Single Shot High Dynamic Range Imaging Using Piecewise Linear Estimators

Cecilia Aguerrebere, Andrés Almansa, Yann Gousseau, Julie Delon, Pablo Muse

▶ **To cite this version:**

Cecilia Aguerrebere, Andrés Almansa, Yann Gousseau, Julie Delon, Pablo Muse. Single Shot High Dynamic Range Imaging Using Piecewise Linear Estimators. 2014 IEEE International Conference on Computational Photography (ICCP), May 2014, Santa Clara, CA, United States. pp.1-10, 10.1109/ICCPHOT.2014.6831807 . hal-01054831

HAL Id: hal-01054831
https://hal.science/hal-01054831

Submitted on 8 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Single Shot High Dynamic Range Imaging Using Piecewise Linear Estimators

Cecilia Aguerrebere, Andrés Almansa, Yann Gousseau
Télécom ParisTech
46 rue Barrault, F-75634 Paris Cedex 13, France
aguerreb,almansa,gousseau@telecom-paristech.fr

Julie Delon
Université Paris Descartes
75270 Paris Cedex 06, France
julie.delon@parisdescartes.fr

Pablo Musé
IIE, Universidad de la República
Herrera y Reissig 565, 11300 Uruguay
pmuse@fing.edu.uy

Abstract

Building high dynamic range (HDR) images by combining photographs captured with different exposure times present several drawbacks, such as the need for global alignment and motion estimation in order to avoid ghosting artifacts. The concept of spatially varying pixel exposures (SVE) proposed by Nayar et al. enables to capture in only one shot a very large range of exposures while avoiding these limitations. In this paper, we propose a novel approach to generate HDR images from a single shot acquired with spatially varying pixel exposures. The proposed method makes use of the assumption stating that the distribution of patches in an image is well represented by a Gaussian Mixture Model. Drawing on a precise modeling of the camera acquisition noise, we extend the piecewise linear estimation strategy developed by Yu et al. for image restoration. The proposed method permits to reconstruct an irradiance image by simultaneously estimating saturated and under-exposed pixels and denoising existing ones, showing significant improvements over existing approaches.

1. Introduction

The idea of using multiple differently exposed images to capture high dynamic range (HDR) scenes can be traced back to the middle of the 19th century, when the French photographer Gustave Le Gray captured a high dynamic range scene at the sea by combining two differently exposed negatives. This idea was introduced in digital photography by Mann and Picard [9] in 1995. Several methods followed, proposing different ways to combine the images [4, 11, 13, 5].

In the case of a static scene and a static camera, the combination of multiple images is a simple and efficient solution for the generation of HDR images. However, several problems arise when either the camera or the elements in the scene move. Global alignment techniques must be used to align images acquired with a hand-held camera and deghosting methods must be used to correct the artifacts due to object motion. These kind of artifacts are particularly annoying on the fused result.

An alternative to HDR imaging from multiple frames was introduced by Nayar and Mitsunaga in [12]. They propose to perform HDR imaging from a single image using spatially varying pixel exposures (SVE). An optical mask with spatially varying transmittance (see Figure 2) is placed adjacent to a conventional image sensor, thus controlling the amount of light that reaches each pixel. This gives different exposure levels to the pixels according to the given transmittance pattern, allowing a single shot to capture an increased dynamic range compared to that of the conventional sensor. In [7], Hirakawa and Simon argue that different sensitivities are already implied by the different translucencies of the three color filters in a regular Bayer Pattern. They propose a clever demosaicking-inspired algorithm to jointly perform demosaicking and HDR imaging from a single shot, with specially taylored color-filter translucencies.

The greatest advantage of the SVE acquisition method is that it allows HDR imaging from a single image, thus avoiding the need for alignment and motion estimation, which is the main drawback of the classical multi-image approach. Another advantage is that the saturated pixels are not organized in large regions. Indeed, some recent multi-image methods tackle the camera and objects motion problems by taking a reference image and then estimating motion relative to this frame or by recovering information from other frames through local comparison with the reference [17, 2].

Figure 1: Example of the acquisition of an HDR scene using spatially varying pixel exposures. **Left:** Tone mapped HDR scene restored from the raw image. **Right top:** Raw image with spatially varying exposure levels. **Right bottom:** Mask of correctly exposed pixels (white) and under or over exposed pixels (black).

A problem encountered by this approach is the need for inpainting saturated and underexposed regions in the reference frame, since the information is completely lost in those areas. The SVE acquisition strategy prevents from having large saturated regions to inpaint. In general, all scene regions are sampled by at least one of the exposures thus simplifying the inpainting problem.

The main drawback of the SVE acquisition is that, unlike the multi-image approach where all scene regions are assumed to be correctly exposed in at least one of the input images, for the brighter and darker regions of the scene some exposure levels will be either too high or too low and the corresponding pixels will be under or over exposed. Hence, those pixels are unknown and need to be somehow reconstructed. Figure 1 illustrates this problem. It shows an example of an HDR scene and the mask of known and unknown pixel values of a single shot of the scene using SVE. Known pixels (white) are the correctly exposed ones and unknown (black) pixels are those either under or over exposed. Moreover, noise reduction is of particular importance in this kind of acquisition setup since the pixels of the lower exposures tend to be quite noisy (mostly in dark regions) thus producing images with high noise levels.

