

Reconstructing Interlaced High-Dynamic-Range Video Using Joint Learning

Inchang Choi, Seung-Hwan Baek, and Min H. Kim, *Member, IEEE*

Abstract—For extending the dynamic range of video, it is a common practice to capture multiple frames sequentially with different exposures and combine them to extend the dynamic range of each video frame. However, this approach results in typical ghosting artifacts due to fast and complex motion in nature. As an alternative, video imaging with interlaced exposures has been introduced to extend the dynamic range. However, the interlaced approach has been hindered by jaggy artifacts and sensor noise, leading to concerns over image quality. In this paper, we propose a data-driven approach for jointly solving two specific problems of deinterlacing and denoising that arise in interlaced video imaging with different exposures. First, we solve the deinterlacing problem using joint dictionary learning via sparse coding. Since partial information of detail in differently exposed rows is often available via interlacing, we make use of the information to reconstruct details of the extended dynamic range from the interlaced video input. Second, we jointly solve the denoising problem by tailoring sparse coding to better handle additive noise in low-/high-exposure rows, and also adopt multiscale homography flow to temporal sequences for denoising. We anticipate that the proposed method will allow for concurrent capture of higher dynamic range video frames without suffering from ghosting artifacts. We demonstrate the advantages of our interlaced video imaging compared with the state-of-the-art high-dynamic-range video methods.

Index Terms—Image reconstruction, deinterlacing, denoising, high-dynamic-range video.

I. INTRODUCTION

HIGH-DYNAMIC-RANGE (HDR) imaging [1] has been broadly used in many state-of-the-art imaging applications [2], [3] to overcome the low-dynamic-range (LDR) limit of image sensors. The fundamental workflow of HDR imaging entails capturing different exposures of a scene and combining them into an HDR

radiance map [4], where a pixel value represents an estimate of scene radiance. A camera response function must be calibrated for conventional cameras unless a raw sensor readout is used [5]. Since HDR imaging requires multiple inputs with different exposures, various imaging hardware designs have been proposed for concurrent capture [6], [7]. These custom-built HDR imaging systems, such as Spheron VR and Arri Alexa, are still expensive for ordinary users. Many alternative algorithmic solutions for HDR imaging have been proposed to reduce temporal-mismatch artifacts, such as *ghosting* [8]–[12]. As addressed by Srikantha and Sidibe [13] and Tursun et al. [14], there is no universal HDR reconstruction algorithm that can remove *ghosting artifacts* or *motion blur* effectively enough for general dynamic scenes. Owing to ghosting artifacts, HDR video has received even less attention, aside from a few studies [15]–[17].

We propose a data-driven interlaced video reconstruction method for extending the dynamic range (DR) of video imaging, by taking advantage of the parallel imaging architecture in conventional cameras. Despite the advantages of capturing multiple exposures concurrently, interlace-based approaches suffer from severe *interlace artifacts* and *noise*, creating challenges in reconstructing high-quality video. Interlaced video imaging has been preliminarily exploited in a few works; however, few extensive investigations of the major problems of deinterlacing and denoising in interlaced video have been reported. Our method reconstructs an extended dynamic range of video by solving the aforementioned two problems: interlaced artifacts and noise, using *jointly learned dictionaries* of artifacts through a multiscale homography flow. We propose a data-driven approach based on dictionary learning. We make use of partially over/underexposed data to reconstruct details using jointly learned dictionaries. For denoising, it appears that additive sensor noise is a major problem in exposure-interlaced video. We tailor dictionary learning to better handle this type of noise, in addition to adopting a multiscale homography flow. Figure 1 provides examples of our joint deinterlacing and denoising. Figure 2 outlines a stepwise workflow of our interlaced video reconstruction strategy.

Our **contributions** are as follows:

- solving the deinterlacing problem via joint sparse coding to make use of partially available data, and
- tailoring sparse representation to better handle additive noise that arises in exposure-interlaced video.

II. BACKGROUND AND RELATED WORK

Extending the dynamic range of imaging, so-called HDR imaging, has been researched extensively in recent decades.

Manuscript received December 12, 2016; revised April 2, 2017 and May 18, 2017; accepted July 18, 2017. Date of publication July 24, 2017; date of current version August 21, 2017. This work was supported in part by Korea NRF under Grant 2016R1A2B2013031 and Grant 2013M3A6A6073718, in part by Samsung Electronics under Grant SRFC-IT1402-02, in part by Cross-Ministry Giga KOREA Project under Grant GK17P0200, and in part by Korea Creative Content Agency in Ministry of Culture, Sports and Tourism, and an ICT Research and Development Program of MSIP/IITP of Korea under Grant R7116-16-1035. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Keigo Hirakawa. (*Corresponding author: Min H. Kim.*)

The authors are with the Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: minhkim@kaist.ac.kr).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The material includes video results, as well as additional figures. The total size of the videos is 66.7 MB. Contact minhkim@kaist.ac.kr for further questions about this work.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2731211

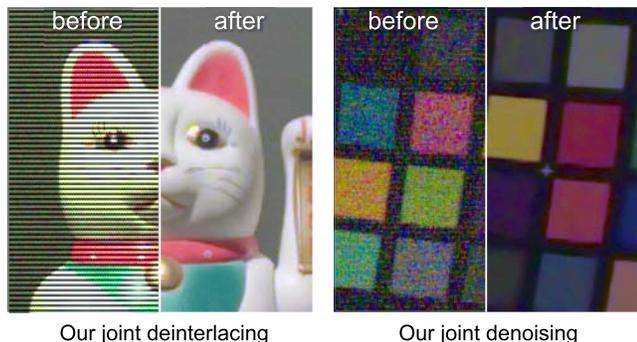


Fig. 1. An example of our extended dynamic range video reconstruction. The left image pair compares the raw interlaced input and the result of our jointly learned deinterlacing. The right image pair compares the noisy original image and the result of our joint denoising.

Most previous works concentrate on reconstructing HDR still images and tone-mapping. For brevity's sake, we refer readers to Reinhard et al. [18] for an overview of this subject. This section reviews HDR imaging only for dynamic scenes and HDR video.

A. HDR Imaging for Dynamic Scenes

1) *Ghosting Artifacts*: Since most HDR reconstruction methods calculate a weighted average of unsaturated readouts of different exposures [1], *ghosting* is inevitable in reconstructing HDR images of dynamic scenes. Many deghosting methods have been studied by aligning input images and reconstructing per-pixel radiance via motion flow [10], [11], [15], [16]. The motion flow is calculated by taking a per-pixel distance into account in an intensity mapping function [9], or a noise distribution of colors in different images [12]. Note that there is no universal HDR reconstruction method that can remove motion blur sufficiently well for general applications in real-life conditions [12]–[14]. The ideal solution for ghost-free HDR imaging of dynamic scenes is to capture different exposures with a consistent shutter time. To this end, gain-interlaced HDR imaging that captures different exposures concurrently is ideal for capturing dynamic scenes, as long as the two practical problems of interlace artifacts and noise are solved. This paper introduces a data-driven solution for these two problems, leading to ghost-free HDR video.

2) *Interlaced HDR*: Complementary-metal-oxide semiconductor (CMOS) image sensors are well known for parallelism of image signal processing features. They are capable of recording multiple exposures by row-wise interlacing. Gu et al. [19] introduced a method that removes a rolling shutter effect using *shutter time-interlaced* modulation for CMOS sensors. Heide et al. [20] developed a unified image-signal processing framework that includes HDR imaging. They proposed seminal time-interlaced HDR applications following Cho et al. [21]. In addition to shutter-time interlacing, Hajisharif et al. [22] proposed an application that records gain-interlaced readout, so-called Dual ISO, using open-source camera firmware, Magic Lantern [23] to reconstruct an HDR image. Many works in HDR imaging [22], [24], [25] have reported high gain amplification leading to severe additive

noise in the form of readout noise and analog-to-digital converter (ADC) noise. Although these prior works proposed interlaced HDR imaging, a few works have been devoted to thoroughly investigating specific major problems of *interlace artifacts* and *noise* that inevitably arise in interlaced HDR imaging.

3) *HDR Video*: Kang et al. [15] proposed an HDR video framework that acquires a time series of different exposures and merges the scene radiance following the optical flow to avoid ghosting artifacts. Based on this foundation, many descendant approaches are differentiated by using different methods of motion estimation. Mangiat and Gibson [26], [27] introduced a block-based motion estimation technique rather than a gradient-based optical flow. Kalantari et al. [17] proposed a patch-based synthesis method to enhance robustness in detecting motion flow. Recently, Gryaditskaya et al. [28] perceptually evaluated the importance of motion and dynamic range in a scene, and presented an HDR video algorithm known as motion-aware exposure bracketing. However, the optical flow calculation could fail when the scene includes fast motion or non-rigidity in the homography. Additionally, there is open-source firmware for Canon cameras [23] that enables HDR video capture by alternating analog gain frame-by-frame. However, an HDR reconstruction method is required to combine multiple exposures via a motion flow and removal of ghosting artifacts, leading to the risk of ghosting due to the asynchronous capture of multi-exposures. In addition to these software solutions, HDR video hardware systems have been proposed by Nayar and Branzoi [29], Tocci et al. [7], and Kronander et al. [30]. However, these custom-built imaging systems are still expensive for ordinary users. Alternatively, we utilize a concurrent capture setup based on interlacing, and propose an interlaced video reconstruction method that effectively removes interlace artifacts and noise, yielding a ghost-free video using a conventional camera.

4) *Sparse Coding for Image Processing*: An image signal can be described as sparse linear combinations of an over-complete dictionary that contains signal-atoms [31]. Many image restoration and denoising algorithms have been introduced in the last decade. These algorithms can be divided into two categories: those employing separate training images for learning the image prior and those using the corrupted image itself to learn the posterior [32]. The former class can be divided into parametric [33] and non-parametric approaches [34]. Li et al. [35] proposed the use of a joint sparse representation to estimate clean sound features from a noisy feature vector of speech data captured inside a vehicle. Yu et al. [36] also introduced a denoising approach that learns a dictionary from multiple image copies rather than a single image posterior. Recently, Serrano et al. [37] used sparse coding to generate HDR images by decoding single coded exposures. Yang et al. [38] proposed a super-resolution method using a joint dictionary learned from a pair of low- and high-resolution training datasets. We were inspired by this approach based on joint sparse representation and adapted it in our interlaced video imaging framework to enrich image quality by minimizing interlace artifacts and shot noise.

