

Simultaneous HDR image reconstruction and denoising for dynamic scenes

Cecilia Aguerrebere, Julie Delon, Yann Gousseau
Télécom ParisTech
46 rue Barrault, F-75634 Paris Cedex 13
aguerreb@telecom-paristech.fr
delon@telecom-paristech.fr
gousseau@telecom-paristech.fr

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

Abstract

High dynamic range (HDR) images are usually generated by combining multiple photographs acquired with different exposure times. This approach, while effective, suffers from various drawbacks. The irradiance estimation is performed by combining, for each pixel, different exposure values at the same spatial position. This estimation scheme does not take advantage of the redundancy present in most images. Moreover, images must be perfectly aligned and objects must be in the exact same position in all frames in order to combine the different exposures. In this work, we propose a new HDR image generation approach that simultaneously copes with these problems and exploits image redundancy to produce a denoised result. A reference image is chosen and a patch-based approach is used to find similar pixels that are then combined for the irradiance estimation. This patch-based approach permits to obtain a denoised result and is robust to image misalignments and object motions. Results show significant improvements in terms of noise reduction over previous HDR image generation techniques, while being robust to motion and changes between the exposures.

1. Introduction

There are physical limitations on the maximal variations of luminosity that a camera sensor can measure. These limitations are due to the fact that each sensor has a finite capacity : above a given quantity of incident photons, the sensor saturates. This phenomenon is well known by photographers and is the reason of backlighting in images. In order to get details simultaneously in shadows and highlights, an efficient technique consists in mixing several images of the same scene taken with different exposure times, creating in this way an High Dynamic Range (HDR) image. Indeed, bright regions can be recovered from short exposures while

dark ones are well captured by longer exposures. The main difficulties of this exposure fusion come from the possible camera motion and from the object motion in the scene. For instance, a hand-held camera will produce misalignments across the frames, while moving objects are the cause of ghosting artifacts.

There is a broad literature dedicated to the creation of HDR images for completely static scenes (the camera and the objects of the scene are fixed) [20, 8, 23, 19, 10, 1]. This problem is simplified but nevertheless particularly interesting. Knowing the physical model of the camera sensor and working with raw data, the most recent approaches [19, 10, 1] model the HDR creation as a statistical estimation problem, for which the maximum likelihood estimator is shown to be optimal [10, 1]. While powerful from the theoretical point of view, in practice these approaches greatly suffer from small camera misalignments and cannot cope with moving objects, which makes them hardly reliable in non controlled environments.

Several approaches have been proposed in the literature to overcome these limitations. Most of these approaches start by globally registering the frames in order to compensate for global camera motion. As a result, the background of the scene can be seen as static, while dynamic objects remain unregistered. The first kind of methods makes use of optical flow techniques in order to compensate the motion across the different frames, generally after a global exposure compensation between the frames [17, 3, 30]. Unfortunately, explicit motion estimation remains an ill-posed and ambiguous problem (especially in the presence of missing regions), and such methods are exposed to alignment errors. A second variety of deghosting methods aims at reducing the influence of pixels of moving objects on the HDR estimation. These pixels can be detected in a first step and completely discarded for the estimation [11, 16, 27, 9]. Alternatively, the pixel weights in the irradiance estimation can be adapted in order to decrease the influence of suspicious pixels [18, 22, 24, 15]. For a recent and complete

review on ghost detection and removal, see [28].

Another limitation of HDR imaging is that images created from different exposures tend to be noisy. This comes from the fact that the HDR estimation is generally performed on a per-pixel basis (possibly after registration). For a given pixel, the estimation only relies on the observed values at the same location for all the available exposures. Moreover, this number is reduced if a moving object occurs or if the pixel is over or underexposed in some frames. The recent work by Heo *et al.* [15] proposes a first solution to this problem, making use of a bilateral filtering during the frame fusion step. However, such a local filtering cannot cope with large motions.

In this paper, we propose to take advantage of the strong self-similarity property of images in order to solve the HDR estimation problem. More precisely, we rely on the redundancy of image patches. This allows to correctly estimate the irradiance at each pixel while avoiding the fastidious image registration and motion estimation (or detection) steps. In particular, the patch similarities are sufficiently robust to handle camera motion and the possibly large motion of objects. As an important by-product, the result of our non local HDR algorithm is much less noisy than the results of state-of-the-art HDR approaches. This denoising capacity is obviously true for the background scene, but also for objects in motion, whose different occurrences through the exposures are correctly retrieved thanks to the patch-based approach.

