Joint Denoising and HDR for RAW Image Sequences

A. Buades, O. Martorell, and M. Sánchez-Beeckman

***Abstract*—We propose a patch-based method for the simultaneous denoising and fusion of a sequence of multi-exposed RAW images. A spatio-temporal criterion is used to select similar patches along the sequence, and a weighted principal component analysis (WPCA) simultaneously denoises and fuses the multi-exposed data. The overall strategy permits to denoise and fuse the set of images without the need to recover each denoised image in the multi-exposure set, leading to a very efficient procedure. Moreover, ghosting removal is included naturally as part of the method by the way patches are selected and the weighted principal component analysis. Several experiments show that the proposed method obtains state-of-the-art fusion results with real RAW data. The method is very flexible, it can be easily adapted to other kinds of noise and extended to video HDR and denoising.**

***Index Terms*—Digital photograph, high dynamic range, image denoising, image fusion, image restoration, image sequences, principal component analysis.**

I. Introduction

A HIGH Dynamic Range (HDR) image has a larger than usual range of luminosity between its brightest and darkest areas. High Dynamic Range imaging refers to the set of methods and techniques that permit to increase the dynamic range of images and videos. We will deal with the combination of several Low Dynamic Range (LDR) images of the same scene acquired with different exposure times, in order to create an HDR image.

HDR methods combine radiance values, obtained from the 8 b color image by inverting the camera response function (CRF). The CRF has to be estimated, generally using the method proposed by Devebec and Malik [21].

For static sequences, many classical methods combine the set of images using a per-pixel weighted average of radiance values. The weights are designed to diminish the contribution of under- and overexposed pixels, e.g. Devebec and Malik [21], Mitsunaga and Nayar [53] or Mann and Picard [50]. In the case of scenes with moving objects or camera motion, methods have to be adapted in order to avoid introducing ghosting effects [35], [39], [42].

Manuscript received 21 June 2023; revised 14 December 2023 and 9 January 2024; accepted 11 January 2024. Date of publication 16 January 2024; date of current version 20 February 2024. The work of M. Sánchez-Beeckman was supported by the Conselleria de Fons Europeus, Universitat i Cultura del Govern de les Illes Balears under Grant FPU2023-011-C. This work was supported in part by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe" through European Union under Grant PID2021-125711OB-I00. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mauricio Delbracio. *(Corresponding author: O. Martorell.)*

The authors are with the Institute of Applied Computing and Community Code (IAC3) E-07122 Palma, Spain, and also with the Department of Mathematics and Computer Science, Universitat de les Illes Balears, E-07122 Palma, Spain (e-mail: o.martorell@uib.cat).

Digital Object Identifier 10.1109/TCI.2024.3354649

While the application of the inverse CRF allows to combine values acquired with different exposure times, it is not adequate for taking care of noise. Most common cameras use a CCD or CMOS sensor device measuring a single colour per pixel. The selected configuration of the sensor usually follows the Bayer colour filter array (CFA) [9]: out of a group of four pixels, two are green, one is red and one is blue, forming a quincunx pattern [7]. We call this the RAW image. At this stage, noise values at neighboring pixels are uncorrelated. Demosaicking is usually performed by combining close values from the same channel or the other two. As a result, the noise gets colour and spatially correlated. The rest of the imaging chain, used to produce an image ready for visualization, consists mainly in colour and gamma corrections and compression. This process enhances the noise in the dark parts of the image, leading to contrasted coloured spots of several pixels. The inversion of the CRF cannot revert all noise correlating stages.

On the other hand, pixel intensities in the RAW domain depend linearly with the number of photons received by the camera sensor, and their noise can be assumed to be spatially uncorrelated. This fact gives rise to two consequences. First, working in the RAW domain can be beneficial for denoising. Second, frames in a burst or sequence of RAW images taken with different exposure times follow the same linear relation between their pixel values and exposure time (up to saturation). Although the radiance values obtained after applying the inverse CRF to an already processed image follow the same relation, inaccuracies on the CRF estimation may introduce inaccuracies in this correspondence. For these reasons, we will directly perform denoising and HDR fusion in the RAW domain, without having the need to estimate a CRF.

We propose a joint denoising and HDR method for removing noise and building an HDR image free of ghosting artifacts. We are inspired by classical video denoising [12] and HDR methods. Video denoising methods take into account a spatio-temporal neighbourhood of each pixel under consideration improving the noise reduction capabilities. The use of patch based methods permits to adapt to the motion avoiding ghosting artifacts. The use of robust estimation techniques, namely PCA, further prevents ghosting and increases the texture and detail reconstruction.

The main novelties of the proposed approach are:

- The joint denoising and fusion without the need of recovering each denoised exposure.
- The removal of ghosting artifacts as a natural part of the algorithm with the use of patch comparison and robust statistical techniques.
- The use of weighted PCA for recovering texture.

- The flexibility of the method, which can be easily adapted to other kinds of noise or images, and directly extended to video HDR and denoising.

The remaining sections of the paper are organized as follows. Section II describes the existing literature on HDR and denoising methods. In Section III we describe the complete method. In Section IV, we discuss the implementation of the method and compare with state-of-the-art algorithms. Finally, we draw some conclusions in Section V.

II. Related Work

A. White Noise Removal in Image Sequences

Local average methods, such as the bilateral filter [68], or patch-based methods like NL-means [11], BM3D [19] and NLBayes [44] and sparse coding algorithms [46] can be easily adapted to image sequences just by extending the neighbouring area to adjacent frames. However, the performance of these methods is improved by introducing motion compensation. These compensated filters estimate explicitly the motion in the sequence and use it to compensate the neighbourhoods, yielding stationary data. The BM3D extension, VBM4D [49], exploits the mutual similarity between 3-D spatio-temporal volumes constructed by tracking blocks along trajectories defined by the motion vectors. In [12], the authors propose combining optical flow estimation and patch-based methods for denoising. The algorithm compensates the failure of these requirements by introducing spatio-temporal patch comparison and denoising in an adapted PCA-based transform.

Recently, neural network methods have appeared making use of motion compensation [65], patch-based processing [20], [71] or being applied recursively in groups of three consecutive frames [66]. Self-supervised methods have also been proposed by using warped neighbouring frames to define the loss function [22].

B. HDR and Deghosting

Many of the early HDR methods applied classical pixel fusion algorithms with different weight and CRF estimation strategies. The most well-known ones were proposed by Mann and Picard [50], Devebec and Malik [21], Khan et al. [42] and Mitsunaga and Nayar [53].

Later HDR methods have devoted their resources to removing ghosting artifacts due to the motion of the camera or the objects in the scene. Global alignment algorithms aim at detecting different global motion models: translation [5], [74], rotation [15], affinity [36] or homography [51], [69]. There exist a variety of methods for estimating the desired motion model: Rad et al. [61] and Yao [78] propose registering the images by translating the problem to the Fourier domain; Mann et al. [51] and Candocia [14] use the comparametric equations to find the desired global transformation; Tomaszewska and Mantiuk [69] and Gevrekci and Gunturk [26] find the transformation by matching image descriptors, such as SIFT [48] or CIFT. Moving object removal algorithms replace selected regions by an estimation of the static background. These methods may fail in presence of dynamic backgrounds, moving objects with occlusions or insufficient number of exposures [29], [42], [57], [64]. Moving object selection methods detect the presence of moving regions and substitute those pixels by corresponding areas from other images [30], [37], [62]. Moving object registration methods remove the majority of artifacts by computing motion between a reference frame and the other images of the stack. The methods in this family can also be classified in two subclasses: optical flow methods [24], [33], [80] or patch-based methods [34], [63]. The methods in the first group aim at finding a pixel-wise motion between images, while the ones in the latter use patch-based strategies to register images and remove artifacts. See Tursun et al. [70] for a detailed review of HDR deghosting algorithms.

More recent methods have begun using deep learning techniques to remove ghosting artifacts. The first method using a convolutional neural network is credited to Kalantari et al. [39], which performs a pixel-wise merging. Wu et al. [75] propose a non-flow-based approach for HDR. Peng et al. [58] propose an improvement of [39] by using a state-of-the-art optical flow method and a more accurate merging network. Yan et al. handle motion between images by an attention module [76] and also propose a non-local network [77]. Niu et al. [55] propose HDR-GAN, a GAN-based model with a novel generator network. Liu et al. [47] align the dynamic frames with a deformable alignment module. Prabhakar et al. [60] propose to use recurrent neural networks to obtain deghosted HDR images. See Wang et al. [72] for a detailed review on learning based methods.

C. Joint HDR and Noise Removal

Sequential filtering and HDR: Min et al. [52] filter the set of images by spatio-temporal motion-compensated anisotropic filters prior to HDR reconstruction. Lee et al. [45] use a sub-band architecture for fusion, with a weighted combination using a motion indicator function to avoid ghosting effects. The low frequency bands are filtered with a multi-resolution bilateral filter while the high frequency bands are filtered by soft thresholding.

Joint HDR and denoising on sRGB data: Akyuz et al. [6] denoise each frame before fusion, by averaging a subset of frames in the radiance domain. Tico et al. [67] combine an initial fusion with the image of the sequence with the shortest exposure in the luminance domain. This combination is performed in the wavelet domain and coefficient attenuation is applied to the coefficients of the difference of luminances.

Ahmad et al. [4] identify noisy pixels and reduce their weight during image fusion. Goossens et al. [27] propose a realistic noise model for HDR imaging that takes into account an accurate noise model of image capture. With that, they modify the HDR weights to improve the PSNR.