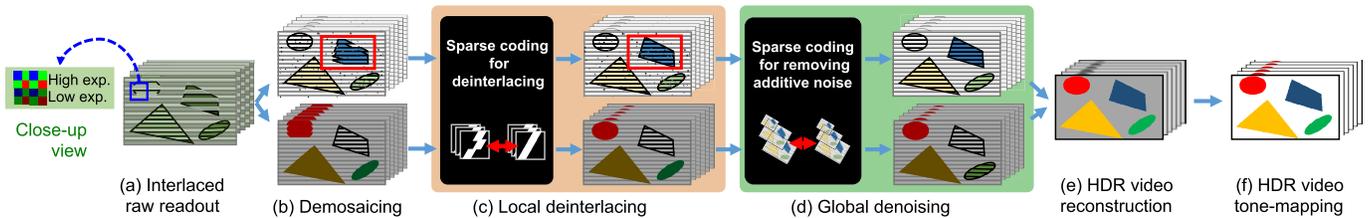


Fig. 2. Schematic diagram of our interlaced HDR video reconstruction strategy. (a) The input is interlaced readout for even and odd macro-rows. (b) It is separated and demosaiced to color images, which suffer from interlace artifacts and additive noise (Section IV-A). (c) We first remove interlace artifacts using a jointly-learned dictionary (Section IV-A). (d) We then remove additive sensor noise using another dictionary via multiscale homography flow (Section IV-B). (e) and (f) Finally, we reconstruct HDR video frames and tone-map them (Section IV-D).

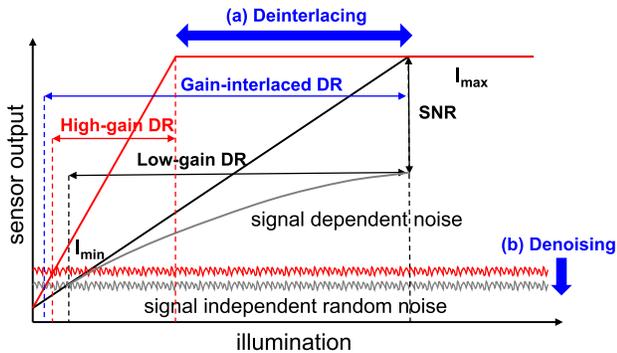


Fig. 3. Schematic diagram explains the theoretical derivation for our method to extend the dynamic range of the camera using our deinterlacing and denoising algorithm from interlaced readout. Adapted from Darmont [39].

III. EXTENDING THE DYNAMIC RANGE

The most common camera options for controlling exposure are aperture size, shutter time, and gain amplitude. Among them, the shutter time [19], [20] and the gain amplitude [22], [24], [25] have been preferred for extending the dynamic range for imaging. Recently, interlacing different exposures has been proposed for taking advantage of the CMOS architecture [19], [20]. Varying either the shutter time or the gain factor is available for interlaced imaging. Even though these approaches can be used to extend the dynamic range, there are several drawbacks. Shutter-interlaced imaging might cause motion blur of moving objects in long-exposure rows and might introduce more noise in short-exposure rows. Gain-interlaced imaging might cause severe readout noise in high-gain rows. Inspired by a previous experimental analysis on the impact of varying the gain and the shutter for HDR imaging by Hasinoff *et al.* [24], we were motivated to choose the gain factor for varying exposure with the objective of avoiding motion blur and ADC noise while extending the dynamic range; *i.e.*, we use the same amount of photons for each row when obtaining the exposure variations in interlaced rows.

A. Analog Gain and Dynamic Range

In this work we use a gain-interlacing approach to extend the dynamic range for video imaging. This section describes our insight regarding how to extend the dynamic range in video imaging through the analog gain control per row.

Figure 3 shows a schematic diagram that explains how to extend the dynamic range using gain interlacing. It compares

the dynamic range differences with two different setups of the low gain (the black line) and the high gain (the red line). The sloped black line indicates the sensor output to low-exposed illumination. The level of signal dependent noise (*e.g.*, photon shot noise) increases nonlinearly as illumination increases, following the output signal. The gray curved line indicates the level of noise. The ratio between the sensor output and signal dependent noise determines the signal-to-noise ratio (SNR). The intersection of signal independent random noise (such as dark-current noise and readout noise) and the sensor output determines the minimum level of the sensible intensity value I_{\min} . The maximum signal level before saturation determines the maximum camera output I_{\max} . From given illumination, the signal range between I_{\min} and I_{\max} determines the *dynamic range* of the camera.

When we control the gain amplitude ratio to make it higher, the slope of the sensor output becomes faster, as shown by the red line. The sensor output results in rapid saturation, proportional to the gain amplitude. In addition, the random noise level increases simultaneously when the gain amplitude increases. Note that the dynamic range is extended toward the lower level of illumination, as observed by Hasinoff *et al.* [24]. However, since the sensor signal is saturated quickly with a high gain, the imaging setup with a high gain results in a smaller dynamic range due to fast saturation.

B. Sensor Noise

The general HDR reconstruction process approximates irradiance from a scene that falls on a sensor by scaling up readouts along with noise. As addressed by Granados *et al.* [25] and Hasinoff *et al.* [24], it is necessary to handle the noise amplification issue for reconstructing high-dynamic-range images. The pixel intensity is proportional to the radiant power Φ of light energy that a sensor collects over an exposure time of t seconds. A pixel value is also proportional to the sensor's analog gain factor g . The raw readout value I varies between the minimum level of signals $\{\Phi t g + n_s g + n_r g + n_a\}$ and its saturation level I_{\max} , where n_s is signal dependent noise such as photon shot noise, n_r is signal independent random noise such as dark-current noise and readout noise, and n_a is gain independent noise such as the ADC noise and the additive noise of the amplifier itself [39]. Signal dependent noise n_s is dominant temporal noise and cannot be reduced by amplifying the pixel output with analog gain.

Therefore, we handle n_s using our temporal denoising algorithm (Section IV-C). We model signal dependent noise n_r as additive noise τ in Equation (6). We therefore alleviate the signal dependent noise n_r using jointly learned dictionaries, which are trained from synthetic random noise (Section IV-B). Note that the remaining term in the signal I , the gain independent noise n_a , is negligible when high amplification is configured so that the analog gain factor g takes a large value.

C. Our Approach

In this work, we attempt to take advantage of both the low and high gain configurations by introducing a data-driven algorithmic solution. For the low luminance level, we adopt an extended dynamic range of the low luminance signals by using the high gain as described above. For the high luminance level, we utilize the original DR in the low gain. Our algorithmic solution allows for *selective reconstruction* of valid signals from gain-interlaced signal outputs, yielding the maximum capacity of the analog circuit. Our algorithm achieves an extended dynamic range in two folds. First, the signal saturation problem in the interlaced higher exposure rows is solved by our *deinterlacing* algorithm via sparse representation (described in Section IV-A). Second, the increased level of random noise is solved by applying our *denoising* algorithm based on sparse coding (Section IV-B) and temporal denoising (Section IV-C). Since our objective is to extend the dynamic range of video imaging for capturing moving objects, we interlace the analog gain amplitude, rather than the shutter time. Therefore, the exposure time of every frame and every row is identical to the given camera setup.

In summary, our joint learning-based strategy allows us to maximize the capacity of the analog circuit by lowering the smallest detectable input signal I_{\min} , while maintaining the details of the largest non-saturating signal I_{\max} , which allows for ghost-free video imaging with an extended dynamic range.

IV. RECONSTRUCTING INTERLACED HDR VIDEO

Taking gain-interlaced raw video readouts as input, our video reconstruction method processes them in several steps, as shown in Figure 2. First, we separate the gain-interlaced raw readouts into two independent raw readouts by exposure. We then perform demosaicing on the two pixel groups of different gains to generate two color video frame groups of different exposures. In order to alleviate interlacing artifacts found in over-/under-exposed pixels, we apply local deinterlacing via sparse representation. Furthermore, we execute global denoising that includes spatial denoising based on sparse coding and temporal denoising based on multiscale homography flow, in order to address different levels of noise in each exposure induced by insufficient photons. Finally, we reconstruct video frames with the extended dynamic range using the deinterlaced and denoised video frames from the previous step, and we perform HDR video tone-mapping to generate a final output video in the sRGB color space. In the following subsections, we provide an in-depth description of each step of our interlaced video reconstruction method.

A. Deinterlacing via Sparse Representation

Since we use the same exposures in every two macro rows (a macro row of a Bayer pattern includes two pixel rows), interlaced HDR readouts capture images of the same exposure in half resolution along the vertical axis. Inspired by Yang et al. [40], we reconstruct two differently exposed images in the full resolution from interlaced readouts using a jointly learned dictionary. We train the joint dictionary from a pair of interlaced and non-interlaced readouts by adopting a joint dictionary learning technique. While Yang et al. focus on increasing image resolution using a jointly learned dictionary, we aim to interlace video frames of different exposures for extending the dynamic range of imaging. Different from the general super-resolution problem, partial information is often available for the missing rows in interlaced HDR. Our proposed method makes use of the partially over/underexposed data to enhance the super-resolution results while learning dictionaries.

1) *Sparse Representation*: A linear relationship between corrupted and clean images can be learned using natural image prior examples in a non-parametric manner. Let a column vector $\mathbf{x} \in \mathbb{R}^m$ be the signal that we want to decompose and let $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_p] \in \mathbb{R}^{m \times p}$ be a set of normalized basis vectors, an overcomplete dictionary of p atoms (a small subset of patches). We can represent \mathbf{x} by adapting \mathbf{D} using a column vector $\boldsymbol{\alpha} \in \mathbb{R}^p$, so-called *sparse code*, such that

$$\mathbf{x} \cong \mathbf{D}\boldsymbol{\alpha}. \quad (1)$$

Extending this idea toward our reconstruction problem, we can formulate a linear relationship $L \in \mathbb{R}^{m \times m}$ from the clean patch \mathbf{x} to the corrupted patch \mathbf{y} , using the sparse linear model:

$$\mathbf{y} \cong L\mathbf{x} = L\mathbf{D}\boldsymbol{\alpha}. \quad (2)$$

Assuming that $\boldsymbol{\alpha}$ is sufficiently sparse and L is a linear transformation, we can recover a clean image patch \mathbf{x} with respect to \mathbf{D} from the corrupted patch \mathbf{y} . Here, the corrupted patch refers to a patch that has interlacing artifacts.