The paper is organized as follows. We start in Section 2 with a short reminder of the state-of-the-art static HDR estimation techniques. In Section 3 we describe the different steps of our non local HDR estimation approach. Experiments and comparisons with state-of-the-art approaches are displayed in Section 4. Conclusions are presented in Section 5.

2. HDR imaging for static scenes and camera

Several methods have been proposed [20, 8, 23, 19, 10, 1] to solve the HDR image generation problem following a pixel-wise approach in the case of static scenes. The basic idea, common to all of them, consists in combining T images acquired with exposure times $\tau_1, \dots, \tau_T$. The irradiance C_p at each pixel p is then computed as a weighted average of the irradiance estimated from each frame,

$$\hat{C}_p = \sum_{i=1}^T w_p^i \left(\frac{f^{-1}(z_p^i)}{\tau_i} \right), \quad \sum_{i=1}^T w_p^i = 1, \quad (1)$$

where f^{-1} is the inverse of the camera response function, z_p^i is the pixel value at position p for the frame acquired with exposure τ_i and w_p^i is the weight assigned to the i -th exposure for pixel p .

The most recent and efficient approaches [10, 1] propose to exploit raw camera values (in that case, f is linear before attaining the saturation threshold) and draw on a precise knowledge of the camera sensor noise to solve the irradiance estimation problem. Under this noise model, non saturated samples are seen as realizations of random variables Z_p^i distributed according to

$$Z_p^i \sim N(ga_p\tau_iC_p + \mu_R, g^2a_p\tau_iC_p + \sigma_R^2), \quad (2)$$

where g is the camera gain, a_p models the photo response non uniformity (PRNU) factor, μ_R and σ_R^2 are the readout noise mean and variance (see for instance [2] for a complete description of this simplified camera acquisition model).

Given that a closed formula for C_p under model (2) cannot be found for the MLE, different numerical solutions have been proposed in the literature for its numerical estimation. The most efficient one is due to Granados *et al.* [10], which propose an iterative algorithm: at each iteration, the irradiance $\hat{C}_p$ is computed as a weighted average of the irradiance estimations from each frame x_p^i with weights equal to the inverse of the variance of each estimation σ_{pi}^2 ,

$$\hat{C}_p = \frac{\sum_{i=1}^T \frac{x_p^i}{\sigma_{pi}^2}}{\sum_{i=1}^T \frac{1}{\sigma_{pi}^2}}, \quad (3)$$

$$x_p^i = \frac{z_p^i - \mu_R}{ga_p\tau_i}, \quad \sigma_{pi}^2 = \frac{g^2a_p\tau_i x_p^i + \sigma_R^2}{(ga_p\tau_i)^2}. \quad (4)$$

The weights are initialized directly from the input samples. Granados *et al.* also consider the dark currents which we neglect here for being substantially smaller than the other considered noise sources [2].

The MLE approach by Granados *et al.* can be considered to be the state-of-the-art in pixel-wise HDR generation for static scenes. Moreover, Aguerreberre *et al.* show in [1] that this MLE approximation performs very close to the Cramér-Rao lower bound of the irradiance estimation problem. This result is proved for a pixel-wise estimation discarding the saturated samples and raw pixel values following model (2). Also it is shown that in most cases, the iterative approach of [10] does not require more than one iteration to yield good results, the estimated irradiance remaining almost unchanged after the first iteration. In the same paper, the irradiance estimation is shown to be improved when performing MLE using a modified likelihood function that takes into account the information provided by the saturated samples [1].

3. HDR imaging for dynamic scenes and handheld camera

The methods presented in the previous section assume that the images have been perfectly co-registered. These

methods work under the hypothesis that the camera is fixed with respect to a static scene. As a consequence, their performance is highly affected by image misalignments and moving objects. Indeed, the HDR estimation problem becomes far more difficult in the case of dynamic scenes. A precise prior registration and some kind of motion estimation are seemingly essential to perform the frame combination if one wants to avoid blur and ghosting artifacts. Now, motion estimation is often subject to ambiguities and fine global registration becomes complicated for large camera motion.

In the following, we propose to solve the HDR estimation problem in the dynamic case by exploiting the inherent redundancy of images. Drawing on the recent works on multi-image denoising [6, 4], the idea is to combine information from pixels potentially sharing the same irradiance value. In the process, we show that the HDR estimation of dynamic scenes does not require to apply any special or dedicated motion estimation method.