On RAW data: Kronander et al. [43] propose a unified framework for HDR reconstruction from raw CFA data. The proposed method is based on an adaptive spatial and cross-sensor filtering using polynomial approximation. During reconstruction, they perform CFA interpolation, resampling and HDR assembly in a single operation.

Aguerrebere et al. [2] propose a filter based on NL-means [11] to jointly denoise and perform HDR on a set of multi-exposed frames. For each patch on a reference image, similar patches are selected using a Chi-square distance function and the method by

Granados et al. [28] is applied to the center pixel of these patches. The fusion method by Granados et al. takes into account the estimated noise model. The selection of similar patches permits to deal with ghosting artifacts.

Deep learning: Chaudhari et al. [16] propose a deep learning method for HDR from raw CFA data. The proposed network is capable of performing demosacking, alignment and fusion of the input images to obtain a deghosted HDR image. Aguerrebere et al. [1] propose a robust framework enabling the use of Gaussian local priors on image patches for solving a useful family of restoration problems by drawing on a hierarchical Bayesian approach. The advantage of the proposed framework is its ability to deal with signal dependent noise, therefore making it suitable for realistic digital photography applications. In particular, it permits to denoise and reconstruct an HDR image from a single view.

III. Joint HDR and Noise removal

The proposed algorithm performs a joint denoising and HDR of a reference RAW image given a series of—not necessarily static—accompanying images taken under a variety of exposure times. We denote the set $\mathcal{I} = \{I_1, I_2, \dots, I_N\}$ as a sequence of noisy RAW images with corresponding exposure times $\{\tau_1, \tau_2, \dots, \tau_N\}$ and assumed to be taken with the same ISO value.

A. Noise Model With Different Exposure Times

The primary sources of noise in a camera sensor are shot noise and read noise. They can be modeled, respectively, with a Poisson and a Gaussian distribution. The sensor, moreover, adds a constant positive offset O to the values it reads in order to avoid the presence of negative values due to noise. By using the normal approximation of the Poisson distribution, the noise model of a read pixel x_{read} is thus usually described as

$$x_{\text{read}} \sim \mathcal{N}\left(x_{\text{true}} + O, a x_{\text{true}} + b\right), \tag{1}$$

where x_{true} is the true pixel value, and a and b depend on the sensor characteristics and on the ISO value used to capture the image [25], [32].

We will perfom an estimation of the noise distribution on the input set of images $\mathcal{I}$ to be used later on the denoising and fusion step. We first pack each one of these images into a 4-channel image of half the width and height of the original RAW image, containing the red, blue and two green values, denoted as R, B, G_1 and G_2, respectively. For simplicity, we still denote these images as $I_i,\ i = 1, 2, \dots, N$.

We estimate the level of noise that is present in each one of the four channels of the images I_i separately. To do so, we use the same approach as in [13], which adapts the method proposed by Colom et al. [18] and Ponomarenko [59]. Note that, since all images are taken with the same sensor and ISO value, their noise curves for each colour channel are identical. This way, we compute only one curve $\sigma(x)$, which contains the noise standard deviation at pixel intensity x for the images in the sequence. The noise curve is estimated independently for each of the four channels, so in fact we have $\sigma(x) = (\sigma_R(x), \sigma_{G_1}(x), \sigma_{G_2}(x), \sigma_B(x))$.

B. Normalization and Variance Stabilization

Let $I_{\text{ref}} \in \mathcal{I}$ be chosen as the reference image of the sequence, which we choose to be the middle-exposed one. Since each frame in $\mathcal{I}$ is in the RAW domain, we can assume that pixel intensities scale linearly with their exposure time. We perform an initial normalization step by equalizing the exposure of each image to the reference one. This is accomplished by removing the constant black offset O, scaling the resulting image, and adding O back

$$\hat{I}_i = O + \frac{\tau_{\text{ref}}}{\tau_i}\left(I_i - O\right). \tag{2}$$

We use $\hat{\mathcal{I}}$ to refer to the set of these images. Another set of images $\hat{\mathcal{J}}$ is built as

$$\hat{J}_i = \min\left\{O + \frac{\tau_{\text{ref}}}{\tau_i}(I_i - O), V_{\text{sat}}\right\}, \tag{3}$$

where V_{sat} denotes the saturation value of the sensor. The images in $\hat{\mathcal{J}}$ will be used for registration and patch selection, while those in $\hat{\mathcal{I}}$ will be combined in order to obtain the HDR image. This is due to the fact that areas that are saturated in the reference image might have considerably larger pixel values in the normalized images with darker exposures. The difference in intensities can mislead the optical flow computation, so we solve the problem by re-saturating those bright values. Similarly, saturated areas in the long exposure images will have, once normalized, smaller values than the reference non-saturated ones. Altough these may lead to erroneus optical flow estimation, in Section III-C we will make sure that these areas are not taken into account in the subsequent fusion process by carefully choosing an HDR weighting function depending on the original values in $\mathcal{I}$.

At this point, we want to apply a transformation to each image in $\hat{\mathcal{I}}$ and $\hat{\mathcal{J}}$ such that all intensity values in the same image have the same noise variance. Such transformation $f_c(x)$ is known as variance stabilization transform (VST) and is written as

$$f_c(x) = \int_0^x \frac{dt}{\sigma_c(t)}, \tag{4}$$

being $c \in \{R, G_1, G_2, B\}$ the image channel of the packed CFA and x the intensity value to be transformed [13]. When assuming a linear variance model (1), this stabilization is known as the Anscombe transform [8]. We apply such a VST to each one of the normalized images in $\hat{\mathcal{I}}$ and $\hat{\mathcal{J}}$ using the same estimated noise curve $\sigma(x)$. It can be proved that, if the noise variance is approximated linearly, the uniform variance of the ith stabilized image, $\hat{\sigma}_i^2$, equals τ_{ref}/τ_i. Note that this variance does not hold for saturated pixels; this is not a problem for us, since these pixels are not used when performing denoising and HDR.

C. Joint Denoising and HDR

From now on, we assume that every frame $\hat{I}_i \in \hat{\mathcal{I}}$ and $\hat{J}_i \in \hat{\mathcal{J}}$ has a uniform noise standard deviation $\hat{\sigma}_i$, while any mention of $I_i \in \mathcal{I}$ refers to the original, pre-normalization image using the RGGB space.

We use a colour transformation from RGGB to a decorrelated colour space, which we denote by YUVW, as proposed in [13].

Assuming that noise values at the sensor data are uncorrelated for different colour channels and pixel locations, since the matrix in [13] is an orthonormal matrix, its application to $\hat{I}_i$ and $\hat{J}_i$ for all $i \in \{1, \dots, N\}$ conserves the decorrelation and uniform standard deviation properties. Moreover, its inverse transform is given by its transpose.

Once we have applied the YUVW transformation, we compute the optical flow between the middle-exposed frame $\hat{J}_{\text{ref}}$ and every other $\hat{J}_i \in \hat{\mathcal{J}}$ on the brightness channel Y using the total variation approach in [79]. Moreover, we check the consistency of the obtained values by also computing the inverse flows and making sure that they are approximately reciprocal. Pixels that do not verify this condition are either occluded or in violation of the optical flow's colour constancy assumption, so we mark them as invalid for the rest of the algorithm.

For every overlapping $r \times r$ patch P in the chosen reference, we denote by $\mathcal{P}$ its motion compensated extension to the temporal dimension, that is, a 3D volumetric patch. We consider the patch's values $\hat{\mathcal{J}}(\mathcal{P})$ and select the set of its K most similar extended patches in the clipped images $\hat{\mathcal{J}}$. Invalid points are not considered in this search.

Now, we consider all the $K \cdot N$ 2D patches taking part in the selected spatio-temporal blocks, $\{Q_j\}$, and their associated values $\{\hat{J}_{k_j}(Q_j)\}$, for each $j \in \{1, \dots, N \cdot K\}$, where $k_j \in \{1, \dots, N\}$ refers to the index of the image containing the jth patch. For each of these patches, we compute its associated weight

$$w_j = w_{\text{sim}}\left(\hat{J}_{\text{ref}}(P), \hat{J}_{k_j}(Q_j)\right) \cdot w_{\text{HDR}}\left(I_{k_j}(Q_j)\right), \quad (5)$$

where

$$w_{\text{sim}}\left(\hat{J}_{\text{ref}}(P), \hat{J}_{k_j}(Q_j)\right) = \exp\left(\frac{-\|\hat{J}_{\text{ref}}(P) - \hat{J}_{k_j}(Q_j)\|^2}{h^2}\right) \quad (6)$$

so that patches that are too dissimilar to P lose influence on the filtering step (depending on a parameter h), and $w_{\text{HDR}}(I_{k_j}(Q_j))$ is any HDR weighting function adapted to the RAW range of intensities, evaluated on the original RGGB image I_{k_j} where Q_j is located. In particular, this HDR weighting will discard the use of any saturated patch independently of its colour value after exposure normalization.

Let $\mathbf{X}$ be the matrix whose j-th row is composed of the flattened pixel values of the 2D patch Q_j evaluated on the image $\hat{I}_{k_j}$. Each column $i \in \{1, \dots, r^2\}$ of that matrix contains the values of a particular position inside the selected patches. Similarly to traditional HDR methods, we compute a weighted average of the pixel intensities of each one of these positions using the weights $\{w_j\}$ as a measure of confidence. A vector encoding those averages is given by the barycenter

$$\mathbf{b} = \frac{1}{\sum_{j=0}^{KN} w_j} \sum_{j=0}^{KN} w_j \mathbf{x}_j, \quad (7)$$

where $\mathbf{x}_j^T$ is the jth row vector of $\mathbf{X}$.