2) *Joint Learning*: The sparse optimization yields two jointly-learned dictionaries $\hat{\mathbf{D}}$ for corrupted images and \mathbf{D} for clean ones that we use for removing the interlace artifacts and denoising. If the interlace artifacts are linearly transformed into a sufficiently small patch size, 6×6 , we can jointly optimize two different dictionaries of corrupted patches \mathbf{y}_i and clean patches \mathbf{x}_i by sharing common sparse code (coefficients):

$$\min_{\hat{\mathbf{D}}, \mathbf{D}, \boldsymbol{\alpha}} \sum_{i=1}^n \frac{1}{2} \|\mathbf{y}_i - \hat{\mathbf{D}}\boldsymbol{\alpha}_i\|_2^2 + \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda\psi(\boldsymbol{\alpha}_i), \quad (3)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{p \times n}$ is the shared sparse code; $\hat{\mathbf{D}}, \mathbf{D} \in \mathbb{R}^{m \times p}$ subject to $\forall j = 1, \dots, p, \|\mathbf{d}_j\| \leq 1$ are the dictionaries of 512 atoms for the corrupted and the clean patches, respectively; the number of patches n is set to 100,000 from natural image examples; λ is set to 0.15 for learning in our experiments.

We used a super-resolution dataset for \mathbf{x} provided by Yang et al. [40], and created a corrupted image set \mathbf{y} synthetically, using L for training $\hat{\mathbf{D}}$ jointly. Note that choosing $\psi(\cdot)$ is a selection problem that induces sparsity in $\boldsymbol{\alpha}$, traditionally

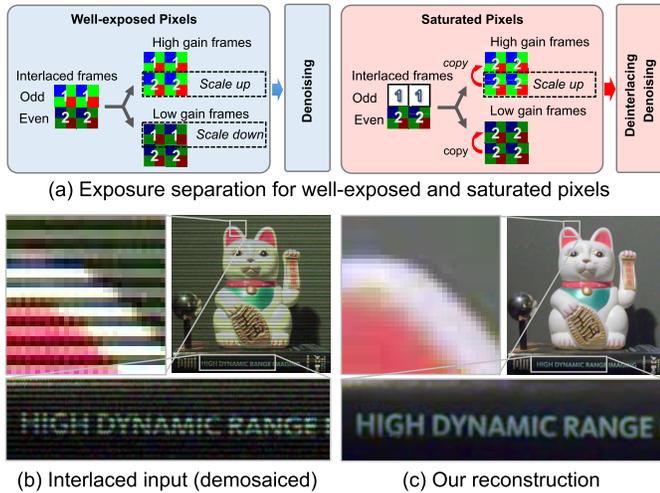


Fig. 4. (a) shows a schematic diagram of the exposure separation, prior to the joint-learning-based reconstructions that remove spatial artifacts. The saturated pixels are first substituted by neighboring rows via a weighted sum operator while unsaturated pixels with different exposures are scaled to match the exposure levels of unsaturated pixels. (b) shows demosaiced interlaced input, and (c) represents the final output of our reconstruction.

either the l_0 pseudo-norm $\|\alpha\|_0$ or the l_1 norm $\|\alpha\|_1$. We use the l_1 norm for training dictionaries and the l_0 norm for later reconstructing sparse coding using orthogonal matching pursuit (OMP) [31] with consideration of computational efficiency.

3) *Sparse Reconstruction*: For reconstructing a complete image, we apply sparse representation exclusively within the expanded boundary of the poorly exposed regions including over-exposed and under-exposed pixels. The expanded boundary is computed by dilating the poorly exposed regions with a disk, of which the radius is twice that of the patch size. We split the corrupted input image into overlapping patches. For each patch $\tilde{\mathbf{y}}$ we solve the following optimization problem to obtain a sparse code $\tilde{\alpha}$ using the jointly-learned corrupted dictionary $\tilde{\mathbf{D}}$ for a corrupted patch:

$$\min_{\tilde{\alpha}} \frac{1}{2} \|\tilde{\mathbf{y}} - \tilde{\mathbf{D}}\tilde{\alpha}\|_2^2 + \lambda\psi(\tilde{\alpha}). \quad (4)$$

Once we obtain the common sparse code $\tilde{\alpha}$, we reconstruct the clean patch $\tilde{\mathbf{x}}$ with sharpened detail by multiplying the clean joint dictionary \mathbf{D} with the common sparse code:

$$\tilde{\mathbf{x}} = \mathbf{D}\tilde{\alpha}. \quad (5)$$

4) *Learning With Partial Information*: Different from the general super-resolution problem, partial information from differently exposed rows is often available for the missing rows in interlaced HDR. In order to utilize partial information in sparse representation, we first classify interlaced macro-row signals into two groups, following Kang et al. [15], who segmented each frame into the well-exposed and poorly-exposed regions.

One group consists of well-exposed pixels, which we can easily reconstruct with the corresponding gain factors [see Figure 4(a)]; e.g., a frame of a high-gain factor (2^5 , ISO 3200) can be completed by copying and scaling the

pixel levels in the low-gain macro rows (2^0 , ISO 100) with the high-gain factor (2^5) and leaving the high-gain macro rows unchanged. The low-gain exposure can be completed in the opposite way.

The other group categorizes saturated pixels to be recovered via our sparse coding framework. Whereas the general super-resolution algorithms fill in missing information directly using a dictionary, we fill in saturated rows first using neighboring partial information by taking unsaturated pixel intensities of the differently exposed rows [Figures 4(a)]. The partial information is set up by copying the unsaturated pixels in neighboring macro-rows to the saturated pixels. This operation causes jaggy weaving artifacts, but we remove the artifact using the joint dictionary that has been trained with images corrupted by the same copying operation. The partial information coupled with the joint corrupted-clean dictionary pair has a significant impact on recovering original sharp edges while reconstructing interlaced readouts. Note that this locally applied operation makes it possible to save original details in interlaced frames without saturation, while we apply denoising globally across the entire image region.

B. Denoising via Sparse Representation

It appears that additive sensor noise is a major problem that degrades image quality along with interlace artifacts in interlaced HDR. As shown in the colorchecker in Figure 1, severe noise and color artifacts caused by the additive noise are unavoidable especially for the dark area in the scene. We therefore tailor sparse representation to better handle sensor noise for each frame. In addition to per-frame noise reduction, we also apply a multiscale homography flow to reduce temporal noise.

1) *Dictionary for Noise*: In most previous works, the dictionary for denoising is learned as a posterior in the input image itself such as K-SVD [31] and structural clustering such as BM3D [41]. Yu et al. [36] introduced a denoising approach that learns a dictionary from multiple image copies rather than a single image posterior. Recently, Singh et al. [42] proposed combining self patch-based denoising with super-resolution. We were inspired to extend our joint representation idea to denoising by modeling corrupted image patches \mathbf{y} as the sum of clean patches \mathbf{x} and additive noise τ :

$$\mathbf{y} \cong \mathbf{x} + \tau. \quad (6)$$

Note that this term τ is additive noise in linear sensor readout, which is different from synthetic Gaussian noise commonly used in testing denoising algorithms. Therefore, instead of simply adding Gaussian noise, we generate gain-interlaced noise by simulating the camera pipeline of our gain-interlacing camera. Given images in a training set, we first transform their colors to the linear sRGB space. The linear sRGB training images are then subsampled to constitute Bayer-patterned raw readouts. We then simulate Gaussian noise independently in four Bayer patterns and two interlaced-gain values. The standard deviations of these eight Gaussian noise have been estimated from real interlaced raw readouts using Liu et al. [43]. We multiplied the gain ratio to the standard deviations of low gain values to simulate noise in resulting

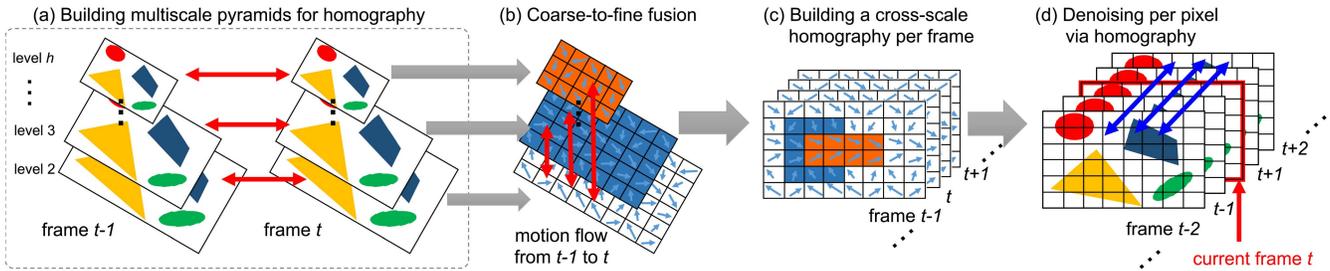


Fig. 5. Temporal denoising process using multiscale homography flow. We first build multiscale homography pyramids for consecutive frames by estimating homography between adjacent gaussian pyramids (a). We then perform coarse-to-fine refinement to merge each multiscale homography pyramid into a cross-scale homography flow map (b). The refinement is based on the accuracy of homography, computed by comparing the warped and the target frame. The cross-scale homography enables to register multiple consecutive frames (c). We perform temporal denoising by selectively filtering the aligned frames (d).

HDR images. After adding noise, we perform a demosaicing algorithm [44] to produce corrupted noisy RGB color images.