3.1. The proposed non local approach

Given the T frames, we first choose a reference image (c.f. Section 3.1.4 for the reference frame selection) and estimate the irradiance on each pixel by combining the information of similar pixels present in all frames, i.e. the reference frame and the frames corresponding to the other exposure times. Similar pixels are found by comparing the neighborhood of the pixel in the reference image with those of pixels in all frames. Finding similar pixels through patch comparison makes the method robust to image misalignments, resulting from hand-held camera motion, and to moving objects without the need of image co-registration or motion detection. We propose to work with raw data in order to exploit the potential of the statistical noise model in (2).

3.1.1 Finding similar pixels

The first step of the method is to determine which pixels share the same irradiance value. For this purpose we propose to use a variation of the patch comparison used by the NL-MEANS method [5]. This variation is adapted to the noise model (2). The original NL-MEANS algorithm aims to denoise images corrupted with Gaussian noise of constant variance. Under model (2), the noise is Gaussian distributed but the variance depends on both the irradiance and the exposure time. Therefore, the noise variance is different for each pixel (different C_p) in each frame (different τ).

Patches must be normalized by the exposure time in order to be comparable. Thus the comparison is done in the irradiance domain. The proposed distance between a patch centered at pixel p in the reference frame and a patch cen-

tered at pixel q in another frame is given by

$$d(p, q) = \frac{1}{N} \sum_{j=1}^N \frac{(x_{pj} - x_{qj})^2 - 2\sigma_{pj}^2}{2\sigma_{pj}^2}, \quad (5)$$

where x_{pj} (resp. x_{qj}) is the irradiance of the j -th pixel of the patch centered at pixel p (resp. at q) in the reference frame (resp. in the other frame) and σ_{pj}^2 is the irradiance variance, given by

$$x_{pj} = \frac{z_{pj} - \mu_R}{ga_p\tau}, \quad \sigma_{pj}^2 = \frac{g(z_{pj} - \mu_R) + \sigma_R^2}{(ga_p\tau)^2}, \quad (6)$$

with z_{pj} the value of the j -th pixel of the patch centered at pixel p , τ its exposure time and N the number of non saturated pixels in the patch. Saturated samples are discarded and only those patches with at least a minimum percentage of non-saturated samples are considered. The camera parameters (g , a_p , μ_R and σ_R^2) needed to compute $d(p, q)$ are assumed to be known from a previous camera calibration step [10].

It can be shown that if the two patches centered at p and q come from the same (un-noisy) underlying patch, the distance $d(p, q)$ follows a Chi-square distribution. This result is used to threshold the distance judiciously and compute binary weights for the patches. If the distance between patches p and q is small enough, pixels p and q are assumed to follow the same distribution and thus q is used for the irradiance estimation of p (q is assigned weight 1). Otherwise q is discarded (q is assigned weight 0).

3.1.2 Irradiance estimation

The irradiance $\hat{C}_p$ at pixel p is computed as the first step of the MLE iterative approximation given by (3). The combined samples are the K pixels found to be similar to p according to the patch distance (5)

$$\hat{C}_p = \frac{\sum_{k=1}^K \frac{x_p^k}{\sigma_{pk}^2}}{\sum_{k=1}^K \frac{1}{\sigma_{pk}^2}}, \quad (7)$$

$$x_p^k = \frac{z_p^k - \mu_R}{ga_p\tau_k}, \quad \sigma_{pk}^2 = \frac{g(z_p^k - \mu_R) + \sigma_R^2}{(ga_p\tau_k)^2}, \quad (8)$$

where z_p^k is the k -th pixel found to be similar to pixel p and τ_k its exposure time.

3.1.3 Implementation details

The proposed algorithm works on raw data. Hence, the patch distance as well as the irradiance estimator must be computed combining pixels of the same color channel.

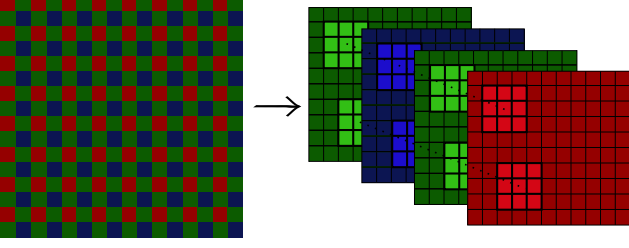


Figure 1: The Bayer pattern is decomposed into four sub-images in order to speed up the similar patch search.