Since $\mathbf{X}$ has been constructed using similar patches, we expect the values of each column to be strongly correlated. That is, the information of the image is highly redundant, and we can expect to find a low-dimensional subspace that captures most of the variance of the data. At the same time, not every patch is equally reliable due to possible saturation and dissimilarities with the reference patch being processed. For these reasons, we carry out a Weighted Principal Component Analysis on $\mathbf{X}$, which generalizes the standard PCA by taking into account the patch weights $\{w_j\}$.

We start by centering $\mathbf{X}$ column-wise by subtracting the vector $\mathbf{b}$ to each of its rows:

$$\bar{\mathbf{X}} = \mathbf{X} - \mathbf{1}_{KN \times 1}\mathbf{b}^T. \quad (8)$$

Then, we consider the eigendecomposition of the weighted covariance matrix

$$\frac{V_1}{V_1^2 - V_2}\bar{\mathbf{X}}^T\mathbf{W}\bar{\mathbf{X}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (9)$$

where $\mathbf{W} = \text{diag}(w_1, \dots, w_{KN})$, and V_1 and V_2 are the sum of the weights and the sum of their squares, respectively. If we compute the Singular Value Decomposition of the weighted centered matrix

$$\mathbf{W}^{\frac{1}{2}}\bar{\mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (10)$$

it holds that $\mathbf{S}^2 = \frac{V_1^2 - V_2}{V_1}\mathbf{\Lambda}$ and that $\mathbf{V}$ is the same orthogonal matrix in both (9) and (10). Furthermore, the matrix $\mathbf{US} = \mathbf{W}^{\frac{1}{2}}\bar{\mathbf{X}}\mathbf{V}$ contains the principal components of $\mathbf{W}^{\frac{1}{2}}\bar{\mathbf{X}}$. The first few eigenvalues in $\mathbf{\Lambda}$ (and, thus, singular values $\{s_i\}$ in $\mathbf{S}$) in decreasing order concentrate most of the variability of the group of patches. This has a double implication: first, singular values from a certain point onwards are mostly noise; second, those up to that aforementioned point hold much more information on the structure of $\mathbf{X}$ than the weighted average $\mathbf{b}$ does. Hence, we filter $\bar{\mathbf{X}}$ by discarding the principal components whose associated singular values are lower than a carefully chosen threshold δ^2. More formally, let $\tilde{\mathbf{D}}$ be the diagonal matrix whose entries are

$$\tilde{D}_{ii} = \begin{cases} 1 & \text{if } \frac{V_1}{V_1^2 - V_2}s_i^2 \geqslant \delta^2, \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

so that $\tilde{\mathbf{S}} = \mathbf{S}\tilde{\mathbf{D}}$ is a hard-thresholded version of $\mathbf{S}$. With it, we can approximate the weighted centered matrix as

$$\mathbf{W}^{\frac{1}{2}}\tilde{\mathbf{X}} = \mathbf{U}\mathbf{S}\tilde{\mathbf{D}}\mathbf{V}^T = \mathbf{W}^{\frac{1}{2}}\bar{\mathbf{X}}\mathbf{V}\tilde{\mathbf{D}}\mathbf{V}^T. \quad (12)$$

The $\mathbf{W}^{\frac{1}{2}}$ factor in (12) cannot be directly removed on both sides of the equality because the matrix is not generally invertible due to the existence of saturated patches. However, we only need the reconstruction of the reference patch P to be correct. Thus, if P is not saturated, we reconstruct an approximation of the fused, noise-free patch group as

$$\hat{\mathbf{X}} = \bar{\mathbf{X}}\mathbf{V}\tilde{\mathbf{D}}\mathbf{V}^T + \mathbf{1}_{KN \times 1}\mathbf{b}^T, \quad (13)$$

where the last term re-centers the data around the original intensity values. When P is saturated in I_{ref}, the step from (12) to (13) is invalid, so we simply take $\mathbf{b}$ as the reconstructed pixel values of the patch. This still performs the fusion using the information from the correctly exposed patches.

The aforementioned process is done independently for each channel and each patch on the reference frame. We keep the

reconstructed values of each patch, and we aggregate them to form the fused image. Lastly, we apply the colour space conversion back from YUVW to RGGB and we undo the variance stabilization transform to obtain the resulting image.

We note that, for noise-free images, using a window size of $r = 1$ and setting $w_{\text{sim}} = 1$ with a full cancelation of the principal components—that is, keeping only the barycenter $\mathbf{b}$—is equivalent to a classic HDR procedure with weights w_{HDR}. Indeed, the computation of the vector $\mathbf{b}$ is a spatio-temporal generalization of the classic HDR averaging with additional noise removal when using all the weight factors. The weight $w_{\text{sim}}(\hat{J}_{\text{ref}}(P), \hat{J}_{k_j}(Q_j))$ permits the noise removal but also avoids creating ghosting artifacts. While the barycenter estimation would be enough to remove noise, increase dynamic range and avoid ghosting effects, texture and details might be over-smoothed. The use of PCA avoids such a detail smoothing and improves the deghosting capabilities of the method.

D. Imaging Chain

The denoised and HDR image that results from the procedure still has the structure of a 4-channel packed CFA. For that reason, we must unpack it and interpolate the missing red, green and blue values. We accomplish this by demosaicking the image, for which we can use any standard method, such as [23, Section II].

After demosaicking, the signal strengths of the three colours can be somewhat unbalanced due to how the camera sensors read different types of light. Therefore, to conform them to a more realistic hue, a white balancing step is performed by multiplying each channel by a different value—chosen depending on the processed image. Moreover, to be able to visualize the image on a standard display, we apply a linear transformation to turn the camera RGB values into coordinates in the sRGB colour space.

Finally, we adjust the contrast and brightness of the HDR image and compress its range so that it fits in 8 bits per channel. We use the local tone mapping algorithm in [56]. Local tone mapping algorithms have the disadvantage that they magnify image noise considerably; this is not a problem for us, given that our images have already had their noise removed.

E. Implementation Details

The same parameters have been used in all experiments. We keep a fixed window size of $r = 7$, and a parameter $h = 60$ (fixed for 12 b RAW images). The HDR weighting function is

$$w_{\text{HDR}}\left(I_j(P)\right) = \min_{c \in \mathcal{C}} w_c\left(I_j(P)\right) \tag{14}$$

with $\mathcal{C} = \{R, G_1, G_2, B\}$ and

$$w_c\left(I_j(P)\right) = 1 - \left(2 \cdot \frac{I_j^c(P) - O}{V_{\text{sat}} - O} - 1\right)^{12}, \; c \in \mathcal{C} \tag{15}$$

where O the black offset value, V_{sat} denotes the saturation value of the sensor and $I_j^c(P)$ is the central value of channel c in patch $I_j(P)$. The weights in (15) are obtained by modifying the weights from Khan et al. [42]. The resulting images are processed with Matlab's `localtonemap` function, which implements the tone mapping algorithm in [56].

The value of the threshold is estimated as $\delta^2 = \kappa\bar{\sigma}^2$, where $\bar{\sigma}^2$ is the average variance of the noise in the selected set of patches. This $\bar{\sigma}^2$ depends on the ratio of exposure times $\{\tau_1, \tau_2, \ldots, \tau_N\}$ as

$$\bar{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \frac{\tau_{\text{ref}}}{\tau_i}}. \tag{16}$$

The value of κ has been empirically set to $\kappa = 2$.

IV. Results and Discussion

We compare the proposed method with state of the art methods with RAW data. We have used images taken by ourselves (Fig. 1), images provided by Aguerrebere et al. [2] (Fig. 2) along with the implementation of their method[1] and images provided by Karajuzovic-Hadziabdic et al. [40], [41] (Fig. 3). All the examples are non static, the scenes vary due to camera motion or changes in object position. The dataset provided by Karajuzovic-Hadziabdic et al. also contains a static set corresponding to each sequence, from which a ground truth HDR image can be estimated. We use it to perform a quantitative evaluation of the proposed method by comparing our method's obtained result with the ground truth image. Although the HDR and denoising process is applied in the RAW data, in all cases we display the processed images by applying the processing pipeline described in subsection III-D.

To assess the quality of our method, we compare it with state-of-the-art methods for HDR in both a qualitative and quantitative manner. We set it side by side with the classic methods from Sen et al. [63] and Aguerrebere et al. [2], and the recent deep learning methods DeepHDR [75], HDR-GAN [55] and HDRRNN [60]. We use the implementations provided by the authors themselves[2], [3], [4], [5], [6]. All compared methods take into account the dynamic nature of the sequences and claim to avoid any ghost artifact creation.