Using corrupted noisy and clean training images, we build a jointly-learned dictionary by pairing corrupted noisy image patches with clean patches of natural image examples. Note that we use the noise image patches for \mathbf{y}_i in Equation (3) for learning dictionaries. We have pre-trained two dictionaries for a pair of ISO 100/3200 and a pair of ISO 100/6400. Once we have determined the joint dictionary $\hat{\mathbf{D}}$ for noise, we obtain sparse code $\tilde{\alpha}$ from the new image using Equation (4). We then use the sparse code $\tilde{\alpha}$ to replace the joint dictionary $\hat{\mathbf{D}}$ with \mathbf{D} to reconstruct a noise-free frame \mathbf{x} using Equation (1).

C. Temporal Denoising

Since we target video applications, we can further improve our performance by applying temporal denoising. We were motivated to make use of temporal sequences. This additional step allows us to further reduce the noise level without introducing ghosting artifacts that arise in HDR video. In many HDR image and video algorithms, *motion flow* has been used to combine multiple exposures of moving scenes in various ways, e.g., detecting an optical flow [10], [15], computing hierarchical homography [15], and estimating patch-based motion flow [16], [17]. However, we found that these approaches often failed in estimating the motion flow through a hierarchical homography flow and optical flow [15], [45], resulting in ghosting artifacts. See Figure 9 for examples. We were therefore motivated to use the motion flow estimate exclusively for *denoising*, rather than reconstructing HDR video. We reconstruct each HDR video frame individually by exploiting concurrent capture based on gain-interlacing.

1) *Multiscale Homography Flow*: We compute a multiscale homography flow map per frame, following Liu et al. [45]. Motion flow calculation relies on detecting features that imply motion. However, we found that many feature detection algorithms fail with feature-less patches in a small sized image. We choose a *coarse-to-fine* approach [46], [47] in calculating the homography flow, different from the hierarchical approach [15], [45]. Figure 5 describes the workflow of temporal denoising. Our motion calculation begins by building a Gaussian pyramid of each frame, where the diameter of the smallest level h is no less than 400 pixels. For a pair of neighboring frames, we search for descriptors of feature points in each level using a corner-detection algorithm [48]

and a descriptor extraction method [49]. We then compute Hamming distances to match these descriptors and estimate a homography transformation of each level of the frame pair using RANSAC [50]. As our denoised image still contains some noise, we begin by calculating each homography matrix from the upper level of the coarsest input to the finest one in the Gaussian image pyramid.

2) *Coarse-to-Fine Refinement*: Once the multiscale homographies are built, we exploit the accuracy of homographies across scales using a coarse-to-fine approach to combine them to a multiscale homography matrix for a transition between frames. We translate the pixels at each level using the corresponding homography and compare the pixel values between the frames. If the similarity between registered pixels of the previous and the next frames is below the threshold (0.1% of the maximum pixel level) at the finest level, we select the homography estimate at the next coarse level of the pyramid and iterate testing the similarity of the next coarse level until it reaches the threshold. We iterate this operation from the finest to the coarsest level for all pixels with a group of nine neighboring frames, including the next four frames, and the current frame in the middle. This multiscale homography flow calculation allows us to compute the motion flow more robustly. Unlike hierarchical homography graphs [15], [45], we chose a global multiscale homography in calculating the motion flow as we observed that child homographies with few features often fail due to lack of rigidity.

3) *Temporal Denoising*: The multiscale homography per frame transition allows us to evaluate the pixel values along the time line for denoising. We register all the pixels via the motion flow using the homography and compare the signal levels of each pixel. For each pixel in all frames, we first compute the mean value of pixels lying along its motion flow path. In this process, only those pixels with an intensity difference to the target below a threshold value are counted. The scene dependent threshold is empirically adjusted. The threshold level is varied from 0.3–4.0% in our experiments. The selective mean value is then assigned as the intensity of the target pixel. This operation can be executed either before or after HDR reconstruction.

D. HDR Reconstruction and Tone-Mapping

Once we deinterlace and denoise all frames of different exposures, we combine all frames of different exposures into



Fig. 6. A hardware system prototype with an SK-Hynix mobile image sensor.

an HDR video frame using a common HDR reconstruction formula [1]:

$$\mathbf{L}_{ij} = \left\{ \sum_{k=1}^2 \omega(\mathbf{X}_{ijk}) \mathbf{X}_{ijk} / \mathbf{g}_k \right\} / \left\{ \sum_{k=1}^2 \omega(\mathbf{X}_{ijk}) \right\},$$

where \mathbf{L}_{ij} is the HDR radiance pixel at pixel location (i, j) ; \mathbf{X}_{ijk} is the deinterlaced and denoised signal with a gain factor \mathbf{g}_k in an exposure level k . $\omega()$ is a pyramid weighting function that accounts for under- and over-saturation levels of pixels. Note that our current HDR video method does not require any deghosting while reconstructing HDR video. The color transformation matrix from raw RGB to sRGB signals and white balancing factors are multiplied after reconstructing HDR to reduce quantization.

1) *HDR Tone-Mapping*: An HDR tone-mapping operator is necessary for displaying the reconstructed HDR video frames on an ordinary LDR display, as the dynamic range of the reconstructed HDR video frames is higher than that of the display [51], [52]. We implemented and compared two HDR tone-mapping operators applicable for HDR video. One is a global tone-mapping operator proposed by Kim et al. [53], [54]; the other is a local-temporal tone-mapping operator by Aydin et al. [55]. Both operators produced high-quality frames with comparable quality. We found that the jointly operated tone-mapping operator [55] performs stronger compression of the dynamic range than the global operator. We mostly use Aydin's operator for high-dynamic-range scenes, while we employ Kim's operator for high fidelity of color reproduction. Since we do not have rapid luminance changes in captured scenes, the adaptive parameters in the operators are determined from the first frame of the scene and used for the remaining frames to avoid flickering.

V. RESULTS

We implemented our interlaced HDR video algorithm using two conventional cameras: a custom-built mobile camera module (2.4 megapixels) with an SK-Hynix Hi231 image sensor (see Figure 6) and a Canon EOS 5D Mark III camera of 22.3 megapixels, equipped with the Dual-ISO module of an open-source camera firmware, Magic Lantern [23], since it allows analog-gain controls on commercial DSLR cameras. odd macro-rows were set to a 16 or 64 times higher gain factor (ISO 1600 or 3200) adaptively, depending on the scene's dynamic range, i.e., manually changing the gain setting considering the dynamic range of the scenes.

A. Validation of Gain Interlacing

In order to validate our approach that extends the dynamic range via gain-interlacing, we built an HDR characterization setup that consists of two transparent color targets and neutral density filters, following Kim and Kautz [5]. As shown in Figure 7, two photographically enlarged IT8.7/1 transparent color charts (576 patches in total) are placed on a diffuse light box in a dark room. The bottom one is covered with four sheets of two-stop neutral density filters (256 times less light comes through the filters). The dynamic range is calculated by a logarithmic ratio of the largest luminance and the smallest detectable luminance: $20 \log_{10} \frac{L_{\max}}{L_{\min}}$ [dB]. This experimental setup exhibits a dynamic range of 99.57 dB (16.54 f -stops) overall (max. 5708 cd/m^2 ; min. 0.06 cd/m^2 across both gray-scales). The dynamic range of this scene is close to the dynamic range of the human eye, approximately 100 dB [56]. Typical ordinary imaging sensors can capture a dynamic range from 60 to 70 dB through a single shot in general [57]. As shown in Figures 7(a) and (b), ordinary capture cannot cover the full dynamic range due to the lack of sensitivity and the well capacity of the photodiode, respectively.

These charts were captured by a Canon EOS 5D Mark III camera (equipped with a 50 mm lens, the aperture of which is set to $f/4.5$) and a spectroradiometer, Jeti Specbos 1200 (calibrated luminance accuracy: ± 0.05 at 1000 cd/m^2 ; xy chromaticity repeatability: ± 0.0005 [58]) simultaneously. Each side of the captured HDR images is tone-mapped manually side by side with respect to its exposure level in order to visualize the captured signals.

LDR images (a) and (b) are separated from the interlaced readout of ISO100 and 6400, equipped with the Dual-ISO module [23]. Both images present saturation in over/under-exposed areas. Image (c) is a one-shot HDR image reconstructed from the interlaced readout using the proposed method. Images (d) and (e) are two traditional HDR images captured using two-shot LDR input images by varying the ISO (100 and 6400) and the shutter time (1/5 and 1/320), respectively. These three HDR images (c), (d) and (e) capture the full dynamic range of 99.57 dB in the setup without saturation.

B. Validation of HDR Video Reconstruction

1) *Sharpness vs. Noise*: We first compare our interlaced HDR reconstruction with three state-of-the-art interlaced HDR methods proposed by Hajisharif et al. [22], Heide et al. [20] and Cho et al. [21]. As shown in Figure 8, our proposed method outperforms these latest interlaced HDR methods in terms of sharpness and noise. Hajisharif et al. reconstruct missing exposure and color information assuming the local color distribution changes smoothly in a Gaussian window. When they use a narrow Gaussian window, edges become sharper, sacrificing noise reduction and vice versa. Heide et al. integrate all the image signal processes for camera into a single optimization problem, based on multiple regularization priors such as total variation [59], BM3D denoising [41], and cross-channel gradient correlation. When they increase