For each frame, the Bayer pattern is decomposed into four sub-images (red, blue and the two green channels separately considered). Patches are then considered as $n \times n \times 4$ -dimensional, where n is the patch size and the third dimension represents the four color channels (see Figure 1). The distance between two $n \times n \times 4$ -dimensional patches $d_T(p, q)$ is the mean distance among corresponding channels

$$d_T(p, q) = \frac{1}{4} \sum_{ch=1}^4 d(p_{ch}, q_{ch}), \quad (9)$$

where $d(p_{ch}, q_{ch})$ is computed using (5). Note that this is equivalent to computing (5) between two patches of size $2n \times 2n$ in the original Bayer pattern. Then the two pixels are assumed to follow the same distribution if $d_T(p, q)$ is below a given threshold as specified in Section 3.1.1. In this way, a list of patches similar to p is created.

The central pixel of a patch corresponds to four different pixels in the Bayer pattern (one for each color channel). The irradiance estimator for each of these pixels is independently computed. For instance, the estimator of the red channel is computed combining the center pixels of the red components of selected patches. Note that the similar patches are found once and used to compute the four irradiance estimators. Another option is to do all processing on the Bayer image directly: center patches at each pixel, compute distances according to (5) and combine samples of the corresponding channels. The advantage of the proposed implementation is that the list of similar patches is computed once and used to estimate four irradiance values, thus making it four times faster.

Following [6] similar patches are sought on a limited search window. Since images are acquired with a hand held camera, a rough global translation estimation is first performed in order to compensate large global translations and allow a reasonable sized search window. The global euclidean distance between translated versions of each frame and the reference frame is computed in order to estimate the translation motion. Notice that this translation estimation is quite rough and only intends to correctly center the patch search window. In all our tests the search window is of size

11×11 pixels (which is equivalent to a window of 22×22 in the Bayer image).

3.1.4 Selection of the reference image

The selection of the reference image has a major impact on results. Indeed, the irradiance values of over or underexposed regions of the reference image cannot be recovered using the proposed method, since the patches lying on those regions do not contain reliable information. Thus these regions need to be identified and filled judiciously. We do not address this aspect of the problem in this paper. We assume that it has been solved in a first stage by a well chosen inpainting technique, as for instance the ones described in [29, 26], which aim at filling these regions in the reference frame with information taken from other frames. By doing so, we assume that an overexposed region of the reference frame will appear as not saturated in at least one of the shorter exposures. Conversely, we assume that underexposed regions can be retrieved from longer exposures. In this process, the reference image has to be chosen carefully, in order to limit the size of the regions to be inpainted. In practice, we choose this reference image as the shortest exposure containing no underexposed region. The HDR generation procedure described in Section 3 can then be directly applied using the filled reference image.

In this work, we make use of a classical inpainting technique, known as Poisson image editing [25], in order to fill the overexposed regions in the reference image. The information is retrieved from other frames by comparing large patches surrounding the region to be filled with similar patches in the other frames. Another interesting possibility to complete the missing regions would be to rely on the work recently published by Sen *et al.* [26], which proposes to generate an HDR image by choosing a reference frame and by filling the missing information through the minimization of a global energy. The unknown information is retrieved from the other frames keeping the coherence with the known regions of the reference image.

4. Results

4.1. Static scene / Static camera

Releasing the hypothesis of aligned images is one of the key points that makes our method robust to camera motion and object motion. However, the price to pay may be quite high since that strong hypothesis carries a large amount of information implying that several samples per pixel can be reliably combined. In practice, that lost information is to be recovered by combining information from neighboring pixels. In this section we present the results of an experiment conducted to evaluate the impact on performance of releasing this assumption on a perfectly static case. Four



Figure 2: Tone mapped version of the HDR image used as ground-truth for the synthetic test presented in Section 4.1. From [13].

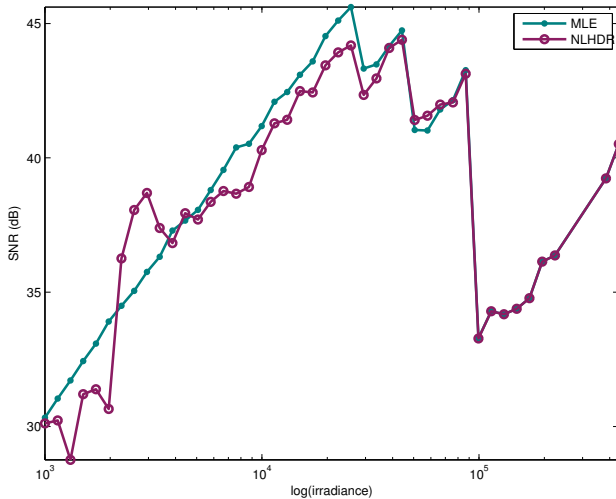


Figure 3: **Synthetic data.** SNR curves obtained using the classical MLE method [10] (green) and the proposed approach (violet). SNR values obtained with the non local approach are very close to those obtained with the classical method. Even though our approach releases the alignment hypothesis, it is close to the optimal result in a case where images are perfectly aligned.