A. Ablation Study

We study the effect each separate type of weight has on the HDR process. In particular, we notice that the introduction of patch similarity weights helps against the presence of ghosting that usually appears when applying HDR algorithms on moving images. This strong point arises from the fact that a weighting based solely on pixel intensities can mistakenly merge patches that have been aligned inaccurately. Those patches give up their influence when attaching an extra weight that compares them to the reference patch. First, we keep none of the coefficients of the weighted PCA, so that the method only centers the patch to the weighted barycenter. Fig. 4 illustrates the deghosting effect

[1] [Online]. Available: https://perso.telecom-paristech.fr/gousseau/hdr_denoising/

[2] [Online]. Available: https://web.ece.ucsb.edu/~psen/hdrvideo

[3] [Online]. Available: https://perso.telecom-paristech.fr/gousseau/hdr_denoising/

[4] [Online]. Available: https://github.com/elliottwu/DeepHDR

[5] [Online]. Available: https://github.com/nonu116/HDR-GAN

[6] [Online]. Available: https://github.com/Susmit-A/HDRRNN

Fig. 1. Our images, acquired with a Canon 100D. The camera is slightly moving, which causes an approximate global image translation.

Fig. 2. Images provided by Aguerrebere et al. [2] used in our experiments. For the images in the first row, there is a significant gap in time, causing people and cars to change positions or even disappear from the scene. For the second row, motion is mostly caused by the change of camera position, and people moving on the bridge.

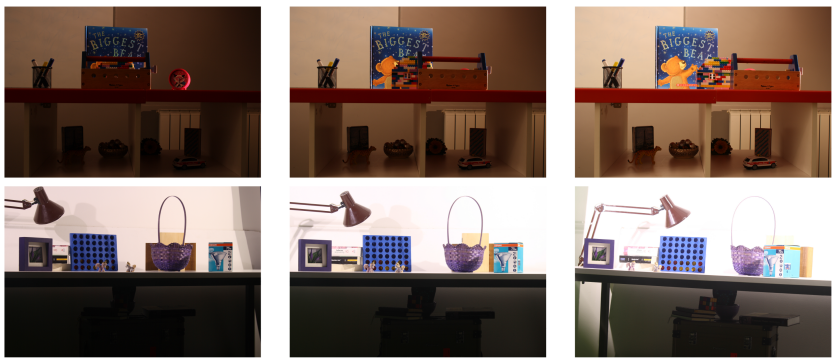


Fig. 3. Images provided by Karajuzovic-Hadziabdic et al. [40], [41] used in our experiments. Both the camera and the objects are moving in the scene.

of the similarity weights: the small dots that appear in Fig. 4(b) on top of the wooden object disappear when we include them (Fig. 4(c)). Finally, adding the weighted PCA instead of the plain barycenter centering removes the ghosting on both the bird and the leopard.

Fig. 5 illustrates the impact that the exposure time of the reference image has on the solution. In the example in the first row, changing the reference produces the same results. In the second one, however, we can observe some pink spots appearing around originally saturated areas when using the image with the longest exposure time as reference. Non-saturated areas are unnaffected. This phenomenon occurs because of an incorrect exposure equalization in saturated areas, which impedes an accurate alignment and matching of patches with optical flow. Indeed, this is the main limitation of our algorithm and the reason we choose as reference the image with the middle exposure.

Next, we test the denoising capabilities of the weighted PCA. Fig. 6 demonstrates how its use (Fig. 6(b)) adds considerable detail to the resulting image compared to the plain weighted average (Fig. 6(c)).

Finally, we compare the use of a different w_{HDR} weighting function, specifically Granados et al. [28]. This HDR function, which is optimal in the estimation of the true HDR value [3], does

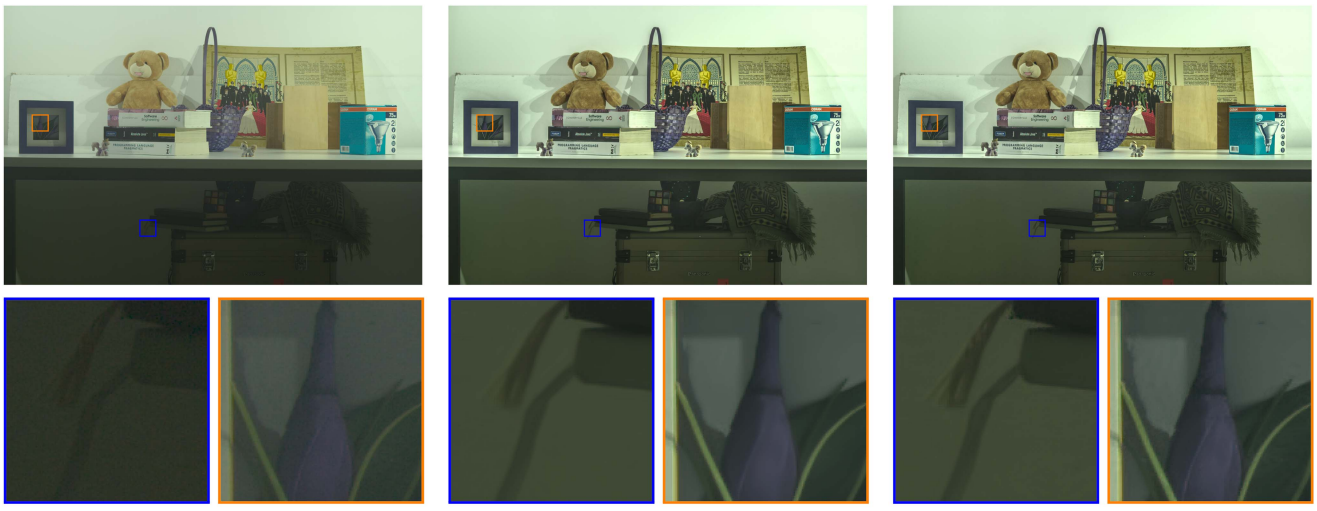
(a) (b) (c) (d)

Fig. 4. Ghosting effects are corrected with the introduction of similarity weights and the weighted PCA. (a) Tone mapped reference noisy image. (b) HDR without similarity weights (only barycenter centering). (c) HDR with similarity weights (only barycenter centering). (d) Final result with similarity weights and weighted PCA.

(a) Under-exposed frame as reference (b) Middle-exposed frame as reference (c) Over-exposed frame as reference

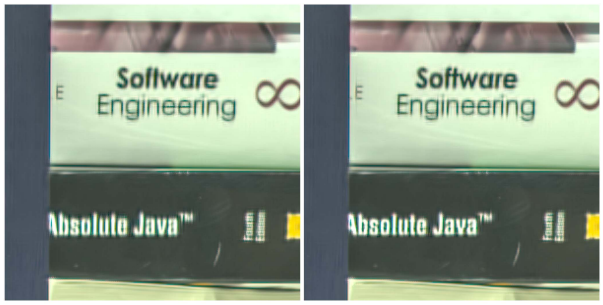
(d) Under-exposed frame as reference (e) Middle-exposed frame as reference (f) Over-exposed frame as reference

Fig. 5. HDR fusion varying the reference frame.



(a) Reference image (b) HDR only with barycenter centering (c) HDR with WPCA

Fig. 6. Use of the weighted principal components permits to increase detail and texture preservation.



(a) Granados et al. [28] weight term (b) Our weight term

Fig. 7. Comparison of the proposed scheme with different w_{HDR} terms. Using the weight in Granados et al., the bright words on the black book and the dark words on the bright one are blurred because it does not vanish near saturation values. Therefore, a classical HDR function avoiding the use of values near saturation has been preferred.

not vanish near the saturation value. As a consequence, details on a black or white saturated background tend to be blurred. This is noticeable in Fig. 7.

B. Visual Comparison With State-of-The-Art

The methods by Sen et al. [63], DeepHDR [75], HDR-GAN [55] and HDRRNN [60] do not take into account noise removal. For this reason, to compare them with the rest of methods, we denoise each of the RAW images before applying the corresponding HDR algorithm. In order to be fair, we use the same denoising process we propose, but applied to each image of the set independently and neglecting the w_{HDR} weight. Fig. 8 displays the result of applying all the methods to the original RAW images without the previous denoising. It is clear that only Aguerrebere et al. [2] and our method are able to remove the noise. Fig. 9 displays the HDR result applied to the denoised exposures (Aguerrebere et al. and ours are applied to the original images). The same colour, texture and HDR effect is reproduced, but noise is removed.

Since deep learning methods take full colour images as input and in order to obtain comparable results, we apply the proposed imaging chain to the RAW images except for the tone mapping step. By keeping floating point values, we can feed directly these images to the networks as radiance data without introducing any quantization effect. Finally, all images are displayed after being applied the same tone mapping function (Matlab's `localtonemap`).

Due to hardware limitations, we could not feed the full resolution images into the deep learning architectures. We used cropped regions of size 2560×2560, which are the ones displayed in the following figures.

Fig. 9 compares the methods in a case with no strong saturation in any of the exposures. HDRRNN reconstructs an image with a different colour than the rest. Aguerrebere et al. has several artifacts and excessive blurring on the pen. Unlike the other methods, where denoising is performed with our proposed strategy, Aguerrebere et al. apply their own noise removal technique based on NLMeans [11], whose averaging strategy is less accurate than the principal component analysis. This difference causes that blurring effect to reoccur in challenging images. All the other methods, including the proposed one, produce a good looking result without artifacts.

Fig. 10 displays a more complex example where both camera and object motion appear. Only Aguerrebere et al. and the proposed method are able to correctly reconstruct the image in the saturated parts, while other methods produce non good looking results on those areas. However, Aguerrebere et al. excessively blurs some details on the boat. All the methods are able to deal with ghosting artifacts except HDRRNN.

Fig. 11 illustrates how DeepHDR and HDRRNN have trouble with saturated areas: the first one produces a pattern on those areas and the second one outputs a blueish colour. Aguerrebere et al. correctly recovers saturated areas but removes much texture in the dark areas. Our method seems to give the best overall quality noise removal and HDR reconstruction.