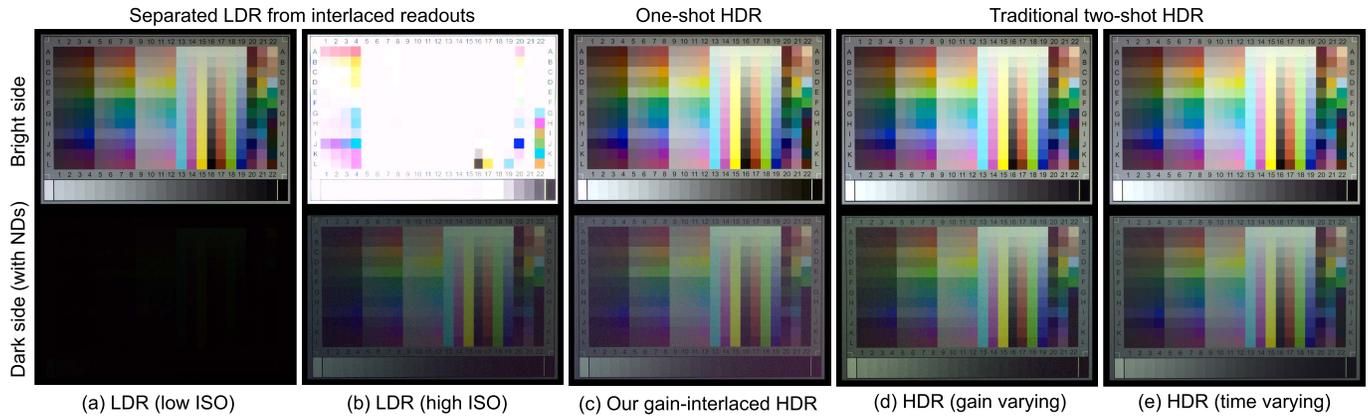


Fig. 7. Two photographically enlarged IT 8.7/1 transparent charts are placed on a diffuse light box in a dark room, where the bottom one is covered with four sheets of two f -stops neutral density filters (8 f -stops less light comes through the films. Overall dynamic range: 99.57 dB, 16.54 f -stops). These patches were captured by a Canon EOS 5D camera and a spectroradiometer (Jeti Specbos 1200). Note that captured HDR images are manually tone-mapped side by side with respect to their exposure levels for visualization. (a) and (b) are LDR input images separated from the interlaced readout of ISO 100 and 6400, equipped with the Dual-ISO module [23]. (c) shows our interlaced HDR image reconstructed from the interlaced readout, compared with traditional HDR images captured by varying the shutter time and ISOs, respectively. (d) and (e) compare one-shot interlaced HDR with traditional HDR images reconstructed from two-shot images by varying the ISO and the shutter time, respectively. Three HDR images (c), (d) and (e) capture the full dynamic range (99.57 dB) of the setup without suffering from saturation, compared to the saturated LDR input signals (a) and (b).

Sharpness/noise comparison in interlaced HDR

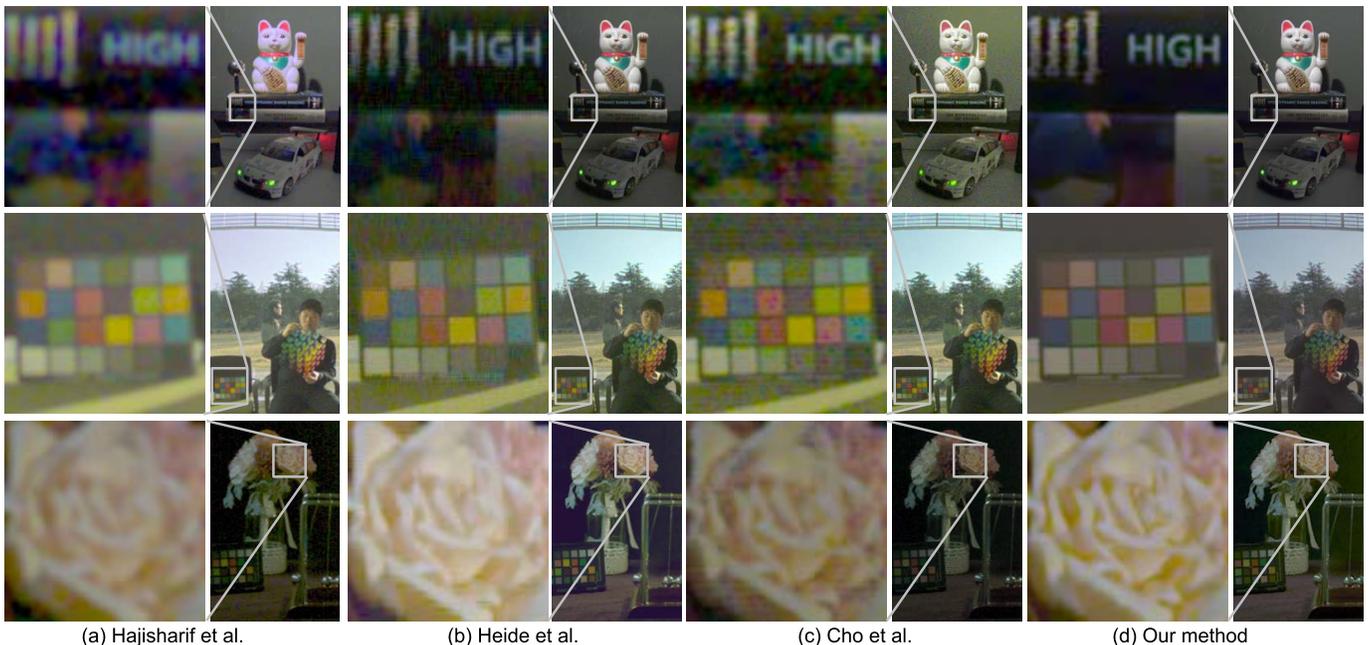


Fig. 8. We compare our interlaced HDR video method with three interlaced HDR video methods, captured by two different cameras (the first and the second row: the SK-Hynix mobile camera and the bottom row: Canon 5D Mark III camera with Magic Lantern). Hajisharif et al. [22] interpolate missing information of multiple exposures smoothly using Gaussian assumption, which results in blurred edges. Heide et al. [20] employ the BM3D prior to reduce noise, and missing information is linearly interpolated. Cho et al. [21] deal with interlacing artifacts and noise by applying bilateral filters. These three methods show a trade-off between denoising and deinterlacing along sharp edges in images, while our proposed method solves this trade-off and achieves clearer video with significantly less noise than prior works.

noise reduction, sharp edges are sacrificed and vice versa. Cho et al. alleviate interlacing artifacts by bilateral filter-like interpolation. They also handle the noise problem with a bilateral filter. Figure 8(c) shows that the gain-interlaced noise cannot be effectively removed, as the interlacing artifacts are often recognized as edge structures in bilateral filtering. These three methods present a tradeoff between denoising

and deinterlacing along sharp edges, while our proposed method preserves sharp edges without suffering from noise. In contrast, our proposed method exploits two different dictionary pairs separately, where they contain external priors from natural images. This allows effective removal of additive noise while preserving sharp edges. Refer to the supplemental video for more results.

Ghosting artifact comparison in HDR video

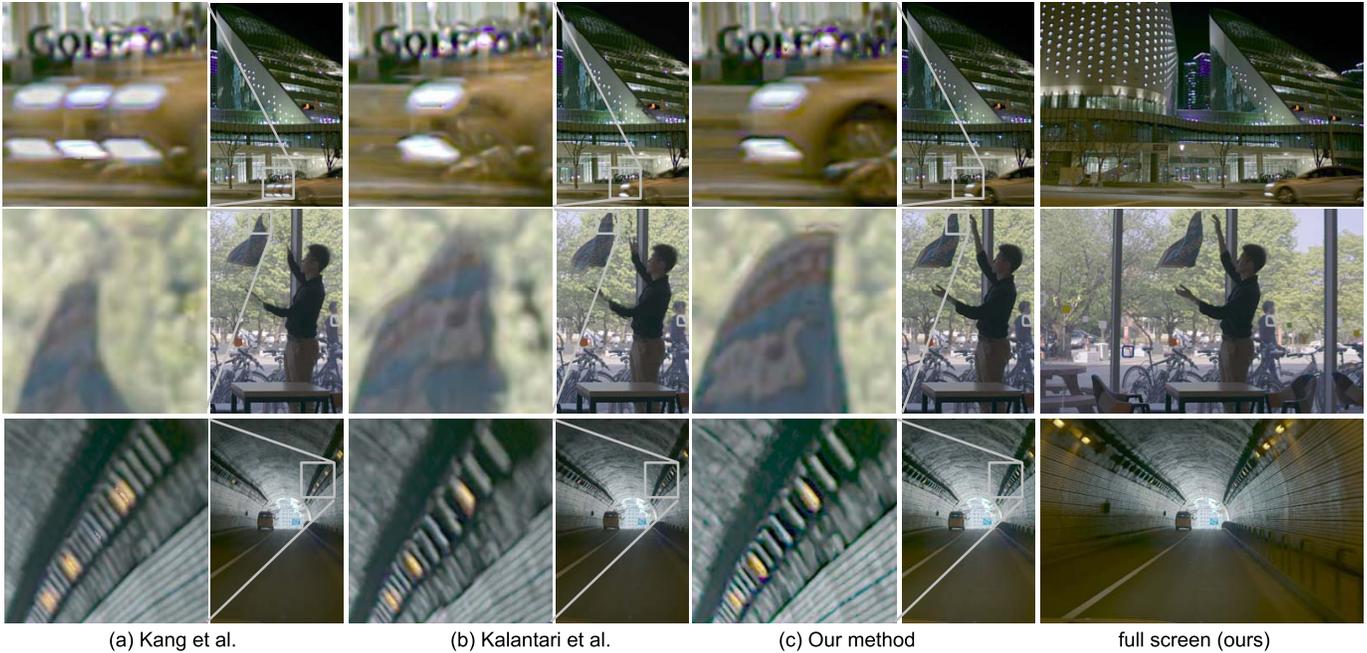


Fig. 9. We compare our interlaced HDR video method with two other temporally-modulated HDR video methods captured by a Canon 5D Mark III with Magic Lantern. Kang et al.’s method is based on homography computation, where local motion often fails in contrast to global motion, resulting in ghosting artifacts. Kalantari et al.’s method is based on patch-based reconstruction of multiple exposures across time, but still suffers from jittering and minor ghosting artifacts. Since jittering is invisible in paper representation, refer to the supplemental video for this artifact. Our method uses concurrent capture using interlaced exposures. There is marginal motion blur in our method caused by a long shutter time to capture night scenes but no ghosting.