In the approach proposed by Nayar and Mitsunaga [12], the varying exposures follow a regular pattern as shown in Figure 2. Two methods are proposed to reconstruct the under and over exposed pixels. The so called aggregation approach consists in averaging the local irradiance values produced by the correctly exposed pixels. The interpolation approach consists in using a bi-cubic interpolation to simultaneously retrieve the unknown pixels and denoise the known ones. A generalization of this kind of pixel varying acquisition, and its application to high dynamic range and multi-spectral imaging is presented in [18].

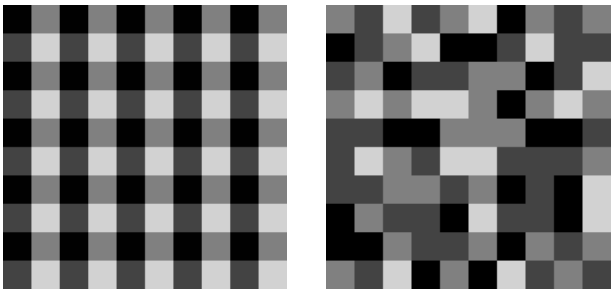

Figure 2: Regular (left) and non-regular (right) optical masks for an example of 4 different filters.

Motivated by the aliasing problems of regular sampling patterns, Schöberl et al. [15] propose to use spatially varying exposures in a non-regular pattern. Figure 2 shows examples of both acquisition patterns. The reconstruction of the irradiance image is then performed using a frequency selective extrapolation algorithm [16] which iteratively generates a sparse model for each image patch as a weighted superposition of the two-dimensional Fourier basis functions. In [14], Schöberl et al. present a practical methodology for the construction of a spatially varying exposures mask with a non-regular pattern.

In this work, we propose a new method to reconstruct the irradiance information of a scene from a single shot acquired with spatially varying pixel exposures following a random pattern. We take advantage of the Gaussian mixture models (GMM), which have been proven accurate at representing natural image patches [19, 8], to reconstruct the unknown pixels and denoise the known ones. The proposed reconstruction method is an extension to the SVE acquisition strategy of the general framework introduced by Yu et al. [19] for the solution of image inverse problems. This allows us to greatly improve the irradiance reconstruction with respect to the previous approaches.

The paper is organized as follows. Section 2 presents the SVE acquisition model. Section 3 introduces the irradiance reconstruction problem and the proposed solution. A summary of the performed experiments is presented in Section 4. Conclusions are presented in Section 5.

2. Spatially varying exposure acquisition model

In this section we introduce a noise model for images captured using the SVE acquisition strategy. This image model is afterward used to develop the irradiance reconstruction method.

As presented in [12, 18, 14], an optical mask with spatially varying transmittance can be placed adjacent to a conventional image sensor to give different exposure levels to the pixels. This optical mask does not change the acqui-

sition process of the sensor, whether using a conventional CCD or CMOS sensor. The main noise sources for this kind of sensors are: the Poisson photon shot noise, which can be approximated by a Gaussian distribution with equal mean and variance; the thermally generated readout noise, which is modeled as an additive Gaussian noise; the spatially varying gain given by the photo response non uniformity (PRNU); dark currents and quantization noise [1, 3]. Therefore, we consider the following noise model for the non saturated nor under-exposed raw pixel value Z_p at position p

$$\mathrm{Z}_p \sim \mathcal{N}(go_pa_p\tau\mathrm{F}_p + \mu_R, g^2o_pa_p\tau\mathrm{F}_p + \sigma_R^2), \quad (1)$$

where g is the camera gain, o_p is the variable gain due to the optical mask, a_p models the PRNU factor, τ is the exposure time, F_p is the irradiance reaching pixel p, μ_R and σ_R^2 are the readout noise mean and variance. Dark currents and quantization noise are neglected. Some noise sources not modeled in [3], such as blooming, might have a considerable impact in the SVE acquisition strategy and should be considered in a more accurate image modeling.

Two main aspects must be defined for the SVE acquisition strategy. One is the number of different filters to be used, i.e. the different exposure levels to capture. This is related to the problem of how many exposure times should be used in the classical HDR acquisition strategy. The solution to this problem depends on the scene. Since the acquisition using SVE uses an a priori fixed optical mask, the number of different exposures is fixed. In general, 2 to 4 images are used for HDR imaging. An optical mask with 4 different exposure levels appears a reasonable choice [12].

The second choice is whether the spatial distribution of the different filters is done randomly or with a regular pattern. This determines the way the scene irradiance is sampled. Figure 2 shows examples of the two sampling strategies. This point is important in the acquisition strategy since, due to unknown under and over exposed pixels, some regions of the image will almost certainly be sub-sampled and some kind of interpolation will be needed to retrieve these pixels values. If the sampling pattern is regular, aliasing artifacts will appear due to the characteristics of the spectrum of the pattern (delta functions at the sampling frequencies). On the contrary, the spectrum of a random pattern is concentrated in a single delta and has negligible values for the rest of the frequencies, thus avoiding aliasing. This fact led us to choose a random pattern to perform the acquisition.