Finally, Fig. 12 shows how DeepHDR, HDR-GAN and HDR-RNN have ghosting artifacts. The first row of Fig. 3 shows that the big wooden object moves along the shelf in the different images, occluding other objects in the scene. A close look on the object shows that the three aforementioned methods produce different artifacts: HDR-GAN generates a ghosting shadow on the background, HDRRNN produces strange colours and DeepHDR generates a wrong texture in the wooden surface. The other three methods produce a good looking result.

C. Numerical Comparison With State-of-The-Art

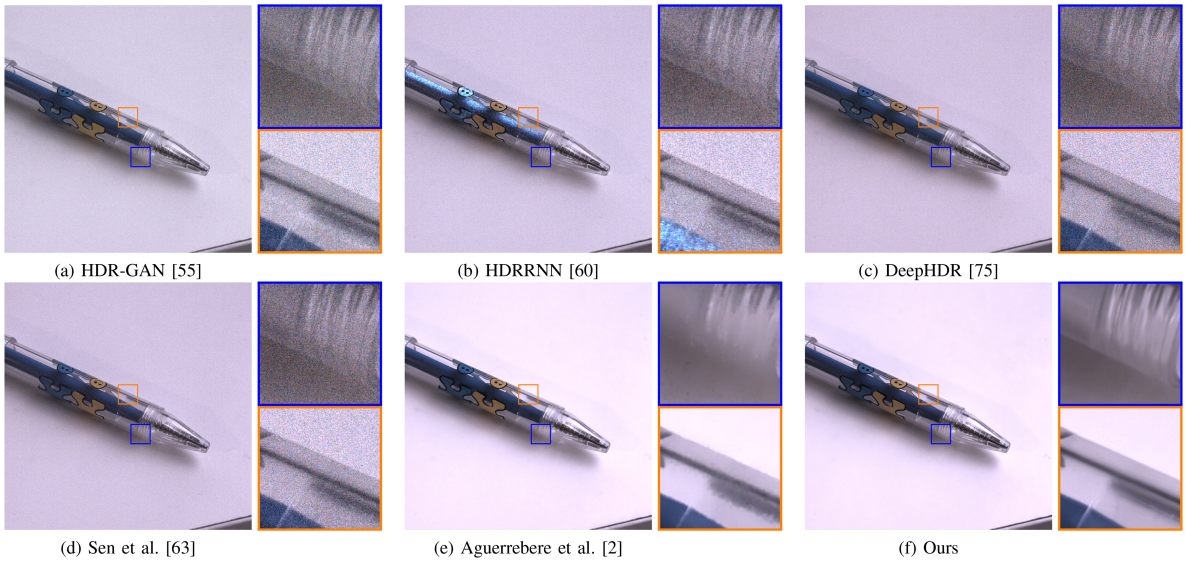
We perform the quantitative comparison by comparing the results obtained with sequences provided by Karajuzovic-Hadziabdic et al. [40], [41] with the corresponding ground truth. We use the Peak Signal-to-Noise Ratio (PSNR), SSIM [73], and Perceptual Index (PI) [10] measures. As in [17], [38], we apply the following tonemap transformation to the images before computing the PSNR and SSIM

$$T = \frac{\log(1+\mu S)}{\log(1+\mu)}, \tag{17}$$

being $\mu = 0.01$, S the fused and denoised image after the demosaicking and white balance, T the tone mapping in which the measures are computed. The PI measure compares the final tone mapped images in sRGB 8-bit format, so we compute it after the full imaging chain.

Like it was done for the visual comparison, a preliminary noise removal stage is applied image by image if the method itself does not include it. The results of the different measure are displayed in Tables I, II and III for PSNR, SSIM and PI, respectively. The numerical results in PSNR and SSIM reaffirm the conclusions drawn during the visual comparison: our method produces on average the best results, with no ghosting artifacts nor noise.

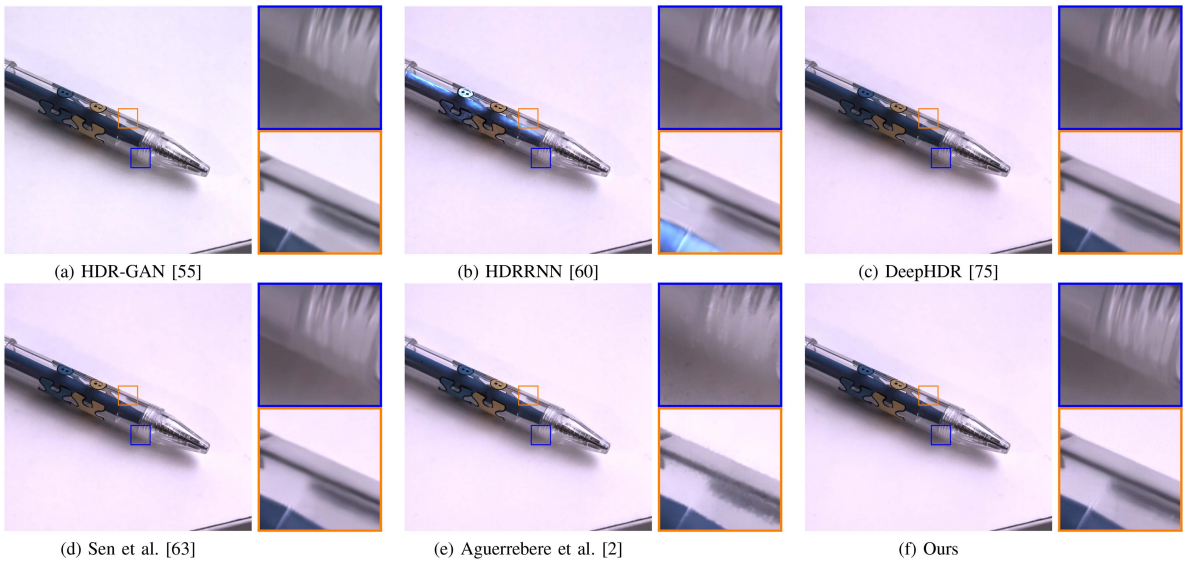
The PI measure attributes a superior performance to HDR-RNN [60] despite some of its results not being visually pleasant. In fact, the metric evaluates the final appearance of the image based on traits such as sharpness and colour. The objective of Sen et al., Aguerreberre et al., and our method is to fuse the images in order to increase the dynamic range; the final degree of sharpness and colour might be adjusted by the subsequent imaging



(a) HDR-GAN [55] (b) HDRRNN [60] (c) DeepHDR [75]

(d) Sen et al. [63] (e) Aguerrebere et al. [2] (f) Ours

Fig. 8. HDR results on an original image sequence. Only Aguerrebere et al. and ours are able to remove noise. For that reason, we denoise each frame before applying the rest of HDR methods as displayed in Fig. 9.



(a) HDR-GAN [55] (b) HDRRNN [60] (c) DeepHDR [75]

(d) Sen et al. [63] (e) Aguerrebere et al. [2] (f) Ours

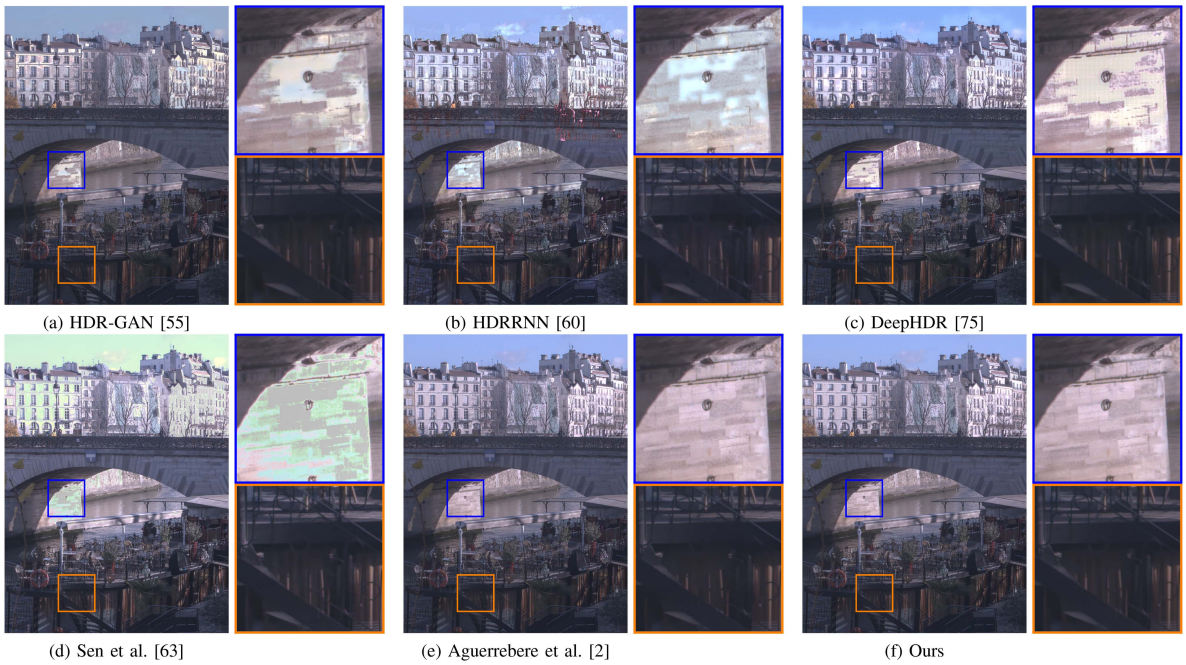
Fig. 9. HDR results on a denoised image sequence. Compare to Fig. 8.