2) *Ghosting Artifacts*: We also compared our method with two state-of-the-art HDR video methods proposed by Kang et al. [15] and Kalantari et al. [17], shown in Figure 9. The two compared methods are based on temporal modulation of multiple exposures across time. Since dynamic scenes are being captured, it is necessary to register corresponding multi-exposure pairs in both methods. The hierarchical homography proposed by Kang et al. handles large global motion well; however, this method is incapable of handling small motion that is sparsely distributed in scenes, and when scenes are too dark or too bright, estimating homography fails due to limited feature points. This results in flickering or ghosting artifacts. The patch-based reconstruction approach proposed by Kalantari et al. produces intermediate sequences of multi-exposures across time. They also find corresponding patches between neighboring frames via optical flow, which often fails in scenes with complex or fast dynamic motion. Even though their method can reduce ghosting artifacts effectively, temporal inconsistency of patch transformation results in jittering artifacts in video. Refer to the supplemental video for this artifact. In contrast, our method uses concurrently captured interlace input of multiple exposures. It presents only marginal motion blur for shutter time and no ghosting artifacts. For a fair comparison, we separate and convert interlaced and denoised input to sequences of long and short exposures to simulate temporally-modulated input for other HDR video methods.

3) *Motion Estimation*: We validate our motion estimation method, which is based on a multiscale homography flow, compared to Kalantari et al. [17] and Liu et al. [45]. As shown in Figure 10(c), our motion estimate method with temporal denoising can capture a fast moving object without

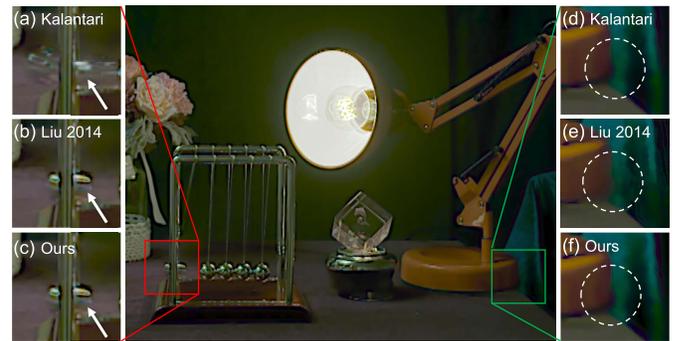


Fig. 10. We compare the performance of our motion estimation algorithm (c) & (f) with Kalantari et al. [17] (a) & (d) and Liu et al. [45] (b) & (e). The patch matching-based motion estimation (a) cannot track the fast movement of a pendulum. As shown in the dotted circle of (e), the hierarchical homography flow blurred an edge of static objects due to wrong motion estimation.

motion artifacts. In contrast, the patch matching-based motion estimation of Kalantari et al. fails to describe the motion of a moving pendulum in (a). Although the hierarchical homography flow proposed by Liu et al. processes the pendulum correctly, as shown in (b), it produces blur artifacts as shown in the circled area in (e). The artifacts show potential failure cases of the homography flow estimation.

4) *Denoising*: Figure 11 compares the real additive noise level of three different steps in our HDR video frame captured by a mobile image sensor. Column (a) shows a naïve HDR video frame without any denoising. This video suffers from severe additive noise due to high gain amplitude. Column (b) shows the result of our spatial denoising step via the jointly-learned dictionary. The PSNR of the noise removal

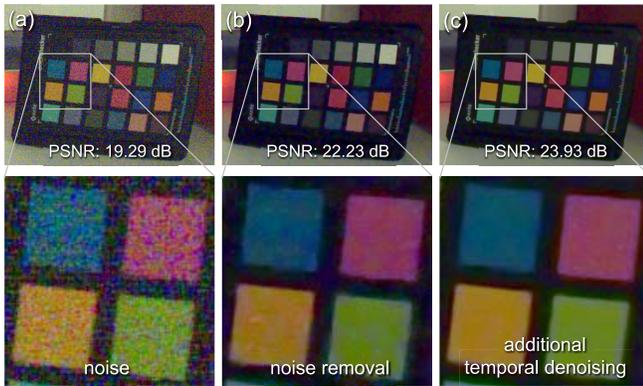


Fig. 11. Our step-wise noise removal. (a) shows a noisy image obtained after deinterlacing. (b) presents the result of our additive noise removal of a frame via sparse coding. (c) includes additional noise reduction for video using nine adjacent frames. (d) is the ground truth generated by averaging 360 consecutive frames of the static region of the cropped color checker in the upper row.

increases from 19.29 dB to 22.23 dB. Column (c) presents the result of additionally applying temporal denoising, through which the PSNR further increases to 23.93 dB. With a real camera, a noise-free image cannot be captured. Hence, the ground truth image (d) is computed by averaging 360 frames of still captures.

5) *Performance*: We reconstruct video clips using a machine that has an Intel i7-3770 CPU 3.40 GHz with 32 GB memory. It took 2.82 minutes for our method to process a frame including deinterlacing and denoising. It took 18.33 minutes for Hajisharif et al., 9.26 minutes for Heide et al., 16.89 minutes for Kalantari et al. and 1.07 minutes for Kang et al. We used non-optimized Matlab codes for all the methods. The implementation of Kalantari et al. is publicly available, provided by the authors, and the other methods are implemented by ourselves.

C. Deinterlacing

In addition to evaluation of HDR video, we validate the performance of our individual algorithm blocks, deinterlacing and denoising. For deinterlacing, we compare our method with four different super-resolution algorithms: direct copying of nearest-neighbor rows as naïve deinterlacing, bicubic interpolation, a self similarity-based super-resolution method [60], and a joint-learning super-resolution method [38], computing average PSNR and structural similarity (SSIM) [61] over all 24 standard images of the Kodak PhotoCD dataset [62]. From the reference Kodak RGB images, we created Bayer-patterned raw images and subsampled odd macro-rows to synthesize interlaced images, which is identical to our imaging configuration. The task is to reconstruct a full resolution from the subsampled readout.

The naïve interpolation of direct copying in column (a) in Figure 12 suffers from severe weaving artifacts with low PSNR and SSIM scores. The bicubic interpolation method in column (b) presents severe blurriness. Two sparse coding-based methods proposed by Peleg and Elad [60] and Yang et al. [38] show similar performance. Peleg and Elad's method is based on self-similarity. However, their method fails to detect similar examples in an interlaced image. Yang et al.'s

method customizes example priors to better handle upsampling while neglecting interlace artifacts. Our method reconstructs interlaced rows successfully, outperforming the other methods that include two state-of-the-art super-resolution methods.

D. Removing Noise

1) *Synthetic Noise*: Denoising has been well studied during the last decade. We compare our noise removal against state-of-the-art denoising methods: BM3D [41], non-local means (NLM) denoising [32], and total variation l_1 (TVL1) denoising [59]. Following the evaluation protocol in previous studies of denoising algorithms, we evaluated our denoising performance. We first generate synthetic gain interlaced noise with different levels in each gain with the reference Kodak image dataset, as described in Section IV-B. The gain interlaced noise was generated with a Gaussian distribution of which the standard deviations are $[1.80, 2.00, 1.00] \times 10^{-3}$ for high gain RGB pixels and $[1.50, 1.50, 1.50] \times 10^{-3}$ for low gain RGB pixels, respectively. For a quantitative evaluation of denoising, the generated noisy images are subjected to denoising algorithms including ours, and after denoising they are compared with clean reference images in terms of PSNR and SSIM metrics. Note that the parameters for the denoising algorithms that we used are optimized for the highest PSNR for a fair comparison.

The overall PSNR performance of BM3D (aver. 31.29 dB) is better than that of our denoising algorithm (aver. 29.59 dB) in this experiment. The SSIM performance of BM3D (aver. 0.92) and ours (aver. 0.91) is competitive. Our noise removal slightly outperforms NLM (aver. 29.18 dB and 0.89) and TVL1 (aver. 30.15 dB and 0.88) in terms of PSNR and SSIM. Refer to Figure 13. In this experiment with synthetic interlaced Gaussian noise, the BM3D method outperforms the others in general. However, as shown in Figure 13, no denoising algorithm in the comparison group results in PSNR and SSIM values that are as high as those yielded by our denoising method. In particular, our denoising method is relatively effective in preserving image structures while removing noise artifacts. The reason *for this* is that the algorithms such as NLM and BM3D, designed to take advantage of non-locality, often fail to find good enough sample patches to extract common structures without noise.

2) *Real Noise in Video*: For synthetic noise, we found that all of these algorithms perform best when the window parameter is closest to the distribution σ that is used to create synthetic Gaussian noise. However, the real noise distribution in a digital camera is unknown in general, and hence it is practically difficult to calibrate the noise distribution due to various noise sources in the sensor. The performance of the algorithms therefore could be different with real noise of a camera in all real-world conditions.

When we learn a joint dictionary for noise, we model the noise term τ as linearly additive noise in raw sensor signals, rather than synthetic Gaussian noise on gamma-corrected sRGB images. We convert the training sRGB images to linear sensor readouts with linearly additive noise. We found that the performance of our HDR video denoising method is superior to that of both BM3D and VBM3D

