High Dynamic Range Imaging of Natural Scenes

Pradeep Sen, Senior Member, IEEE, and Cecilia Aguerrebere

Abstract

High dynamic range (HDR) imaging enables the capture of a wider range of the illumination present in a scene and therefore produces images that more closely resemble what we see with our own eyes. In this paper, we explain the problem of limited dynamic range in the standard imaging pipeline and then present a survey of the state-of-the-art research in HDR imaging, including the history of HDR imaging, specialized cameras that capture HDR images directly, and algorithms for capturing HDR images using sequential stacks of differently-exposed images. Since this last one is the most common method for capturing HDR images using conventional digital cameras, we also discuss algorithms to address artifacts that occur with this method for dynamic scenes. Finally, we discuss systems for the capture of HDR video, and conclude with a review of open problems and challenges in HDR imaging.

Index Terms

High dynamic range (HDR) imaging, HDR performance bounds, HDR deghosting, HDR video

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 to easily see the bright scene outside a window as well as the darkened interior, for example. A digital camera, on the other hand, has a sensor that responds linearly to illumination which, coupled with the limited capacity of the sensor pixels to store energy and the noise present in the acquisition process, fundamentally limits its measurable dynamic range. The low-dynamic range (LDR) of modern digital cameras is a major reason preventing them from capturing images like we see them (Fig. 1). For this reason, there is an entire research community, both in academia and industry, engaged in the development of HDR imaging algorithms and systems that enable the capture of better photographs.

P. Sen is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106, USA, e-mail: psen at ece.ucsb.edu

C. Aguerrebere is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA, e-mail: cecilia.aguerrebere at duke.edu

Manuscript received December 21, 2015
In this paper, we shall describe research in the computational photography community on high dynamic range (HDR) imaging, which enables the capture of a wider range of illumination than is normally captured and produces images that are closer to what we see with our own eyes. In a way, HDR imaging is perhaps the epitome of computational photography: many of the solutions require novel optics, new acquisition processes, and clever algorithms in the back-end to produce better images. As such, this paper will focus only on the acquisition of HDR images, and will not discuss related topics which have been extensively studied such as HDR image representation (how to compress and store HDR images) or tone mapping, which turns an HDR image into a low-dynamic range image suitable for a standard display [2]. Furthermore, because of the strict space limitations of this tutorial, we cannot cover in depth the large body of work that has been done on HDR imaging and refer interested readers instead to textbooks and papers that survey the subject [1]–[6].

I. HISTORICAL BACKGROUND

As early as the mid-1800’s – 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 laid 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 using 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 where everything was properly “exposed” (Fig. 2).
Fig. 2. Two Ways of Life, Oscar Gustave Rejlander, 1857. This is one of the earliest examples of combination printing, wherein differently-exposed negatives are combined together to extend the dynamic range of the final result. In this case, 32 negatives were combined together to complete the final image. (Image in the public domain)

This idea of combining images acquired with different exposures to produce a high-dynamic range result was re-introduced to digital photography in the 1990’s (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. Furthermore, we must formalize colloquial terms such as “brightness” by introducing the appropriate radiometric units that characterize light.

II. THE STANDARD IMAGING PIPELINE AND ITS LIMITED DYNAMIC RANGE

The standard imaging pipeline (Fig. 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 sensor called irradiance \( E \), units: \( W/m^2 \). This irradiance is then integrated over the time that the shutter is open to produce an energy density, commonly referred to as exposure \( X \), units: \( J/m^2 \). 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 it before it is converted. For non-RAW images, the digital value is then mapped through a non-linear
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). Then there 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 whose expected value (as well as its variance) is based on the true irradiance $E(p)$. The spatial non uniformity of the sensor also causes different pixels to respond differently to the same amount of incident photons, which is modeled by the photo-response non uniformity (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 (except for 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]:

$$\mu_{\hat{X}(p)} = ga(p)E(p) \cdot t + \mu_R, \quad \sigma^2_{\hat{X}(p)} = g^2a(p)E(p) \cdot t + \sigma^2_R,$$  

where $g$ is the camera gain, $a(p)$ is the PRNU factor for the pixel and $\mu_R$ and $\sigma^2_R$ are the readout mean and variance, respectively. The Poisson nature of the photon shot noise is responsible for the dependence of the pixel variance 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 is therefore to recover the original high dynamic range irradiance \( E(p) \) from noisy, LDR images like \( Z(p) \). To do this, two main approaches have been proposed: 1) specialized HDR camera systems that measure a larger dynamic range directly (Sec. III), and 2) capturing a stack of differently-exposed LDR images that are merged together to produce an HDR result (Sec. IV).

### III. 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, or 2) systems with prisms, beam-splitters, or mirrors in the optical path to image a number of sensors at different exposures simultaneously. For 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 having a logarithmic response like the human eye [13].