TABLE I

MEASURED PSNR ON THE SEQUENCES PROVIDED BY KARAJUZOVIC-HADZIABDIC ET AL. [40], [41], AS WELL AS THE MEAN VALUE OVER ALL THE DATA

	HDR-GAN	HDRRNN	Deep HDR	Sen et al.	Aguerrebere et al.	Ours
setup1 complex	31.6416	21.4905	18.1983	44.2303	47.9788	**48.1834**
setup1 occlusion	33.0574	23.1981	17.9995	48.9452	48.9463	**49.6900**
setup2	24.6789	17.3641	16.3514	43.6248	49.9234	**50.3669**
setup2 multiview	23.8690	14.8090	18.8549	38.7167	48.7816	**49.3969**
setup3 crop 1	29.7557	26.3483	20.2023	37.7122	43.0010	**47.1868**
setup3 crop 2	24.2645	21.4709	15.7963	31.0629	37.6382	**48.4950**
setup4 handheld	14.8426	10.2060	16.5914	39.3937	**44.3254**	44.1320
Mean	26.0157	19.2696	17.7134	40.5265	45.7992	**48.2073**

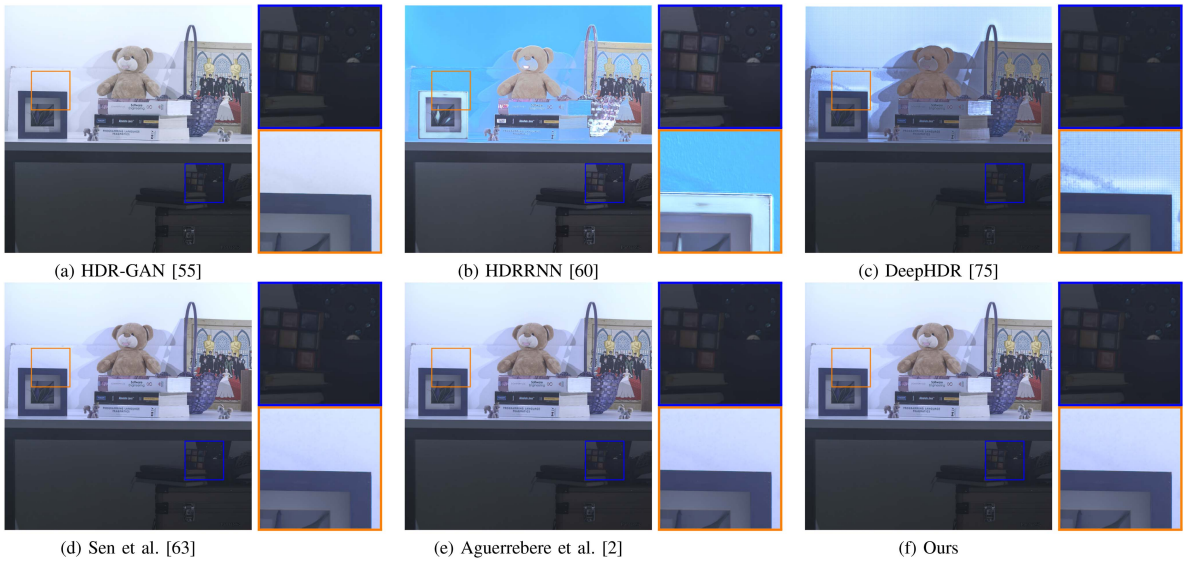
Methods HDR-GAN, HDRRNN, DeepHDR and Sen et al. are applied to the previously denoised raw sequences by a single image version of the proposed denoising method. For each sequence, we mark in bold the method with a higher (better) value.



(a) HDR-GAN [55] (b) HDRRNN [60] (c) DeepHDR [75]

(d) Sen et al. [63] (e) Aguerrebere et al. [2] (f) Ours

Fig. 10. HDR results on a denoised image sequence provided by Aguerrebere et al. [2].



(a) HDR-GAN [55] (b) HDRRNN [60] (c) DeepHDR [75]

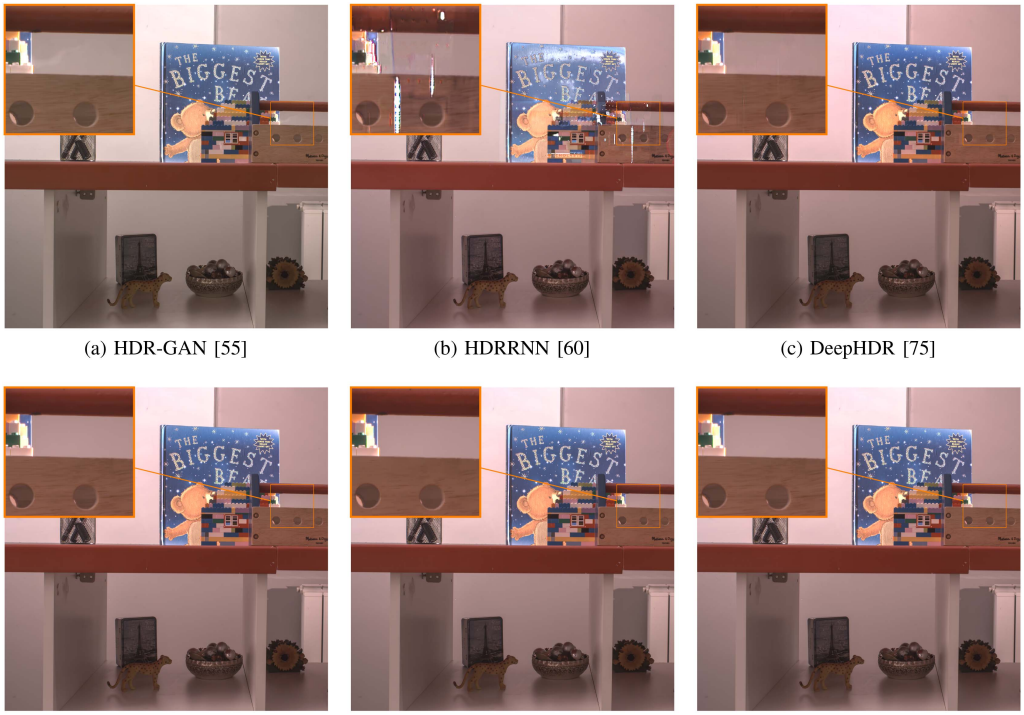
(d) Sen et al. [63] (e) Aguerrebere et al. [2] (f) Ours

Fig. 11. HDR results on the denoised image sequence setup2 provided by Karajuzovic-Hadziabdic et al. [40], [41].

TABLE II
MEASURED SSIM [73] ON THE SEQUENCES PROVIDED BY KARAJUZOVIC-HADZIABDIC ET AL. [40], [41], AS WELL AS THE MEAN VALUE OVER ALL THE DATA

	HDR-GAN	HDRRNN	Deep HDR	Sen et al.	Aguerrebere et al.	Ours
setup1 complex	0.982157	0.908305	0.640575	0.999752	0.999894	**0.999937**
setup1 occlusion	0.985632	0.936527	0.642013	0.999965	0.999917	**0.999972**
setup2	0.971500	0.881667	0.656106	0.999903	0.999911	**0.999935**
setup2 multiview	0.968366	0.722157	0.654697	0.999868	0.999811	**0.999974**
setup3 crop 1	0.963307	0.909900	0.615736	0.996364	0.997052	**0.998036**
setup3 crop 2	0.966712	0.916591	0.661117	0.995009	0.997364	**0.999723**
setup4 handheld	0.636871	0.395103	0.595764	0.998475	0.999565	**0.999596**
Mean	0.924935	0.810036	0.638001	0.998477	0.999073	**0.999596**

Methods HDR-GAN, HDRRNN, DeepHDR and Sen et al. are applied to the previously denoised raw sequences by a single image version of the proposed denoising method. For each sequence, we mark in bold the method with a higher (better) value.



(a) HDR-GAN [55] (b) HDRRNN [60] (c) DeepHDR [75]

(d) Sen et al. [63] (e) Aguerrebere et al. [2] (f) Ours

Fig. 12. HDR results on the denoised image sequence setup1 occlusion provided by Karajuzovic-Hadziabdic et al. [40], [41].