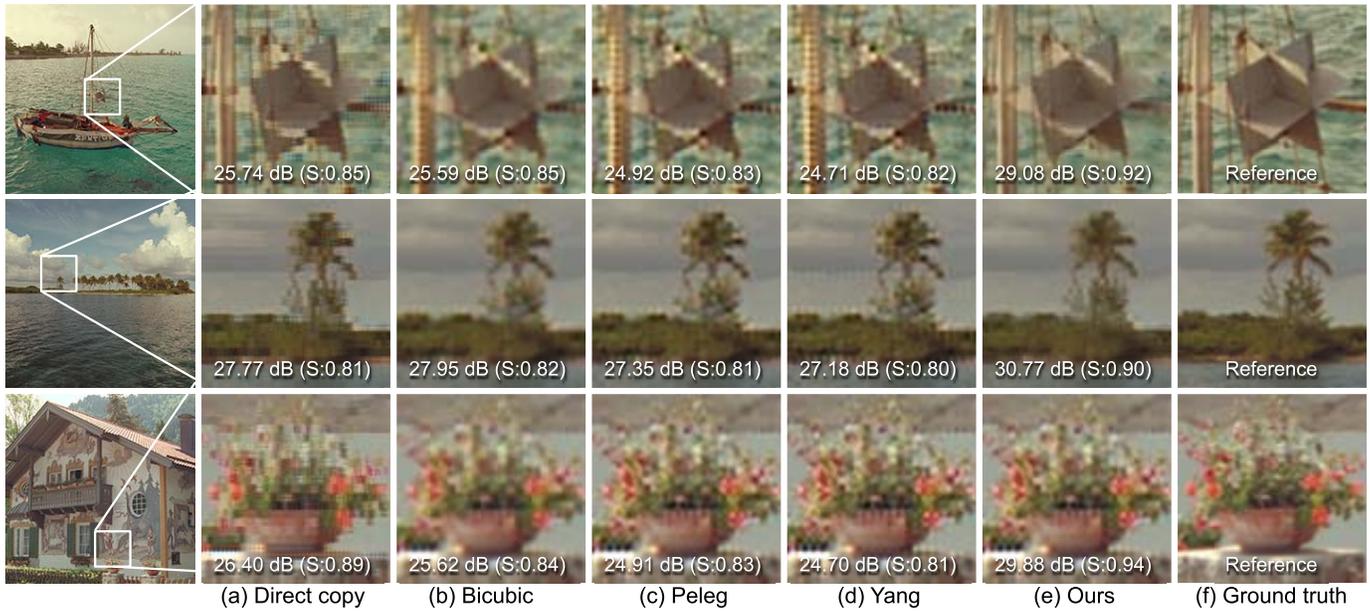


Fig. 12. We compare our deinterlacing against other relevant methods by measuring the PSNR to the reference, in addition to SSIM (in parentheses): directly copying neighboring rows (naïve interlacing), bicubic interpolation, a self-training sparse-coding method [60], and a joint-learning approach [38]. The PSNR and SSIM values are computed for the entire area of each image. Refer to the supplemental material for more examples.

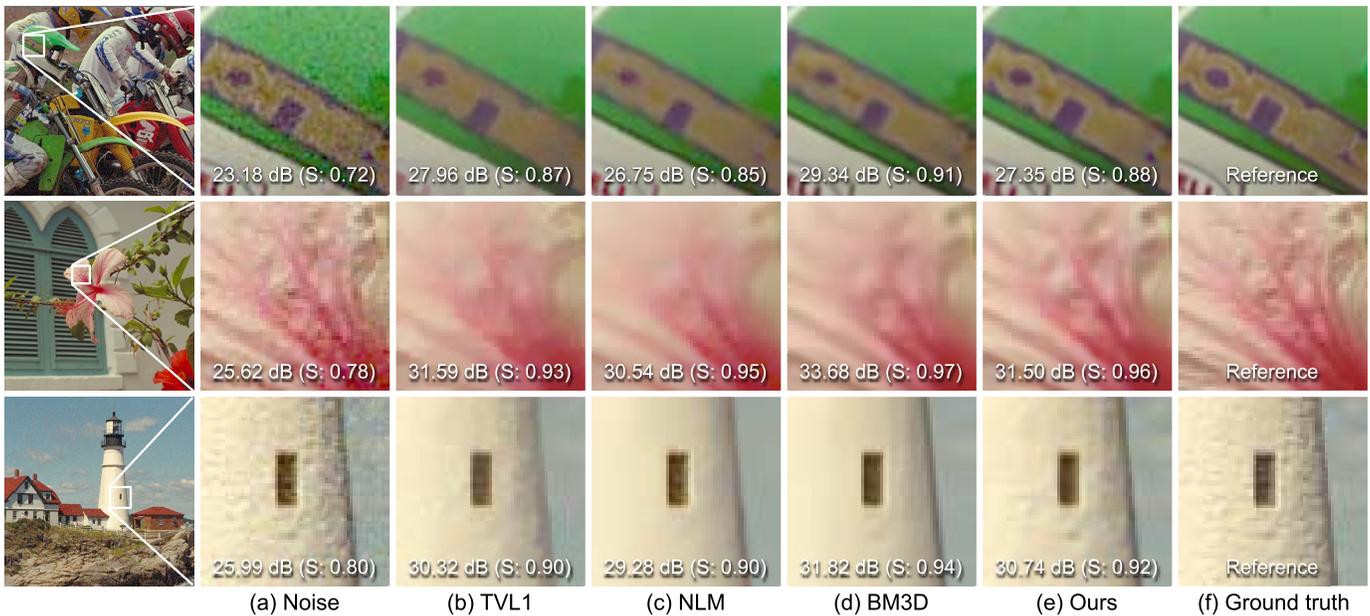


Fig. 13. We present examples of additive gain interlaced noise removal in the Kodak dataset by our method (without temporal denoising), compared with total variation l_1 denoising [59], NLM denoising [32] and BM3D [41]. The gain interlaced noise in (a) was generated with Gaussian distributions, of which standard deviations are $[2.95, 3.26, 1.60] \times 10^{-3}$ for high gain RGB pixels and $[2.43, 2.43, 2.45] \times 10^{-3}$ for low gain RGB pixels, respectively. Parameters for each method were optimized for the highest PSNR for a fair comparison.

(video versions of BM3D [41]) as well as NLM and multi-channel TVL1 denoising. Figure 14 presents video denoising examples from the SK mobile camera (2.4MP), which has a higher level of noise than the Canon DSLR camera. For a noise-free reference image from a real camera, we capture video footage (360 frames) of a static scene and average the frames as a single frame. Note that our method cannot model the real complex distribution of sensor noise in such various aspects. More physically accurate noise modeling via sparse coding remains as future work.

VI. DISCUSSION

Our joint dictionaries learn not only the natural image prior of clean images patches, but also simultaneously learn the representation of corrupted image patches. Furthermore, since they are trained jointly, the mapping between clean and corrupted image patches is learned as well. Based on this insight, our interlaced HDR video reconstruction can outperform state-of-the-art reconstruction approaches, as demonstrated in Section V. In contrast, other reconstruction methods compared in our experiments either use priors

Real noise removal comparison in video

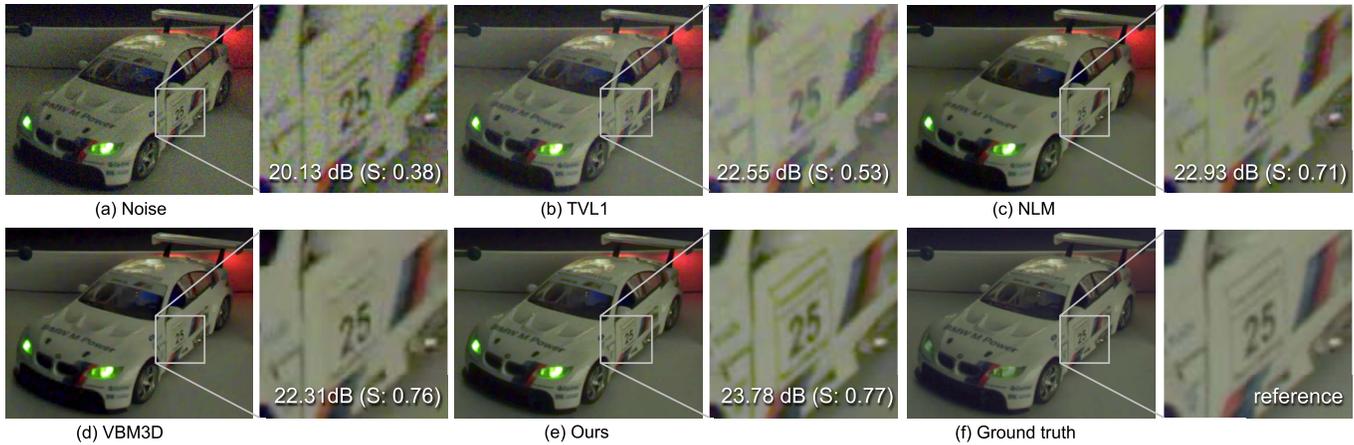


Fig. 14. For real additive noise removal, we compare the performance of our complete additive noise removal with the state-of-the-art video denoising methods: (a) noisy raw input, (b) total variation l_1 denoising for video, (c) NLM denoising [32], (d) Video BM3D [41] and (e) multi-channel NLM. Our complete noise removal handles real additive noise better, outperforming VBM3D. Here, the ground truth image (f) of an SK mobile camera (2MP) was made by averaging static 360 frames in the same scene.

from input images, or are based on global optimization methods.

We capture two different exposures using different gain amplitudes for each macro-row. This allows us to maximize the capacity of the analog circuit of the camera through sparse coding. We diversify the interlaced exposures up to six f -stops (e.g., ISO 100 and 6400). The degree to which we can extend the dynamic range depends on the maximum capacity of the analog circuit in the sensor. In contrast, the maximum dynamic range of traditional HDR video that varies exposure time row by row or frame by frame is practically limited by *motion blur* in the longer exposure rows or frames. For instance, Kalantari et al. [17] use two or three frames as input with exposure intervals of three f -stops to build an HDR video frame, while capturing thirty frames per second. While both gain-interlaced and time-varied HDR video methods present similar extendability of the dynamic range in capturing HDR video in practice, the proposed gain-interlaced approach does not suffer from ghosting artifacts or motion blur.

The recent success of deep convolutional neural networks, CNNs, included impressive achievements in various image processing and computer vision problems [63]–[65]. Since CNNs are capable of learning complex nonlinear representation of observation, which sparse coding is incapable of handling, we expect that CNNs would be beneficial for interlaced HDR video reconstruction. We leave this as future work to train deep CNNs that reconstruct HDR video from interlaced sensor readouts in an end-to-end manner.

VII. CONCLUSION

In interlace-based HDR imaging, two specific problems of interlace artifacts and additive sensor noises arise, degrading video quality significantly despite the benefit of concurrent capture in preventing ghosting artifacts. We sought to develop a learning-based solution that focuses on these two specific problems by making use of partial information of under-/over-exposure in interlaced HDR input. We also tailored sparse

representation to better handle additive noise via multiscale homography flow in an interlaced HDR video stream. The proposed method allows users to obtain interlaced HDR video using a conventional camera. Gain-controlled interlacing will become widely available soon. We believe that the proposed method would be beneficial to users who do not have access to an expensive HDR video system, allowing for high-quality HDR video using a conventional camera.