3. Irradiance reconstruction

In order to reconstruct the dynamic range of the scene we need to solve an inverse problem, that is, to find the irradiance values from the input pixel values. Several widely known methods solve image inverse problems decomposing the image into patches so as to take advantage of accurate models developed to represent patches. These models assume that the patches are redundant in the image and that all patches can be represented by a limited number of classes. In particular, Yu et al. [19] introduced a general framework to solve this kind of problems using piecewise linear estimators (PLE). They propose to decompose the image into patches and model these patches using a GMM. Then an expectation-maximization-like iterative procedure is introduced to alternately reconstruct the patches and update the GMM parameters. In this work we propose to use an extension of the work by Yu et al. [19], also based on a GMM for image patches, which is adapted to the acquisition model with variable exposure.

3.1. An inverse problem

The problem we want to solve is that of estimating the irradiance image $\mathbf{F}$ from the input image $\mathbf{Z}$, knowing the exposure levels and the camera parameters. Let us consider Y_p the normalization of the input pixel Z_p to the irradiance domain

$$\mathrm{Y}_p = \frac{\mathrm{Z}_p - \mu_R}{go_pa_p\tau}. \quad (2)$$

We take into account the effect of saturation and underexposure by introducing the exposure degradation factor U_p given by

$$\mathrm{U}_p = \begin{cases} 1 & \text{if } \mu_R < \mathrm{Z}_p < z_{sat}, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

with z_{sat} equal to the pixel saturation value. From (1), Y_p can be modeled as

$$\mathrm{Y}_p \sim \mathcal{N}\left(\mathrm{U}_p\mathrm{F}_p, \frac{g^2o_pa_p\tau\mathrm{U}_p\mathrm{F}_p + \sigma_R^2}{(go_pa_p\tau)^2}\right), \quad (4)$$

Notice that (4) is the distribution of Y_p for a given U_p, since U_p is itself a random variable that depends on Z_p. The exposure degradation factor must be included in (4) since the variance of the over or under exposed pixels no longer depends on the irradiance F_p but is only due to the readout noise σ_R^2.

Then the problem of irradiance estimation can be stated as retrieving $\mathbf{F}$ from the image $\mathbf{Y}$, which implies denoising the known Y_p pixel values ($\mathrm{U}_p = 1$) and estimating the completely unknown ones ($\mathrm{U}_p = 0$).

3.2. Piecewise linear estimators for noise with variable variance

In order to reconstruct $\mathbf{F}$ from $\mathbf{Y}$ we extend the general framework proposed by Yu et al. [19], by adapting it to the noise present in the raw irradiance values given by (4).

Patch model Based on [19], we decompose the irradiance image $\mathbf{Y}$ into overlapping patches $\mathbf{y}_i$ of size $\sqrt{N} \times \sqrt{N}$, $i = 1, \ldots, I$ with I the number of patches in the image. From (4), each patch $\mathbf{y}_i$ taken as a column vector of size $N \times 1$ can be modeled according to

$$\mathbf{y}_i = \mathbf{U}_i \mathbf{f}_i + \mathbf{\Sigma}_{w_i}^{1/2} \mathbf{w}_i, \tag{5}$$

where the degradation operator $\mathbf{U}_i$ is a $N \times N$ diagonal matrix with the diagonal elements equal to the degradation image $\mathbf{U}$ restricted to the patch i, $\mathbf{f}_i$ is the patch on the irradiance image we seek to estimate, $\mathbf{\Sigma}_{w_i}$ is a $N \times N$ diagonal matrix with the j-th diagonal element given by

$$(\mathbf{\Sigma}_{w_i})_j = \frac{g^2 o_j a_j \tau (\mathbf{U}_i \mathbf{f}_i)_j + \sigma_R^2}{(g o_j a_j \tau)^2}, \tag{6}$$

where $(\mathbf{U}_i \mathbf{f}_i)_j$ is the j-th element of vector $\mathbf{U}_i \mathbf{f}_i$ and $\mathbf{w}_i$ is a Gaussian noise with zero mean and identity covariance matrix.

A GMM is chosen to describe image patches with K Gaussian distributions $\mathcal{N}(\mu_k, \mathbf{\Sigma}_k)_{1 \leq k \leq K}$ parametrized by their means μ_k and covariance matrices $\mathbf{\Sigma}_k$. Each patch $\mathbf{f}$ is assumed to be drawn independently from one of these Gaussians, whose probability density functions are given by

$$p(\mathbf{f}) = \frac{1}{(2\pi)^{N/2} |\mathbf{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{f} - \mu_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{f} - \mu_k) \right). \tag{7}$$

To simplify notation, we consider in the following $\mu_k = 0$ $\forall k = 1, \ldots, K$ since we can always center the patches with respect to their means.