TABLE III
PERCEPTUAL INDEX (PI) [10] MEASURED ON ALL SEQUENCES USED IN THE EXPERIMENTATION AS WELL AS THE MEAN VALUE OVER ALL THE EXAMPLES

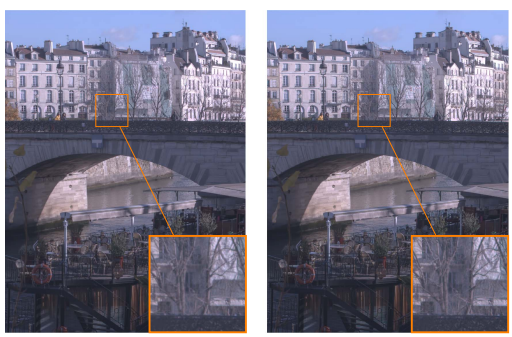
Sequence	HDR-GAN	HDRRNN	Deep HDR	Sen et al.	Aguerrebere et al.	Ours	Ours sharpened
example 1	2.89	**2.64**	2.86	2.88	3.08	**2.84**	2.99
example 2	3.61	2.99	**2.92**	**2.87**	3.50	3.61	3.12
pen crop 1	8.00	7.40	7.14	7.84	8.72	**6.58**	**6.57**
pen crop 2	7.08	6.53	**6.24**	6.84	7.55	6.44	**6.42**
setup1 complex	5.71	**5.12**	**4.94**	5.74	5.76	5.67	5.23
setup1 occlusion	5.16	5.18	**4.54**	5.15	5.21	5.06	**4.61**
setup2	5.31	**4.53**	4.96	5.25	5.23	5.15	**4.71**
setup2 multiview	6.14	**4.47**	5.58	6.02	5.87	5.90	**5.36**
setup3 crop 1	7.33	6.93	6.46	**6.12**	8.07	6.56	**5.25**
setup3 crop 2	7.81	**6.69**	**6.05**	7.60	7.79	7.74	7.21
setup4 handheld	5.67	**3.57**	5.37	5.83	5.86	5.88	**5.36**
Mean	5.88	**5.10**	5.19	5.65	6.06	5.58	**5.17**

Methods HDR-GAN, HDRRNN, DeepHDR and Sen et al. are applied to the previously denoised raw sequences by a single image version of the proposed denoising method. For each sequence, we mark in bold the two methods with lowest (better) value.

pipeline. In order to illustrate this fact, we have included an extra column where we compute the PI of a slightly sharpened version of our result (see Fig. 13 to compare the original and sharpened result). This slight post-processing considerably improves the PI value, obtaining a comparable score to the best learning-based method. This suggests that the PI measure is possibly not the best measure to evaluate the capability of an HDR fusion method.

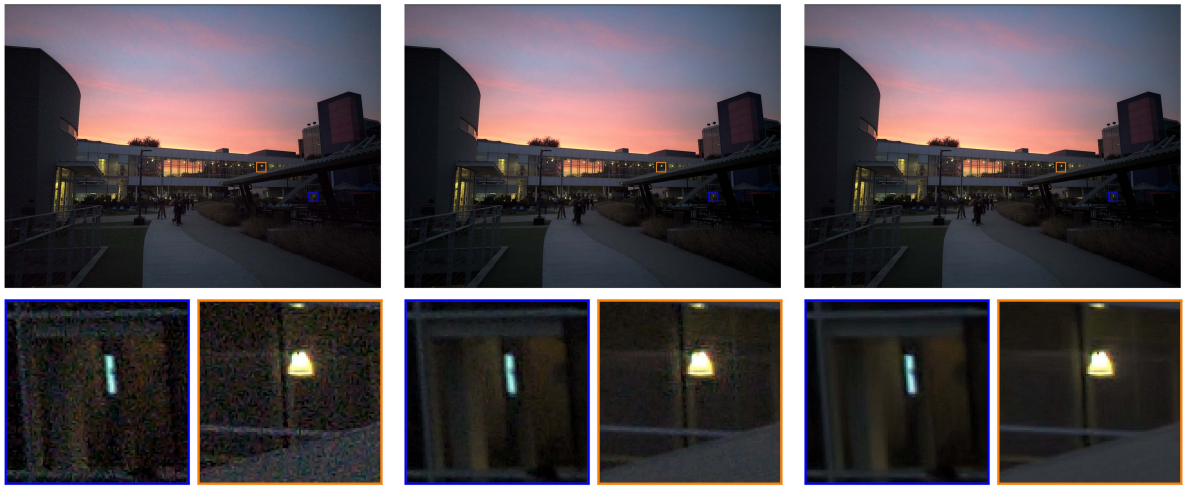
D. Comparison With HDR+

We have tested our method with the HDR+ burst photography dataset [31]. This dataset contains different noisy bursts of dark scenes acquired with the same exposure. Indeed, the use of a burst permits to reduce noise in dark noisy regions and thus safely apply of a subsequent strong enhancement in these regions



(a) Without enhancement (b) With enhancement

Fig. 13. Comparison between the result of our method and after applying a light sharpening. The one on the right has a lower (better) PI value.



(a) Noisy reference (b) HDR+ [31] (c) Ours

Fig. 14. HDR results on a sequence of the HDR+ burst photography dataset [31].

TABLE IV
COMPUTATION TIME BY EACH METHOD FOR THE DENOISING AND HDR OF CROPS OF SIZE 2560 × 2560 OF THE SEQUENCES PROVIDED BY KARAJUZOVIC-HADZIABDIC ET AL. [40], [41]

	Denoise	HDR	Total
HDR-GAN	225,15s	87s	312,15s
Deep HDR	225,15s	38,33s	263,48s
HDRRNN	225,15s	193,5s	418,65s
Sen et al.	225,15s	128,83s	353,98s
Aguerrebere et al.	56,96s		56,96s
Ours	241,16s		241,16s

without enhancing the noise. However, it does not permit to increase the dynamic range. That is, very dark parts will appear quantized after an extreme tonemapping even if the noise was reduced and saturated areas in the very bright regions will remain saturated.

We compare our method with the free implementation of HDR+ given by Monod et al. [54]. Results are shown in Fig. 14. The full burst of 9 images is used as input for both methods. We have kept the same values for all the parameters in our method despite the different nature of the fused data. We observe a similar detail preservation by both methods, while our algorithm is able to remove much more noise.

E. Computation Time Comparison

Table IV compares the average execution time of the different methods for crops of size 2560 × 2560 of the sequences provided by Karajuzovic-Hadziabdic [40], [41]. The measured time includes the whole process, from reading input images to writing the final tonemapped image.

For the methods which do not take into account noise removal, we had to previously denoise each exposure with a single image version of our algorithm. We indicate separately the time of denoising and fusion for these methods. All the methods were run on CPU.

The method by Aguerrebere et al. is the fastest one followed by ours. This is due to the alignment step of our method, which takes an important part of the computation time for images of this size.

V. CONCLUSION

In this paper, we have proposed a patch-based method for the joint denoising and fusion of sequences of RAW images captured with different exposure times. We have used a spatio-temporal criterion to select similar patches along the sequence, over which a weighted PCA-based filtering. The proposed method is able to denoise and fuse the multi-exposed images efficiently in a single step, without the need of recovering each denoised image in the multi-exposure set. Furthermore, the adaptive PCA filtering is able to deal with ghosting artifacts that usually appear due to motion in the scene, which is a common problem in multi-exposed video sequences.

Using both a visual assessment and a numerical evaluation of the results over different datasets, we have shown that the proposed method obtains state-of-the-art fusion results with real RAW data. The method recovers details correctly without over-smoothing or blurring the image, and is capable of preserving the naturalness of the scenes both in saturated and very dark areas.

The current method would benefit from advances in the alignment of dark and noisy images, as well as in the alignment of images of different exposures. The literature in these areas is very scarce and it is in our future work plan.

Another line of research which could lead to progress in this area is the joint application of HDR fusion, denoising, demosaicking and tonemapping of image sequences. The final quality of the processed image has been shown to depend greatly on these stages of the imaging pipeline.

Lastly, we plan to extend the current approach to multi-exposed video sequences, where it might not be possible to take a reference image that is relatively well exposed, leading to an increase of problems in saturated or under-exposed areas.

REFERENCES

[1] C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau, and P. Musé, "A Bayesian hyperprior approach for joint image denoising and interpolation, with an application to HDR imaging," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 633–646, Dec. 2017.

[2] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Muse, "Simultaneous HDR image reconstruction and denoising for dynamic scenes," in *Proc. IEEE Int. Conf. Comput. Photogr.*, 2013, pp. 1–11.

[3] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Best algorithms for HDR image generation. A study of performance bounds," *SIAM J. Imag. Sci.*, vol. 7, no. 1, pp. 1–34, 2014.

[4] A. Ahmad, M.M. Riaz, A. Ghafoor, and T. Zaidi, "Noise resistant fusion for multi-exposure sensors," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5123–5124, Jul. 2016.

[5] A. Oguz Akyüz, "Photographically guided alignment for HDR images," in *Proc. Eurographics*, 2011, pp. 73–74.

[6] A. O. Akyuz and E. Reinhard, "Noise reduction in high dynamic range imaging," *J. Vis. Commun. Image Representation*, vol. 18, no. 5, pp. 366–376, 2007.

[7] D. Alleysson, S. Susstrunk, and J. Herault, "Linear demosaicing inspired by the human system," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 439–449, Apr. 2005.

[8] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[9] B. E. Bayer, "Color imaging array," U.S. Patent 3971065, 1976.

[10] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 334–355.

[11] A. Buades, B. Coll, and J. M. Morel, "A non local algorithm for image denoising," *IEEE Comput. Vis. Pattern Recognit.*, vol. 2, pp. 60–65, 2005.

[12] A. Buades, J. L. Lisani, and M. Miladinović, "Patch based video denoising with optical flow estimation," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2573–2586, Jun. 2016.

[13] A. Buades and J. Duran, "CFA video denoising and demosaicking chain via spatio-temporal patch-based filtering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 4143–4157, Nov. 2020.

[14] F.M. Candocia, "Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1485–1494, Dec. 2003.

[15] L. Cerman and V. Hlavac, "Exposure time estimation for high dynamic range imaging with hand held camera," in *Proc. Comput. Vis. Winter Workshop*, 2006, pp. 76–81.