synthetic images corresponding to four different exposures are generated according to model (2) taking as irradiance ground-truth an HDR image. A tone mapped version of the ground-truth is shown in Figure 2. The camera parameters for the simulation are those of a Canon 400D set to ISO 200. This corresponds to the ideal case of the classical MLE approach: static camera and static scene. Figure 3 shows the SNR curves obtained using the classical MLE method [10]

(green) and the proposed approach (violet). It can be verified that the SNR values obtained with the non local approach are very close to those obtained with the classical approach. Hence, releasing the alignment hypothesis does not have significant impact on the algorithm performance in a case where this hypothesis is perfectly satisfied.

This result was also verified using real data. Pictures of a static scene were acquired using a static camera remotely controlled from a laptop. The data set is composed of four images, corresponding to four exposure times acquired with a Canon 400D camera set to ISO 800. The first row of Figure 4 shows JPEG versions of the input image set¹. Since our method treats raw data, a post-processing stage including demosaicking, white balance and tone mapping must be performed to display the results. To perform demosaicking we choose the technique by Hamilton and Adams [12] because of its simplicity, it does not modify the known samples, and offers better performance than a simple bilinear interpolation. More evolved techniques were considered in order to avoid some artifacts present in Hamilton and Adams results. However, they were discarded since they also perform denoising of the samples which would interfere with our denoising results. For tone mapping we use the technique by Mantiuk *et al.* [21].

Figure 4 shows the tone mapped version of the results obtained for the irradiance estimation of the scene. The second row shows the result obtained using the reference image normalized by the exposure time. The saturated regions of the reference image were filled using the Poisson editing patch based approach described in Section 3.1.4. We present this result for comparison purposes, since the normalized reference frame is the simplest HDR generation method we could use. The third row shows the results obtained using the MLE approach [10] and the fourth row shows the results obtained using the proposed non local estimation approach.

As expected, the noisier result is the one obtained with the normalized reference frame. On the other hand, the result for the non local approach is less noisy than the one obtained using MLE. The non local method gives a good quality result even if it does not make use of the alignment hypothesis. Sub-pixel motion (e.g. due to the camera shutter) deviates samples from perfect alignment thus degrading the MLE performance with respect to the ideal case.

For this real example involving a static scene and a static camera, a ground-truth image could have been computed from several pictures so as to compare the SNR results of the different methods, in a way similar to [10, 14]. We declined to do so since, as shown in [2], a highly accurate knowledge of the camera model parameters is needed in order to obtain unbiased results, putting a strong question mark on the

¹All datasets and results are available at http://perso.telecom-paristech.fr/~gousseau/hdr_denoising