Patch reconstruction Assuming that the class k and the corresponding Gaussian parameters μ_k and $\mathbf{\Sigma}_k$ are known, we propose to estimate the patch $\tilde{\mathbf{f}}_i$ as the *linear estimator* $\tilde{\mathbf{W}} \mathbf{y}_i$ that minimizes the Bayesian mean squared error

$$\tilde{\mathbf{W}} = \arg\min_{\mathbf{W}} \mathrm{E}[(\mathbf{W} \mathbf{y}_i - \mathbf{f}_i)^2]. \tag{8}$$

Notice that since $\mathbf{f}_i$ is a random variable (following Model (7)), the expectation operator is with respect to the joint probability density function $p(\mathbf{y}_i, \mathbf{f}_i)$.

The linear estimator $\mathbf{W} \mathbf{y}_i$ that minimizes the Bayes quadratic risk must satisfy

$$\mathrm{E}[(\mathbf{W} \mathbf{y}_i - \mathbf{f}_i) \mathbf{y}_i^T] = 0, \tag{9}$$

thus (see Appendix A)

$$\tilde{\mathbf{W}} = \mathrm{E}[\mathbf{f}_i \mathbf{y}_i^T] \mathrm{E}[\mathbf{y}_i \mathbf{y}_i^T]^{-1} \tag{10}$$

$$= \mathbf{\Sigma}_k \mathbf{U}_i^T (\mathbf{U}_i \mathbf{\Sigma}_k \mathbf{U}_i^T + \mathbf{\Sigma}_{w_i})^{-1}. \tag{11}$$

Hence we propose to estimate $\tilde{\mathbf{f}}_i$ as

$$\tilde{\mathbf{f}}_i^k = \mathbf{W}_{k,i} \mathbf{y}_i, \tag{12}$$

where $\mathbf{W}_{k,i}$ is the Wiener filter

$$\mathbf{W}_{k,i} = \mathbf{\Sigma}_k \mathbf{U}_i^T (\mathbf{U}_i \mathbf{\Sigma}_k \mathbf{U}_i^T + \mathbf{\Sigma}_{w_i})^{-1}. \tag{13}$$

Notice that the same estimator is obtained if we compute the maximum of the posterior probability $p(\mathbf{f}_i | \mathbf{y}_i, \mathbf{\Sigma}_k)$ ignoring the dependence of $\mathbf{\Sigma}_{w_i}$ on $\mathbf{f}_i$.

In the original framework studied by Yu et al. [19], the noise is assumed to have constant variance (σ^2, i.e. $\mathbf{\Sigma}_{w_i} = \sigma^2 Id$). In this simpler case, the linear estimator (13) fully corresponds to the MAP estimator and can be shown to minimize the Bayesian quadratic risk and not only the risk among linear estimators.

As defined in (6), the noise covariance matrix $\mathbf{\Sigma}_{w_i}$ depends on the irradiance $\mathbf{f}_i$. An iterative procedure could be used to alternatively compute $\mathbf{f}_i$ and $\mathbf{\Sigma}_{w_i}$ from (12). We opt here to compute $\mathbf{\Sigma}_{w_i}$ directly from the input samples, i.e., taking $\mathbf{f}_i = \mathbf{y}_i$, since this approximation of the noise variance was proved robust in previous irradiance estimators [5, 3].

Class selection and update In the previous step, following (12), the class k and its parameters μ_k and $\mathbf{\Sigma}_k$ are supposed to be known. In practice, they must be determined.

The best model $\tilde{k}_i$ is selected as the one maximizing the posterior probability $p(\mathbf{f} | \mathbf{y}_i, \mathbf{\Sigma}_k)$ over k assuming $\mathbf{f} = \tilde{\mathbf{f}}_i^k$

$$\tilde{k}_i = \arg\max_k \left(\log p(\mathbf{y}_i | \tilde{\mathbf{f}}_i^k, \mathbf{\Sigma}_k) + \log p(\tilde{\mathbf{f}}_i^k, \mathbf{\Sigma}_k) \right) \tag{14}$$

$$= \arg\min_k \Big((\mathbf{y}_i - \mathbf{U}_i \tilde{\mathbf{f}}_i^k)^T \mathbf{\Sigma}_{w_i}^{-1} (\mathbf{y}_i - \mathbf{U}_i \tilde{\mathbf{f}}_i^k) \tag{15}$$

$$+ (\tilde{\mathbf{f}}_i^k)^T \mathbf{\Sigma}_k^{-1} \tilde{\mathbf{f}}_i^k + \log |\mathbf{\Sigma}_k| \Big). \tag{16}$$

Given that the Gaussian parameters μ_k and $\mathbf{\Sigma}_k$ are unknown, and following [19], an iterative procedure is proposed to alternately compute $(\tilde{\mathbf{f}}_i^k, \tilde{k}_i)$ and update the GMM parameters. The Gaussian parameters for the K classes are first initialized from synthetic images (see [19] for a detailed explanation of the initialization procedure). At the *estimation step*, $\tilde{\mathbf{f}}_i$ and $\tilde{k}_i$ are computed according to equations (12) and (16) respectively. At the *model estimation step*, the classes parameters μ_k and $\mathbf{\Sigma}_k$ are updated by computing the corresponding maximum likelihood estimators from the patches assigned to each class (the $\tilde{k}_i$ assigned at the previous step),

$$\tilde{\mu}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \tilde{\mathbf{f}}_i, \quad \tilde{\mathbf{\Sigma}}_k = \frac{1}{|C_k|} \sum_{i \in C_k} (\tilde{\mathbf{f}}_i - \tilde{\mu}_k)(\tilde{\mathbf{f}}_i - \tilde{\mu}_k)^T, \tag{17}$$

with C_k the set of all patches assigned to class k and $|C_k|$ its cardinality.