[16] P. Chaudhari, F. Schirrmacher, A. Maier, C. Riess, and T. Köhler, "Merging-ISP: Multi-exposure high dynamic range image signal processing," in *Proc. DAGM German Conf. Pattern Recognit.*, 2021, pp. 328–342.

[17] G. Chen, C. Chen, S. Guo, Z. Liang, K-Y. K. Wong, and L. Zhang, "HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2502–2511.

[18] M. Colom, A. Buades, and J.-M. Morel, "Nonparametric noise estimation method for raw images," *JOSA A*, vol. 31, no. 4, pp. 863–871, 2014.

[19] K. Dabov et al., "BM3D image denoising with shape-adaptive principal component analysis," in *Proc. Workshop Signal Process. Adaptive Sparse Structured Representations*, 2009.

[20] A. Davy, T. Ehret, J.-M. Morel, P. Arias, and G. Facciolo, "A non-local CNN for video denoising," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 2409–2413.

[21] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.

[22] V. Dewil, J. Anger, A. Davy, T. Ehret, G. Facciolo, and P. Arias, "Self-supervised training for blind multi-frame video denoising," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 2724–2734.

[23] J. Duran and A. Buades, "Self-similarity and spectral correlation adaptive algorithm for color demosaicking," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4031–4040, Sep. 2014.

[24] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles, "Generation of HDR Images in Non-Static Conditions based on Gradient Fusion," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, 2012, pp. 31–37.

[25] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[26] M. Gevrekci and B.K. Gunturk, "On geometric and photometric registration of images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, pp. I-1261–I-1264.

[27] B. Goossens, H. Luong, J. Aelterman, A. Pižurica, and W. Philips, "Realistic camera noise modeling with application to improved HDR synthesis," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–28, 2012.

[28] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. PA Lensch, "Optimal HDR reconstruction with linear digital cameras," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 215–222.

[29] M. Granados, H.-P. Seidel, and H. PA Lensch, "Background estimation from non-time sequence images," in *Proc. Graph. Interface*, 2008, pp. 33–40.

[30] T. Grosch et al., "Fast and robust high dynamic range image generation with camera and object movement," *Vis., Model. Vis.*, vol. 277284, pp. 277–284, 2006.

[31] W. Samuel et al., "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, 2016.

[32] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 267–276, Mar. 1994.

[33] I. Hossain and K. B. Gunturk, "High dynamic range imaging of non-static scenes," *Proc. SPIE*, vol. 7876, 2011, Art. no. 78760P.

[34] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 499–512.

[35] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1163–1170.

[36] J. Im, S. Jang, S. Lee, and J. Paik, "Geometrical transformation-based ghost artifacts removing for high dynamic range image," in *Proc. 18th IEEE Int. Conf. Image Process.*, 2011, pp. 357–360.

[37] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Comput. Graph. Appl.*, vol. 28, no. 2, pp. 84–93, Mar.-Apr. 2008.

[38] N. K. Kalantari and R. Ramamoorthi, "Deep HDR video from sequences with alternating exposures," *Comput. Graph. Forum*, vol. 38, no. 2, pp. 193–205, 2019.

[39] K. Nima Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 144.1–144.12, 2017.

[40] K. Karajuzovic-Hadziabdic, J. H. Telalovic, and R. K. Mantiuk, "Multi-exposure image stacks for testing HDR deghosting methods," Apollo - University of Cambridge Repository, 2017.

[41] K. Karajuzovic-Hadziabdic, J. H. Telalovic, and R. K. Mantiuk, "Assessment of multi-exposure HDR image deghosting methods," *Comput. Graph.*, vol. 63, pp. 1–17, 2017.

[42] E.A. Khan, A.O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proc. Int. Conf. Image Process.*, 2006, pp. 2005–2008.

[43] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger, "Unified HDR reconstruction from raw CFA data," in *Proc. IEEE Int. Conf. Comput. Photogr.*, 2013, pp. 1–9.

[44] M. Lebrun, A. Buades, and J.-M. Morel, "A nonlocal Bayesian image denoising algorithm," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1665–1688, 2013.

[45] D.-K. Lee, R.-H. Park, and S. Chang, "Ghost and noise removal in exposure fusion for high dynamic range imaging," *Int. J. Comput. Graph. Animation*, vol. 4, no. 4, pp. 1–18, 2014.

[46] H. Liu, J. Zhang, and R. Xiong, "CAS: Correlation adaptive sparse modeling for image denoising," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 638–647, 2021.

[47] Z. Liu et al., "ADNet: Attention-guided deformable convolutional network for high dynamic range imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 463–470.

[48] G. Lowe, "Sift-the scale invariant feature transform," *Int. J*, vol. 2, pp. 91–110, 2004.

[49] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising using separable 4D nonlocal spatiotemporal transforms," *Proc. SPEI*, vol. 7870, pp. 787003–787003, 2011.

[50] S. Mann and R. Picard, *Beingundigital'with Digital Cameras*, vol. 1. Cambridge, MA, USA: MIT Media Lab Perceptual, 1994.

[51] S. Mann, C. Manders, and J. Fung, "Painting with looks: Photographic images from video using quantimetric processing," in *Proc. 10th ACM Int. Conf. Multimedia*, 2002, pp. 117–126.

[52] T.-H. Min, R.-H. Park, and S.K. Chang, "Noise reduction in high dynamic range images," *Signal, Image, Video Process.*, vol. 5, no. 3, pp. 315–328, 2011.

[53] T. Mitsunaga and K. Shree Nayar, "Radiometric self calibration," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1999, pp. 374–380.
[54] A. Monod, J. Delon, and T. Veit, "An analysis and implementation of the HDR burst denoising method," *Image Process. Line*, vol. 11, pp. 142–169, 2021.
[55] Y. Niu, J. Wu, W. Liu, W. Guo, and R. WH Lau, "HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions," *IEEE Trans. Image Process.*, vol. 30, pp. 3885–3896, 2021.
[56] S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–12, 2011.
[57] M. Pedone and J. Heikkilä, "Constrain propagation for ghost removal in high dynamic range images," in *Proc. 3rd Int. Conf. Comput. Vis. Theory Appl.*, 2008, pp. 36–41.
[58] F. Peng, M. Zhang, S. Lai, H. Tan, and S. Yan, "Deep HDR reconstruction of dynamic scenes," in *Proc. IEEE 3 rd Int. Conf. Image, Vis. Comput.*, 2018, pp. 347–351.
[59] N. N. Ponomarenko, V. V. Lukin, M. S. Zriakhov, A. Kaarna, and J. T. Astola, "An automatic approach to lossy compression of AVIRIS images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 472–475.
[60] K. R. Prabhakar, S. Agrawal, and R. Venkatesh Babu, "Self-gated memory recurrent network for efficient scalable HDR deghosting," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1228–1239, 2021.
[61] A. Ajdari Rad, L. Meylan, P. Vandewalle, and S. Süsstrunk, "Multidimensional image enhancement from a set of unregistered differently exposed images," *Proc. SPIE*, vol. 6498, 2007, Art. no. 649808.
[62] S. Raman, V. Kumar, and S. Chaudhuri, "Blind de-ghosting for automatic multi-exposure compositing," in *Proc. ACM SIGGRAPH ASIA Posters*, 2009, Art. no. 44.
[63] P. Sen, N. Khademi Kalantari, M. Yaesoubi, S. Darabi, D. B Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203.1–203.11, 2012.
[64] S. Silk and J. Lang, "Fast high dynamic range image deghosting for arbitrary scene motion," in *Proc. Graph. Interface* 2012, pp. 85–92.
[65] M. Tassano, J. Delon, and T. Veit, "DVDNet: A fast network for deep video denoising," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 1805–1809.
[66] M. Tassano, J. Delon, and T. Veit, "FastDVDNet: Towards real-time deep video denoising without flow estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1354–1363.
[67] M. Tico, N. Gelfand, and K. Pulli, "Motion-blur-free exposure fusion," in *Proc. IEEE 17th Int. Conf. Image Process.*, 2010, pp. 3321–3324.
[68] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
[69] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in *Proc. Int. Conf. Central Europe Comput. Graph., Vis. Comput. Vis.*, 2007, pp. 49–56.
[70] O. T. Tursun, A. O. Akyuz, A. Erdem, and E. Erdem, "The state of the art in HDR deghosting: A survey and evaluation," in *Computer Graphics Forum*, vol. 34. Hoboken, NJ, USA: Wiley, 2015, pp. 683–707.
[71] G. Vaksman, M. Elad, and P. Milanfar, "Patch craft: Video denoising by deep modeling and patch matching," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2157–2166.
[72] L. Wang and K.-J. Yoon, "Deep learning for HDR imaging: State-of-the-art and future trends," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8874–8895, Dec. 2022.
[73] 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.
[74] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *J. Graph. Tools*, vol. 8, no. 2, pp. 17–30, 2003.
[75] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 117–132.
[76] Q. Yan et al., "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1751–1760.
[77] Q. Yan et al., "Deep HDR imaging via a non-local network," *IEEE Trans. Image Process.*, vol. 29, pp. 4308–4322, 2020.
[78] S. Yao, "Robust image registration for multiple exposure high dynamic range image synthesis," *Proc. SPIE*, vol. 7870, 2011, Art. no. 78700Q.
[79] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Proc. Pattern Recognit.: 29th DAGM Symp.*, 2007, pp. 214–223.
[80] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," in *Computer Graphics Forum*, vol. 30. Hoboken, NJ, USA: Wiley, 2011, pp. 405–414.