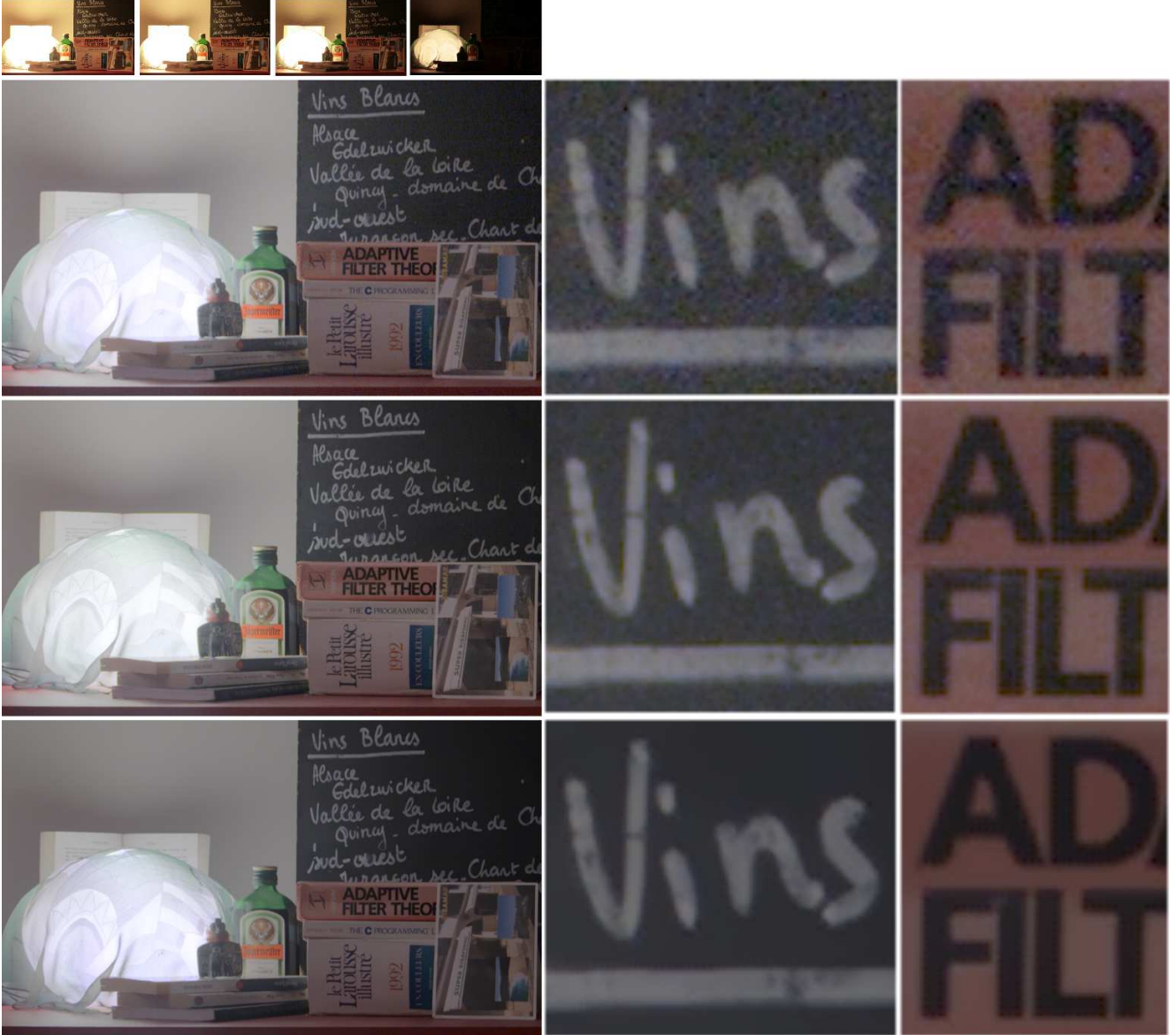


Figure 4: **Real data.** A static scene is acquired using a static camera. **First row:** Input images (JPEG version). **Second row:** Tone mapped reference image normalized by the exposure time. The saturated regions of the reference image were filled using the Poisson editing patch based approach described in Section 3.1.4. **Third row:** Tone mapped irradiance estimation by the MLE approach [10]. **Fourth row:** Tone mapped irradiance estimation of the proposed non local estimation approach. The result of the non local approach is less noisy than the one obtained using MLE. Thus releasing the alignment hypothesis does not have a significant impact on the algorithm performance. Recall that some demosaicking artifacts may appear due to the basic used technique. Please see electronic copy for better color and details reproduction.

validity of such ground-truth.

4.2. Dynamic scene / Hand-held camera

As previously discussed, three main problems must be handled when generating an HDR image: camera motion (global motion), object motion and noise reduction of the irradiance estimation. Several methods addressing these

problems can be found in literature. Nevertheless, to the best of our knowledge, none of them treats them simultaneously. There exists a wide variety of HDR methods that treat camera and object motion [28]. In particular, the work recently presented by Sen *et al.* [26], is shown to be among the state-of-the-art methods treating these issues. On the other hand, the literature is less vast concerning methods

for noise reduction of the irradiance estimation. The application of most classical denoising techniques is not straightforward since the irradiance image does not respect the usually considered hypothesis of additive white Gaussian noise with constant variance. Indeed, the noise distribution varies depending on the estimation method and the variance is seldom constant. Granados *et al.* [10] propose to denoise the result of their irradiance estimation using a bilateral filter. Their irradiance estimation method allows to compute an estimate of the irradiance noise variance. Then the filtering parameter of the bilateral filter is set according to the estimated noise variance. Although Granados *et al.* apply some bilateral filtering after the MLE estimation, we chose not to include their results in the comparison because two critical hypothesis of their approach are not verified. Neither the camera nor the scene are fixed. Therefore, their results present severe ghosting problems on our test sets. In the following we present a comparative evaluation of methods performing HDR imaging for dynamic scenes acquired with hand-held cameras. Next, we present some results comparing our simultaneous HDR - denoising technique against the denoising as a post-processing step after estimating the irradiance.