Others, such as Nayar and Mitsunaga [14] 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 \textit{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 to use a digital micromirror device (DMD) in front of the sensor to modulate the amount of light arriving 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 non-regular filter pattern to avoid aliasing problems. A patch-based approach to single-image HDR with SVE acquisition was also later proposed [18], which uses a piecewise linear estimation strategy to reconstruct an irradiance image by simultaneously estimating over- and under-exposed pixels as well as denoising the well-exposed ones. Finally, there has also 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].

The second category of approaches includes 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 together to form the final HDR result using the stack-based approaches described in the next section. Some systems use pyramid-shaped mirrors, refracting prisms, or beam-splitters to do this [21], although this approach suffers from parallax errors (because each “looks” through the camera lens from a slightly different angle) as well as wasted light (because
In the optical system of Tocci et al. [20], (a) two beam-splitters 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) Sample HDR result captured by the camera (the three captured LDR images are shown 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.

IV. HDR IMAGING USING IMAGE STACKS

For conventional cameras, the most practical approach for high-dynamic range 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 \( Z_1, \ldots, Z_N \), we can merge them together and estimate the irradiance map \( \tilde{F} \) using a simple weighting scheme that takes into account the measured irradiance \( \hat{X}_i(p)/t_i \) from each image:

\[
\tilde{F}(p) = \frac{\sum_{i=1}^{N} w_i(p) \cdot \hat{X}_i(p)/t_i}{\sum_{i=1}^{N} w_i(p)}.
\]

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].

Since 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 that are in the middle of the pixel range and reduces the influence of pixels that are poorly exposed:

$$w_i(p) = \min \left( Z_i(p), 255 - Z_i(p) \right)$$

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 smart phones to extend the dynamic range of their cameras (i.e., “HDR mode”).

A. Fundamental limits on irradiance estimation performance

It is interesting to understand what are the fundamental limits of irradiance estimation performance for stack-based algorithms like these. To study this, the problem of irradiance estimation from an image stack can be posed as a parameter estimation problem from a set of noisy samples. In the case of static scenes, $N$ independent samples $\hat{X}_1(p), \ldots, \hat{X}_N(p)$ following the random model in Eq. 1 are given per pixel, corresponding to exposure times $t_1, \ldots, t_N$. Assuming the camera parameters are known from a calibration stage, the only unknown parameter in Eq. 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, since 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 a 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 arise by combining information from different pixel positions with similar irradiance values, such as in recent patch-based denoising approaches [24], or even considering information from saturated samples [4].

V. 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 in the first place (see Fig. 5). Since this is the most common scenario in imaging, there has been almost 20 years of research into HDR deghosting algorithms.
Fig. 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 (Sec. IV) produces ghosting artifacts because of the motion. (c) HDR results from the patch-based optimization algorithm of Sen et al. [1] (Sec. V-C) contains detail in all regions of the image without artifacts.

which seek to eliminate these artifacts from motion. Specifically, there have been three different kinds of methods proposed for dealing with motion, which we will discuss in the next sections using a taxonomy similar to those in two previous publications by the first author [1], [10]. For space limitations, we limit our discussion to a couple of key algorithms in each category.

A. Algorithms that Align the Different Exposures

These algorithms attempt to deghost the HDR reconstruction by warping the individual images in the stack to match a reference image to eliminate misalignment artifacts. Unlike the rejection methods discussed in the next section, these algorithms can actually move content around in each image and can therefore potentially handle dynamic HDR objects.

The simplest methods of this category assume the images can be aligned with rigid transformations. For example, a common method is to compute 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 pre-process for more complex algorithms such as those described below.

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. In order 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 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 that finding good correspondences 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.

B. Algorithms that Reject Misaligned Information

These algorithms for HDR reconstruction assume the camera is static (or that the images have been pre-registered using a rigid alignment process such as those just described) 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 which pixels are affected by motion and which ones are not. The pixels that do not contain motion artifacts can be merged using the standard HDR merging algorithms described in Sec. IV. 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 together in order to suppress artifacts from moving objects.

To do this, there are two different kinds of rejection methods: 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 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 non-linear camera response curves. In 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 these cases, the algorithm simply reverts back to using the values in the reference image for these pixels.

Gallo et al. [30] improved upon this work by using the log-irradiance domain to do the threshold comparisons. Furthermore, 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, they apply Poisson blending to the final results.

The second category of approaches are rejection algorithms without a reference image, which must select a “static” subset of images at every pixel to merge together to produce HDR values. These methods have a fundamental advantage over those that utilize a single reference image since motion might 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.