The covariance matrix $\tilde{\mathbf{\Sigma}}_k$ may not be well conditioned as a result, for example, of a small number of patches in the class. For this reason a regularization term ε is added to ensure the correct inversion of the matrix [19] ($\tilde{\mathbf{\Sigma}}_k = \tilde{\mathbf{\Sigma}}_k + \varepsilon Id$).

At convergence, the proposed method determines a GMM that represents the set of image patches, it assigns each patch to its corresponding class and restores it accordingly.

The final step of the method consists in combining all the restored patches to reconstruct the image. As it is classical with patch-based methods, the value of each pixel in the final image is the average of the values the pixel takes in all the restored patches that contain it.

The proposed approach is summarized in Algorithm 1.

Algorithm 1: Summary of the proposed method

```
Compute the irradiance image Y from the input image Z using (2).
Compute the degradation mask U from the input image Z using (3).
Decompose Y and U into overlapping patches.
Initialize the K Gaussian parameters μ_k and Σ_k as in [19].
for it=1 to max_its do
    for all patches do
        Compute f̃_i using (12).
        Compute k̃_i using (16) assuming f = f̃_i^k.
    end
    Update μ_k and Σ_k using (17).
    Combine all restored patches to generate the reconstructed image.
end
```

Important algorithm precisions Following [19], the input image is decomposed into regions of size 128×128 and the proposed approach is applied to each region separately. Regions are half-overlapping to avoid boundary effects. Because the image content is more coherent semi-locally than globally, this treatment allows for a better reconstruction with a fixed number of classes K. This semi-local treatment is especially important in the case of HDR images, where the considered dynamic range may be very high and the number of classes needed to represent the image treated as a whole would be very large. In [19], the authors show that 20 classes gives a good trade-off between performance and computational cost. We used $K = 20$ in all our experiments. The algorithm is found to converge in 3 to 4 iterations.

4. Experiments

The proposed reconstruction method was thoroughly tested in several synthetic and real data examples. A summary of the results is presented in this section.

4.1. Synthetic data

Experiments using synthetic data are carried out in order to be able to compare the reconstruction obtained by the proposed method and previous ones from the literature against a ground-truth. This is not possible (or highly prone to errors) using real data. For this purpose, sample images are generated according to Model (1) using the HDR images in Figures 3 and 4 as ground-truth. Both a random and a regular pattern with four equiprobable exposure levels are simulated. For the *lamp* example (Figure 3), the exposure levels are set to $o = \{1, 2, 5, 10\}$, and the exposure time is set to $\tau = 1/250$ seconds. For the *bridge* example (Figure 4), the exposure levels are set to $o = \{1, 10, 20, 40\}$, and the exposure time is set to $\tau = 1/500$ seconds. For both examples, the camera parameters are those of a Canon 400D camera set to ISO 200 [3] ($g = 0.66$, $\sigma_R^2 = 17$, $\mu_R = 256$, $z_{sat} = 4057$). A patch size of 8×8 is used for the *lamp* example and a size of 6×6 for the *bridge* example. In both cases the parameter ϵ is set to 5.

Figure 3 shows the results obtained by the proposed method and by Schöberl et al. [15] for the random pattern, as well as the results obtained by the bi-cubic interpolation proposed by Nayar et Mitsunaga [12] using the regular pattern for the *lamp* example. Three extracts of the image are shown together with their corresponding masks of known (white) and unknown (black) pixels. The percentage of unknown pixels for the first extract is 65% (it is nearly the same for both the regular and non-regular pattern). For the other two extracts most of the pixels are known (99%) so that the proposed method mostly performs denoising in these extracts. Table 1 shows the PSNR values obtained in each extract by each method. The proposed method manages to correctly reconstruct the irradiance information from the input samples. Moreover, its denoising performance is much better than both those of Schöberl et al. and Nayar and Mitsunaga, giving a similar reconstruction quality on the unknown areas.

Figure 4 shows on the right the result obtained by the proposed method for the full test image. On the left, it shows extracts of the results obtained by the proposed approach and by Schöberl et al. [15] for the random pattern as well as the results obtained by the bi-cubic interpolation proposed by Nayar et Mitsunaga [12] using the regular pattern for the *bridge* example. Table 1 shows the PSNR values obtained in each extract by each method. This example shows a quite extreme case in terms of noise. The extracts shown in the second and third rows correspond to

Figure 3: **Synthetic data. First column (top to bottom):** Ground-truth with indicated extracts, full image result obtained by the proposed approach, full image result by Schöberl et al. [15] **Second to fifth column (left to right):** Extracts of the ground-truth, result by the proposed approach, Schöberl et al. [15], Nayar and Mitsunaga [12]. **Sixth column:** Random (top) and regular (bottom) mask for each extract. Black represents unknown and white known pixels. The percentage of unknown pixels for the first extract is 65% (it is nearly the same for both the regular and non-regular pattern). For the other two extract most pixels are known (99%) so that the proposed method mostly performs denoising in these extracts.