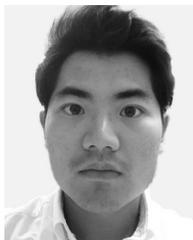
ACKNOWLEDGMENT

The authors would like to thank Xin Tong, Steven Lin (MSR), Seung-Tak Ryu, Jong Ho Park, and Incheol Kim (KAIST) for generous help.

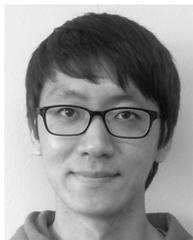
REFERENCES

- [1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. Conf. Comput. Graph. Interact. Techn.*, 1997, pp. 369–378.
- [2] G. Nam, J. H. Lee, H. Wu, D. Gutierrez, and M. H. Kim, "Simultaneous acquisition of microscale reflectance and normals," *ACM Trans. Graph.*, vol. 35, pp. 185:1–185:11, no. 6, 2016.
- [3] G. Nam, H. Lee, S. Oh, and M. H. Kim, "Measuring color defects in flat panel displays using HDR imaging and appearance modeling," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 2, pp. 297–304, Feb. 2016.
- [4] H. Zhao, B. Shi, C. Fernandez-Cull, S.-K. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *Proc. ICCP*, Apr. 2015, pp. 1–10.
- [5] M. H. Kim and J. Kautz, "Characterization for high dynamic range imaging," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 691–697, 2008.
- [6] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE CVPR*, Jun. 2000, pp. 472–479.
- [7] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," in *Proc. ACM SIGGRAPH*, 2011, p. 41.
- [8] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles, "Generation of HDR images in non-static conditions based on gradient fusion," in *Proc. VISAPP*, 2012, pp. 31–37.
- [9] O. Gallo, N. Gelfandz, M. Tico, W.-C. Chen, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proc. IEEE ICCP*, Apr. 2009, pp. 1–7.
- [10] 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, 2011.
- [11] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *Proc. ECCV*, 2012, pp. 499–512.

- [12] M. Granados, K. I. Kim, J. Tompkin, and C. Theobalt, "Automatic noise modeling for ghost-free HDR reconstruction," *ACM Trans. Graph.*, vol. 32, no. 6, p. 201, 2013.
- [13] A. Srikantha and D. Sidibé, "Ghost detection and removal for high dynamic range images: Recent advances," *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 650–662, 2012.
- [14] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, "State of the art reports: The state of the art in HDR deghosting: A survey and evaluation," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 683–707, 2015.
- [15] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.
- [16] 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, p. 202, 2012.
- [17] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen, "Patch-based high dynamic range video," *ACM Trans. Graph.*, vol. 32, no. 6, p. 202, 2013.
- [18] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging*, 2nd ed. San Mateo, CA, USA: Morgan Kaufmann, 2010.
- [19] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *Proc. IEEE ICCP*, Mar. 2010, pp. 1–8.
- [20] F. Heide *et al.*, "FlexISP: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, no. 6, p. 231, 2014.
- [21] H. Cho, S. J. Kim, and S. Lee, "Single-shot high dynamic range imaging using coded electronic shutter," *Comput. Graph. Forum*, vol. 33, no. 7, pp. 329–338, 2014.
- [22] S. Hajsharif, J. Kronander, and J. Unger, "HDR reconstruction for alternating gain (ISO) sensor readout," in *Proc. Eurographics*, 2014, pp. 025–028.
- [23] M. Lantern. (2014). *An Open-Source Camera Firmware for Canon DSLR Cameras*. [Online]. Available: <http://www.magiclantern.fm/>
- [24] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in *Proc. IEEE CVPR*, Jun. 2010, pp. 553–560.
- [25] 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 CVPR*, 2010, pp. 13–18.
- [26] S. Mangiat and J. Gibson, "High dynamic range video with ghost removal," *Proc. SPIE*, vol. 7798, p. 779812, Aug. 2010.
- [27] S. Mangiat and J. Gibson, "Spatially adaptive filtering for registration artifact removal in HDR video," in *Proc. IEEE ICIP*, Sep. 2011, pp. 1317–1320.
- [28] Y. Gryaditskaya, T. Pouli, E. Reinhard, K. Myszkowski, and H.-P. Seidel, "Motion aware exposure bracketing for HDR video," *Comput. Graph. Forum*, vol. 34, no. 4, pp. 119–130, 2015.
- [29] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *Proc. IEEE ICCV*, Oct. 2003, pp. 1168–1175.
- [30] J. Kronander, S. Gustavson, G. Bonnet, A. Ynnerman, and J. Unger, "A unified framework for multi-sensor HDR video reconstruction," *Signal Process., Image Commun.*, vol. 29, no. 2, pp. 203–215, 2014.
- [31] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE CVPR*, Jun. 2005, pp. 60–65.
- [33] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
- [34] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in *Proc. IEEE CVPR*, Jun. 2006, pp. 901–908.
- [35] W. Li, Y. Zhou, N. Poh, F. Zhou, and Q. Liao, "Feature denoising using joint sparse representation for in-car speech recognition," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 681–684, Jul. 2013.
- [36] N. Yu, T. Qiu, and F. Ren, "Denoising for multiple image copies through joint sparse representation," *J. Math. Imag. Vis.*, vol. 45, no. 1, pp. 46–54, 2013.
- [37] 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, pp. 153–163, 2016.
- [38] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [39] A. Darmont, *High Dynamic Range Imaging: Sensors and Architectures*. Bellingham, WA, USA: SPIE Press, 2012.
- [40] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. S. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
- [41] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [42] A. Singh, F. Porikli, and N. Ahuja, "Super-resolving noisy images," in *Proc. IEEE CVPR*, Jun. 2014, pp. 2846–2853.
- [43] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.
- [44] H. S. Malvar, L.-W. He, and R. Cutler, "High-quality linear interpolation for demosaicing of Bayer-patterned color images," in *Proc. IEEE ICASSP*, May 2004, p. iii-485-8.
- [45] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun, "Fast burst images denoising," *ACM Trans. Graph.*, vol. 33, no. 6, p. 232, 2014.
- [46] J. H. Lee, I. Choi, and M. H. Kim, "Laplacian patch-based image synthesis," in *Proc. IEEE CVPR*, Jun. 2016, pp. 2727–2735.
- [47] S.-H. Baek, I. Choi, and M. H. Kim, "Multiview image completion with space structure propagation," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 488–496.
- [48] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.
- [49] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. ECCV*, 2010, pp. 778–792.
- [50] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *ACM Commun.*, vol. 24, no. 6, pp. 381–395, 1981.
- [51] M. H. Kim and L. W. MacDonald, "Rendering high dynamic range images," in *Proc. 17th Annu. EVA Conf. Int. (ECI)*, Jul. 2006, p. 22.
- [52] M. H. Kim and J. Kautz, "Consistent tone reproduction," in *Proc. IASTED*, 2008, pp. 1–8.
- [53] M. H. Kim, T. Weyrich, and J. Kautz, "Modeling human color perception under extended luminance levels," *ACM Trans. Graph.*, vol. 28, no. 3, p. 27, 2009.
- [54] M. H. Kim, "High-fidelity colour reproduction for high-dynamic-range imaging," Ph.D. dissertation, Dept. Comput. Sci., Univ. College London, London, U.K., 2010.
- [55] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic, "Temporally coherent local tone mapping of HDR video," *ACM Trans. Graph.*, vol. 33, no. 6, p. 196, 2014.
- [56] R. Boitard, M. T. Pourazad, P. Nasiopoulos, and J. Slevinsky, "Demystifying high-dynamic-range technology: A new evolution in digital media," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 72–86, Oct. 2015.
- [57] S. Sukegawa *et al.*, "A 1/4-inch 8Mpixel back-illuminated stacked CMOS image sensor," in *Proc. IEEE ISSCC*, Feb. 2013, pp. 484–485.
- [58] T. Morgenstern, G. Bornhoeft, and S. Goerlich, "Miniaturized spectroradiometer," JETI Techn. Instrum. GmbH, Jena, Germany, Tech. Rep., Jun. 2004, 2004. [Online]. Available: http://www.jeti.com/cms/images/jeti_com/down/papers/svetlo2004.pdf
- [59] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock, "An introduction to total variation for image analysis," *Theor. Found. Numer. Methods Sparse Recovery*, vol. 9, pp. 263–340, Jul. 2010.
- [60] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.
- [61] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [62] E. Kodak. (2014). *Kodak PhotoCD Lossless True Color Image Suite*. [Online]. Available: <http://r0k.us/graphics/kodak/>
- [63] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Jun. 2016, pp. 770–778.
- [65] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE CVPR*, Jun. 2015, pp. 769–777.



Inchang Choi received the B.S. and M.S. degrees in computer science from the Korea Advanced Institute of Science and Technology in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree in computer science with a focus on learning-based image reconstruction in computational imaging. His research interests include machine learning and various applications of computational imaging, such as high-dynamic-range imaging and hyperspectral imaging.



Seung-Hwan Baek received the B.S. (*magna cum laude*) degree in computer science and engineering from Sogang University in 2013 and the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology (KAIST) in 2015. He is currently pursuing the Ph.D. degree with the KAIST School of Computing, where he is involved in computational photography leveraging computer vision, computer graphics, and optics. His research interests include imaging, display, image processing, deep learning, optimization, and augmented reality.

He received the Outstanding M.S. Thesis Award. He was a recipient of the Best Application Paper Award and the Best Demo Award at ACCV 2014.



Min H. Kim received the Ph.D. degree in computer science from University College London in 2010 with a focus on color reproduction in computer graphics. He is currently an Associate Professor of computer science with the Korea Advanced Institute of Science and Technology (KAIST), leading the Visual Computing Laboratory. Prior to KAIST, he was a Post-Doctoral Researcher with Yale University. His research interests include computational imaging, computational photography, 3D imaging, and hyperspectral imaging, in addition to color and visual perception. In addition to serving on many conference program committees, such as SIGGRAPH Asia and Pacific Graphics, he has been an Associate Editor in various journals: *ACM Transactions on Graphics*, *ACM Transactions on Applied Perception*, and *Elsevier Computers and Graphics*.