The book by Reinhard et al. [3] proposed one of the earliest methods like this. For every pixel that is deemed to be affected by motion, they 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 $3 \times 3$ 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, they find the biggest irradiance value in the region that is not in the top 2%, which are deemed to be outliers. They then select the longest exposure that includes this value within its valid range to fill in this ghosted region, since the longest exposure will contain less noise. To further suppress artifacts, they 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], who instead of detecting and handling differently the pixels affected by motion, propose to iteratively weight the contribution of each pixel depending on its probability to be static (i.e., belong to the background of the scene). To do this, they assume that most of the pixels are of the static background and therefore determine the probability of a pixel being static by measuring its similarity to the neighborhood around it.

Finally, there have been recent methods that 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 each other. Therefore, they construct a matrix from the different exposures images and essentially minimize its rank to solve for the motion-free image.

The biggest problem with rejection algorithms like all of these is that they cannot handle dynamic HDR content, since 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 object are well exposed in disjoint regions of the different images, these cannot be brought together to produce an acceptable result.
C. Patch-Based Optimization Algorithms

Recently, Sen et al. [1] proposed a new alternative for HDR deghosting based on patch-based optimization which addresses the problems of rejection and alignment methods. Specifically, they formulated an equation that 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:

$$\text{Energy}(E) = \sum_{p \in \text{pixels}} \left[ \alpha_{\text{ref}}(p) \cdot \left( f^{-1}(Z_{\text{ref}}(p))/t_{\text{ref}} - E(p) \right)^2 + (1 - \alpha_{\text{ref}}(p)) \cdot E_{\text{BDS}}(E \mid Z_1, \ldots, Z_N) \right].$$  \hspace{1cm} (3)

The first term states that the desired HDR image $E$ should be close in an $L_2$ sense to the LDR reference $Z_{\text{ref}}$ mapped to the linear irradiance domain by applying the inverse camera response function $f^{-1}$ and dividing by the exposure time $t_{\text{ref}}$. This is only to be done for the pixels where the reference is properly exposed, as given by the $\alpha_{\text{ref}}$ term, which is a trapezoidal function in the pixel value domain (similar to the weighting function in Eq. 2) that favors intensities near the middle of the pixel value range.

In the regions where the reference image $Z_{\text{ref}}$ 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 $E_{\text{BDS}}$ term. This energy term enforces that for every pixel patch in the image stack (given by $Z_1, \ldots, Z_N$), 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 (Fig. 6).

Patch-based optimization algorithms like this are fundamentally different from the alignment algorithms discussed in Sec. V-A, which warp the images to match based on correspondences. As we discussed, alignment methods fail in cases of occlusion or parallax (which happen commonly in dynamic scenes) since 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 instead related to patch-based image synthesis methods (e.g., for single image hole-filling), since 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 proven 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] that tested 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, Aguerrebere et al. [24] focused on reducing the noise of the estimated irradiance. First, their method synthesizes a “reference” containing well-exposed, deghosted information in all parts of the image using Poisson image editing, although the method of Sen et al. [1] could also be used. They then reduce noise 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 Eq. 1. The maximum-likelihood estimator (MLE) of the patch-centers at each pixel is then computed to significantly reduce the noise in the final result.

VI. 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 extremely important. 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 systems such as RED have been pushing the dynamic range of standard sensors. Furthermore, specialized HDR camera systems such as the one of Tocci et al. [20] have been shown to be able to capture 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 these neighboring frames, and unidirectional flows from the 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 together with the reference to produce an HDR image at every frame of the sequence, making sure to reject 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 (Eq. 3) to enforce temporal coherence by performing a bi-
directional similarity between adjacent frames. Furthermore, they use optical flow during the optimization
to constrain the patch-based search, which produces a stream of high-quality HDR frames.

VII. OPEN PROBLEMS AND CHALLENGES

Despite the tremendous progress of the computational photography community on HDR imaging in
the last twenty 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. Furthermore, 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 with 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 use and interact with high-dynamic range 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.

VIII. CONCLUSIONS

This paper presented a summary of 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 presented a survey of the 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 this paper encourages
researchers from areas such as signal processing, solid state devices, and image processing to continue
to pursue this interesting problem.
ACKNOWLEDGMENTS

The authors would like to thank their co-authors on their respective papers. 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.

REFERENCES


**Pradeep Sen** received the B.S. degree from Purdue University, and the M.S. and Ph.D. degrees from Stanford University. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of California at Santa Barbara. He has co-authored over 30 technical publications, including eight SIGGRAPH/ACM Transactions on Graphics publications. His research interests include algorithms for image synthesis, computational image processing, and computational photography. He has received over $2.2 million in funding, including the 2009 NSF CAREER Award.

**Cecilia Aguerrebere** received the B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from Universidad de la República, Uruguay, in 2006, 2011 and 2014 respectively, the M.Sc. degree in applied mathematics from ENS Cachan, France, in 2011, and the Ph.D. degree in Signal and Image Processing from Télécom ParisTech, France, in 2014 (joint Ph.D. program with Universidad de la República, Uruguay). From August 2015, she is with the Electrical and Computer Engineering Department, Duke University, where she holds a Postdoctoral Research Associate position.