	PSNR (dB)		
Lamp	extract 1 (green)	extract 2 (blue)	extract 3 (red)
Proposed method	35.8	50.1	41.9
Schöberl et al.	34.6	43.2	37.0
Nayar and Mitsunaga	35.9	43.9	35.4
Bridge	extract 1 (green)	extract 2 (blue)	extract 3 (red)
Proposed method	30.6	29.1	41.0
Schöberl et al.	25.1	22.5	34.4
Nayar and Mitsunaga	31.3	18.5	31.4

Table 1: PSNR values for the extracts in Figures 3 and 4.

quite dark regions where the signal to noise ratio of the samples is very low, specially for the lower exposure levels. In this extreme conditions, the reconstruction capacity of the proposed method clearly outperforms that of the compared methods.

We observed that for synthetic scenes with a very high dynamic range (e.g. 17 stops), the reconstructed HDR images could present some artifacts. This limitation never occurred in the experiments using real data that we conducted. We suspect that the Gaussian mixture model used in the PLE approach is not fully adapted when the dynamic range of image patches is too large. We are currently working on a refinement of the stochastic model taking into account this specificity.

4.2. Real data

The feasibility of the SVE random pattern has been shown in [14] and that of the SVE regular pattern in [18]. Nevertheless, these acquisition systems are still not available for general usage. However, as stated in Section 2, the only variation between the classical and the SVE acquisition is the optical filter, i.e. the amount of light reaching each pixel. Hence, the noise at a pixel p captured using SVE with an optical gain factor o_p and exposure time τ/o_p and a pixel captured with a classical camera using exposure time τ should be very close. We take advantage of this

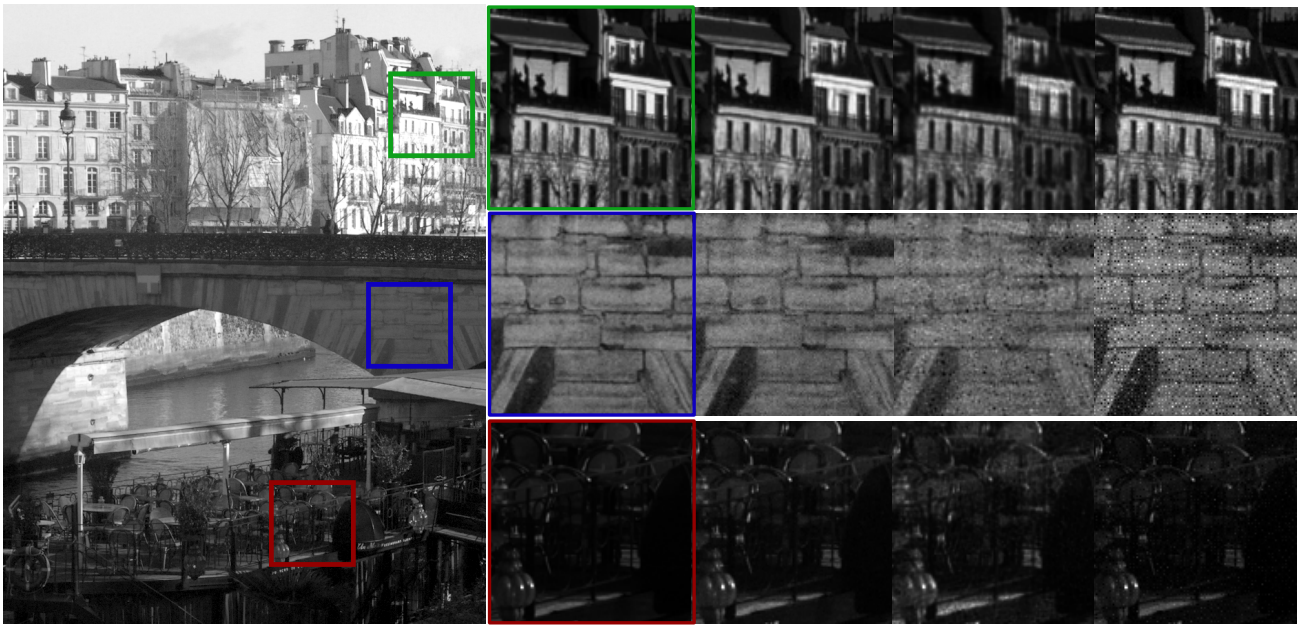

Figure 4: **Synthetic data. Left:** Result obtained by the proposed method for the full test image with indicated extracts. **Right (left to right):** Ground-truth, result by the proposed approach, Schöberl et al. [15], Nayar and Mitsunaga [12]. The extracts shown in the second and third rows correspond to quite dark regions where the signal to noise ratio of the samples is very low, specially for the lower exposure levels. In this extreme conditions, the reconstruction capacity of the proposed method clearly outperforms that of the compared methods.

fact in order to evaluate the reconstruction performance of the proposed approach using real data. For this purpose we generate an SVE image drawing pixels at random from four raw images acquired with different exposure times. The four different exposure times simulate the different filters of the SVE optical mask. The images are acquired using a remotely controlled camera and a tripod so as to be perfectly aligned. Otherwise, artifacts may appear from the random sampling of the four images to composite the SVE frame. Notice that the SVE image thus obtained is very similar to the one obtained if such an optical filter was placed adjacent to the sensor.