Evaluating simultaneous HDR imaging and denoising

Here we present the results obtained by the proposed HDR image generation method on three sets of images. Each set is composed of pictures of a dynamic scene acquired with a hand-held camera. We compare our results to those obtained by Sen *et al.* [26]. We also present the results obtained using the reference image normalized by the exposure time. The saturated regions of the reference image are filled using the Poisson editing patch based approach described in Section 3.1.4. We present this result for comparison purposes, since the normalized reference frame is the simplest HDR generation method we could use.

As already explained in Section 4.1, results are displayed using the demosaicking technique by Hamilton and Adams [12] and the tone mapping technique by Mantiuk *et al.* [21]. The first case is a static scene except for one moving object (the postcard). Images are acquired using a hand-held camera (Canon 400D set to ISO 800). The second row of Figure 5 shows a tone mapped version of the irradiance estimation obtained using our non local estimation approach. Three extracts of the results obtained for the moving object are shown in the first row (right side) of Figure 5. It can be verified that the result presents no ghosting artifacts. The second and third rows (right side) of Figure 5 present extracts of the results obtained with the normalized reference frame only (left), Sen *et al.* approach (center) and the proposed non local estimation method (right). The irradiance estimation of the proposed approach is far less noisy than the one obtained using the reference frame only or the

method by Sen *et al.* This example allows to verify both, the effectiveness in noise reduction and the robustness to camera and object motion of the proposed approach.

Figures 6 and 7 present the results obtained for the two other test sets. These two cases present dynamic scenes where several objects are moving: pedestrians in the bridge and people next to the boat for the first example and pedestrians in the street and moving cars for the second example. Images are acquired using a hand-held camera (Canon 400D set to ISO 800 for the bridge scene and to ISO 400 for the street scene). The result for the proposed method is shown in Figures 6 and 7 (left). Extracts of the results obtained for the moving objects are presented in Figure 8 (left for the bridge scene and right for the street scene). It can be verified in both examples that no ghosting artifacts appear. The denoising capacity of the method is shown on various extracts. As in the previous example, the third and fourth rows of Figures 6 and 7 present extracts of the results obtained with the normalized reference frame only (left), Sen *et al.* approach (center) and the proposed non local estimation method (right).

It is interesting to remark that the non local approach is not likely to spread errors possibly introduced by the saturated region filling method. Recall that overexposed regions in the reference image must be filled before performing the non local estimation. If artifacts appear after hole filling, the created patches are not likely to appear elsewhere and thus won't have a significant effect on the estimation.

On other ways to denoise irradiance maps Another possibility to denoise the irradiance map is to apply existing denoising techniques *after* estimating the irradiance. However, as previously mentioned, the application of most of these techniques is not straightforward since the irradiance image does not respect the usually made assumption of additive white Gaussian noise with constant variance. The noise distribution varies depending on the estimation method and the variance is seldom constant. This complicates the setting of parameters. Values chosen to correctly denoise bright regions will remove details in the dark regions while values correctly denoising dark regions will not denoise bright regions. To illustrate this behavior, the classical NL-MEANS algorithm² is applied to the normalized reference image and to the result obtained by Sen *et al.* The NL-MEANS parameters are chosen so as to obtain a good visual compromise between removing noise and keeping details. Figure 9 shows extracts of the obtained results. The proposed approach manages to correctly denoise the bright regions (see the street panel) while still preserving details on the dark regions (see the tree branches). On the contrary, this is not the case for the two other examples. The denois-

²Implementation from [7].



Figure 5: **Real data.** Static scene except for one moving object (the postcard) acquired using a hand-held camera. **Left side.** **First row:** Input images (JPEG version). **Second row:** Tone mapped irradiance estimation using the proposed non local estimation approach. **Right side.** **First row:** Extracts of the moving object (the postcard). No ghosting artifacts appear. **Second / Third row:** Extracts of the normalized reference image (left), the method by Sen *et al.* [26] (center) and the proposed non local approach (right). The irradiance estimation of the proposed approach is considerably less noisy than the one obtained using the reference frame only or the method by Sen *et al.* Please see the electronic copy for better color and details reproduction.

ing technique manages to remove part of the noise on the bright regions but at the cost of blurring the dark zones.

For the same reason, applying a classical denoising technique to each raw LDR image before the irradiance estimation is not a good option. Besides, information is lost when denoising each LDR image independently (e.g. details are lost, blurred edges) which might be kept in a multi-image denoising approach. Moreover, after denoising, the statistical model known for input samples is no longer valid. Thus the nearly optimal irradiance estimation obtained with the MLE estimator is no longer justified and an alternative optimal estimator should be found (which is not obvious given the new unknown model).

