202:1–202:8, Nov. 2013.
- [6] O. T. Tursun, A. O. Akyuz, A. Erdem, and E. Erdem, "The state of the art in HDR deghosting: A survey and evaluation," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 683–707, 2015.
- [7] B. C. Madden, "Extended intensity range imaging," Univ. of Pennsylvania, Tech. Rep. MS-CIS-93-96, 1993.
- [8] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IS&T*, 1995, pp. 442–448.
- [9] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.
- [10] O. Gallo and P. Sen, "Stack-based algorithms for HDR capture and reconstruction," in *High Dynamic Range Video, From Acquisition to Display and*

Applications, 2nd ed., F. Dufaux, P. L. Callet, R. Mantiuk, and M. Mrak, Eds. Amsterdam, The Netherlands: Elsevier, 2016, ch. 3.

- [11] V. Brajovic and T. Kanade, "A sorting image sensor: an example of massively parallel intensity-to-time processing for low-latency computational sensors," in *Proc. ICRA*, vol. 2, pp. 1638–1643, Apr. 1996.
- [12] H. Zhao, B. Shi, C. Fernandez-Cull, S.-K. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, Apr. 2015, pp. 1–10.
- [13] U. Seger, U. Apel, and B. Höflinger, "HDR-Imagers for natural visual perception," in *Handbook of Computer Vision and Application*, B. Jähne, H. Haußecker, and P. Geißler, Eds. New York: Academic, 1999, vol. 1, pp. 223–235.
- [14] S. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2000, vol. 1, pp. 472–479.
- [15] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable imaging: Towards a flexible camera," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 7–22, 2006.
- [16] K. Hirakawa and P. Simon, "Single-shot high dynamic range imaging with conventional camera hardware," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 1339–1346.
- [17] M. Schöberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup, "High dynamic range video by spatially non-regular optical filtering," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2012, pp. 2757–2760.
- [18] C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau, and P. Musé, "Single shot high dynamic range imaging using piecewise linear estimators," in *Proc. IEEE Int. Conf. Computational Photography (ICCP)*, 2014, pp. 1–10.
- [19] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia, "Convolutional sparse coding for high dynamic range imaging," *Comput. Graph. Forum*, vol. 35, no. 2, 2016.
- [20] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 41:1–41:10, July 2011.
- [21] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," *Int. J. Comput. Vis.*, vol. 58, no. 1, pp. 7–17, June 2004.
- [22] A. Badki, N. K. Kalantari, and P. Sen, "Robust radiometric calibration for dynamic scenes in the wild," in *Proc. IEEE Int. Conf. Computational Photography (ICCP)*, Apr. 2015, pp. 1–10.
- [23] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 215–222.
- [24] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Simultaneous HDR image reconstruction and denoising for dynamic scenes," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2013, pp. 1–11.
- [25] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proc. Int. Conf. Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 2007, pp. 49–56.
- [26] L. Bogoni, "Extending dynamic range of monochrome and color images through fusion," in *Proc. IEEE Intl. Conf. Pattern Recognition (ICPR)*, 2000, pp. 3007–3016.
- [27] M. Gupta, D. Iso, and S. Nayar, "Fibonacci exposure bracketing for high dynamic range imaging," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2013, pp. 1473–1480.
- [28] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 405–414, Apr. 2011.
- [29] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," in *Proc. Int. Symp. Vision, Modeling and Visualization*, 2006, pp. 277–284.
- [30] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE Intl. Conf. Computational Photography (ICCP)*, 2009, pp. 1–7.
- [31] E. Khan, A. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2006, pp. 2005–2008.
- [32] C. Lee, Y. Li, and V. Monga, "Ghost-free high dynamic range imaging via rank minimization," *IEEE Signal Processing Lett.*, vol. 21, no. 9, pp. 1045–1049, Sept. 2014.
- [33] T.-H. Oh, J.-Y. Lee, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 37, no. 6, pp. 1219–1232, June 2015.
- [34] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Computer Vision and Pattern Recog. (CVPR)*, June 2013, pp. 1163–1170.
- [35] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, July 2003.