This protocol does not allow us to take scenes with moving objects. Let us emphasize, however, that using a real SVE device, this, as well as the treatment of moving camera, would of course not be an issue.

Given the procedure we use to generate the SVE image form the input raw images, the Bayer pattern of the latter is kept in the generated SVE image. The proposed irradiance reconstruction method is thus applied to the raw SVE image with an overlap of $\sqrt{N} - 2$ between patches (i.e. a shift of two pixels) in order to compare pixels of the corresponding color channels. A patch size of 6×6 is used for the examples in Figures 6 and 7, and a patch size of 8×8 for the example in Figure 5. The ϵ parameter is set to 5 for all experiments. The demosaicking method by Adams and Hamilton [6] is then used to obtain a color image from the reconstructed irradiance. To display the results we use the tone mapping technique by Mantiuk *et al*. [10].

A comparison against the methods by Nayar and Mitsunaga and Schöberl et al. is not presented since they do not precise in their works how to treat raw images with a Bayer pattern (how to treat color) and therefore an adaptation of their methods should be made in order to process our data.

Figures 5 to 7 show the results obtained in three real scenes, together with the input raw images and the mask of known (white) and unknown (black) pixels[1]. Recall that among the unknown pixels, some of them correspond to saturated pixels and some of them to under exposed pixels. The proposed method manages to correctly reconstruct the unknown pixels even in extreme conditions where more than 70% of the pixels are missing.

These examples show the capacity of the proposed approach to reconstruct the irradiance information in both very dark and bright regions simultaneously. See for instance the example in Figure 6, where the dark interior of the building (which can be seen through the windows) and the highly illuminated part of another building are both cor-

[1] A reduced version of the images is included in the pdf due to file size restrictions. Originals are available at `http://perso.telecom-paristech.fr/~gousseau/single_shot_hdr`

Figure 5: **Real data. Left:** Tone mapped HDR image obtained by the proposed approach (11.4 stops). **Middle top:** Raw image with spatially varying exposure levels. **Middle bottom:** Mask of unknown (black) and known (white) pixels. In the regions with unknown pixels, the percentage of missing pixels varies between 25% to 40%. **Right:** Extracts of the scene

rectly reconstructed (please consult the pdf version of this article for better visualization).

5. Conclusions

In this work, we have proposed a novel approach for the generation of HDR images from a single shot using spatially varying pixel exposures. The SVE acquisition strategy allows the creation of HDR images without the drawbacks of multi-image approaches, such as the need for global alignment and motion estimation to avoid ghosting problems. Nevertheless, existing restoration methods from HDR SVE images lacked a mechanism for jointly denoising and interpolating the image effectively. The proposed method follows a recent and popular trend in image restoration, modeling patch distributions by Gaussian Mixture Models. We make use of the piecewise linear estimators proposed by Yu et al. [19], and we extend the approach to the case of a complete camera noise model, where noise variance is both variable and dependent on the signal.

The proposed method could also be applied to reconstruct the irradiance map when using the acquisition technique proposed by Hirakawa and Simon [7]. This strategy can be seen as a very practical implementation of SVE, with certain constraints on the optical filters.

The resulting method manages to simultaneously denoise and reconstruct the missing pixels, even in the presence of (possibly complex) motions, improving the results obtained by existing methods. Examples with real data acquired in very similar conditions to those of the SVE acquisition show the high capabilities of the proposed approach. The presence of artifacts was noted in the HDR reconstruction of synthetic scenes with a very high dynamic range. This limitation never occurred in the experiments using real data that we have conducted. We suspect that the Gaussian mixture model used in the PLE approach is not fully adapted when the dynamic range of image patches is too large and we are currently working on a refinement of the stochastic model taking into account this specificity. More precisely, we are currently developing a famework generating the PLE strategy in the spirit of the recent state-of-the-art denoising method [8], but allowing the treatment of missing data.

Let us conclude by observing that, in the proposed approach, both saturated and under-exposed pixels are equally treated as missing pixels. However, valuable information exists in the fact that a pixel is either saturated or under-exposed [3]. Hence, future work should explore the possibility of different treatments for each of these two kind of pixels. It would not be surprising that this strategy, if well implemented, may improve current results.

A. Appendix

We look for the linear estimator $\mathbf{W}\mathbf{y}_i$ that minimizes the Bayes quadratic risk. Thus $\mathbf{W}$ must satisfy $\mathrm{E}[(\mathbf{W}\mathbf{y}_i - \mathbf{f}_i)\mathbf{y}_i^T] = 0$ and we have (the dependence on the patch position i is avoided to simplify notation)

$$\mathbf{W} = \mathrm{E}[\mathbf{f}\mathbf{y}^T]\mathrm{E}[\mathbf{y}\mathbf{y}^T]^{-1}. \tag{18}$$

From the patch model (5), the (p, q) element of matrix $\mathrm{E}[\mathbf{f}\mathbf{y}^T]$ is given by

$$\mathrm{E}[\mathbf{f}\mathbf{y}^T]_{p,q} = \mathrm{E}[\mathbf{f}_p(\mathbf{U}\mathbf{f})_q^T + \mathbf{f}_p(\mathbf{\Sigma}_w^{1/2}\mathbf{w})_q^T] \tag{19}$$

$$= (\mathbf{\Sigma}_k\mathbf{U}^T)_{p,q}, \tag{20}$$

Figure 6: **Real data. Left:** Tone mapped HDR image obtained by the proposed approach (15.6 stops). **Right top:** Extracts of the scene. **Right bottom:** Mask of unknown (black) and known (white) pixels. In the brightest part of the building 73% of the pixels are unknown. Despite this fact, the reconstructed HDR image does not exhibit any visible artifact.

Figure 7: **Real data. Left:** Tone mapped HDR image obtained by the proposed approach (13.4 stops). **Right top:** Raw image with spatially varying exposure levels. **Right bottom:** Mask of unknown (black) and known (white) pixels. In the lamp area 70% percent of the pixels are unknown.

since $\mathbf{w}_q$ is independent of $\mathbf{f}_p$ and has zero mean. From the patch model (5), the (p, q) element of matrix $\mathrm{E}[\mathbf{y}\mathbf{y}^T]$ is given by

$$\mathrm{E}[\mathbf{y}\mathbf{y}^T]_{p,q} = \mathrm{E}[(\mathbf{U}\mathbf{f})_p(\mathbf{f}\mathbf{U})_q^T + (\mathbf{U}\mathbf{f})_p(\boldsymbol{\Sigma}_w^{1/2}\mathbf{w})_q^T \quad (21)$$

$$+ (\boldsymbol{\Sigma}_w^{1/2}\mathbf{w})_p(\mathbf{U}\mathbf{f})_q^T + (\boldsymbol{\Sigma}_w^{1/2}\mathbf{w})_p(\boldsymbol{\Sigma}_w^{1/2}\mathbf{w})_q^T] \quad (22)$$

$$= (\mathbf{U}\boldsymbol{\Sigma}_k\mathbf{U}^T)_{p,q} + (\boldsymbol{\Sigma}_w)_{p,q}. \quad (23)$$

Hence we have,

$$\mathbf{W} = \boldsymbol{\Sigma}_k\mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}_k\mathbf{U}^T + \boldsymbol{\Sigma}_w)^{-1}. \quad (24)$$

References

[1] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Study of the digital camera acquisition process and statistical modeling of the sensor raw data. *Preprint HAL* `http://hal.archives-ouvertes.fr/docs/00/73/35/38/PDF/camera_model.pdf`, 2012.

[2] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Simultaneous HDR image reconstruction and denoising for dynamic scenes. In *Computational Photography (ICCP), 2013 IEEE International Conference on*, pages 1–11, 2013.

[3] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Best algorithms for HDR image generation. a study of performance bounds. *SIAM Journal on Imaging Sciences*, 7(1):1–

34, 2014.

[4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pages 369–378, 1997.

[5] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch. Optimal HDR reconstruction with linear digital cameras. In *CVPR*, pages 215–222, 2010.

[6] J. Hamilton and J. Adams. Adaptive color plan interpolation in single sensor color electronic camera. US Patent 5,629,734, 1997.

[7] K. Hirakawa and P. Simon. Single-shot high dynamic range imaging with conventional camera hardware. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1339–1346, Nov 2011.

[8] M. Lebrun, A. Buades, and J. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013.

[9] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

[10] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3):68:1–68:10, Aug. 2008.

[11] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *CVPR*, pages 1374–1380, 1999.

[12] S. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 472–479, Jun 2000.

[13] M. A. Robertson, S. Borman, and R. L. Stevenson. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *J. Electronic Imaging*, 12(2):219–228, 2003.

[14] M. Schöberl, A. Belz, A. Nowak, J. Seiler, A. Kaup, and S. Foessel. Building a high dynamic range video sensor with spatially nonregular optical filtering, 2012.

[15] M. Schöberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup. High dynamic range video by spatially non-regular optical filtering. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2757–2760, 2012.

[16] J. Seiler and A. Kaup. Complex-valued frequency selective extrapolation for fast image and video signal extrapolation. *Signal Processing Letters, IEEE*, 17(11):949–952, 2010.

[17] 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.*, 31(6):203:1–203:11, Nov. 2012.

[18] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. *IEEE Transactions on Image Processing*, 99, Mar 2010.

[19] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *Image Processing, IEEE Transactions on*, 21(5):2481–2499, 2012.