5. Conclusions

In this paper we present a new method for HDR image generation which copes simultaneously with three important problems: irradiance estimation noise, camera motion (hand-held camera) and multiple objects motion (dynamic scenes). Previous methods successfully handle these problems independently, but to the best of our knowledge none of them treats them all. The noise reduction capacity and robustness to camera and object motion of the proposed approach was experimentally verified in various real cases. The results show good denoising performance and no ghosting artifacts.

6. Acknowledgments

We would like to thank the reviewers for their valuable remarks. We also thank Miguel Granados and Nima Kalantari for kindly providing their code, and Mauricio Delbracio, Baptiste Mazin, Alasdair Newson and Guillaume Tardivel for their valuable input. This work has been partly founded by the FUI grant CEDCA.

References

- [1] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé. Best algorithms for HDR image generation. a study of performance bounds. *Preprint HAL* <http://hal.archives-ouvertes.fr/hal-00733853/>, 2012.
- [2] 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.
- [3] L. Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 3, pages 7–12. IEEE, 2000.
- [4] G. Boracchi and A. Foi. Multiframe raw-data denoising based on block-matching and 3-D filtering for low-light imaging and stabilization. In *Proc. Int. Workshop on Local and Non-Local Approx. in Image Processing*, volume 1, 2008.



Figure 6: **Real data.** Dynamic scene (pedestrians in the bridge and people next to the boat) acquired using a hand-held camera. **Left:** Tone mapped irradiance estimation using the proposed non local approach. No ghosting artifacts appear. **Right first row:** Input images (JPEG version). **Right second row:** Extracts of the the normalized reference image. **Right third row:** Extracts of the results by Sen *et al.* [26]. **Right fourth row:** Extracts of the results by the proposed non local approach. The result obtained by the proposed approach is far less noisy than the one by Sen *et al.* Please see the electronic copy for better color and details reproduction.

- [5] A. Buades, B. Coll, and J. M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
- [6] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *Int. J. Comput. Vision*, 76(2):123–139, 2008.
- [7] A. Buades, B. Coll, and J.-M. Morel. Non-Local Means Denoising. *Image Processing On Line*, 2011.
- [8] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pages 369–378, 1997.
- [9] O. Gallo, N. Gelfandz, W. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–7. IEEE, 2009.
- [10] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. Seidel, and H. Lensch. Optimal HDR reconstruction with linear digital cameras. In *CVPR*, pages 215–222, 2010.
- [11] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen*, pages 277–284, 2006.
- [12] J. Hamilton and J. Adams. Adaptive color plan interpolation in single sensor color electronic camera. US Patent 5,629,734, 1997.
- [13] S. Hasinoff, F. Durand, and W. Freeman. Noise-optimal capture for high dynamic range photography. <http://people.csail.mit.edu/hasinoff/hdrnoise/>. Accessed: 13/08/2012.
- [14] S. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, pages 553–560, 2010.
- [15] Y. Heo, K. Lee, S. Lee, Y. Moon, and J. Cha. Ghost-free high dynamic range imaging. *Computer Vision–ACCV 2010*, pages 486–500, 2011.
- [16] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *Computer Graphics and Applications, IEEE*, 28(2):84–93, 2008.
- [17] S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics (TOG)*, 22(3):319–325, 2003.



Figure 7: **Real data.** Dynamic scene (pedestrians in the street and moving cars) acquired using a hand-held camera. **Left:** Tone mapped irradiance estimation using the proposed non local approach. No ghosting artifacts appear. **Right first row:** Input images (JPEG version). **Right second row:** Extracts of the normalized reference image. **Right third row:** Extracts of the results by Sen *et al.* [26]. **Right fourth row:** Extracts of the results by the proposed non local approach. The result obtained by the proposed approach is significantly less noisy than the one by Sen *et al.* Please see the electronic copy for better color and details reproduction.



Figure 8: **Real data.** Extracts of the results obtained by the proposed non local estimation method for moving objects present on the scenes of Figures 6 (left) and Figure 7 (right). Notice that no ghosting artifacts appear in neither example.

- [18] E. Khan, A. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. In *Image Processing, 2006 IEEE International Conference on*, pages 2005–2008. IEEE, 2006.
- [19] K. Kirk and H. J. Andersen. Noise characterization of weighting schemes for combination of multiple exposures.

In *BMVC*, pages 1129–1138, 2006.

- [20] 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.

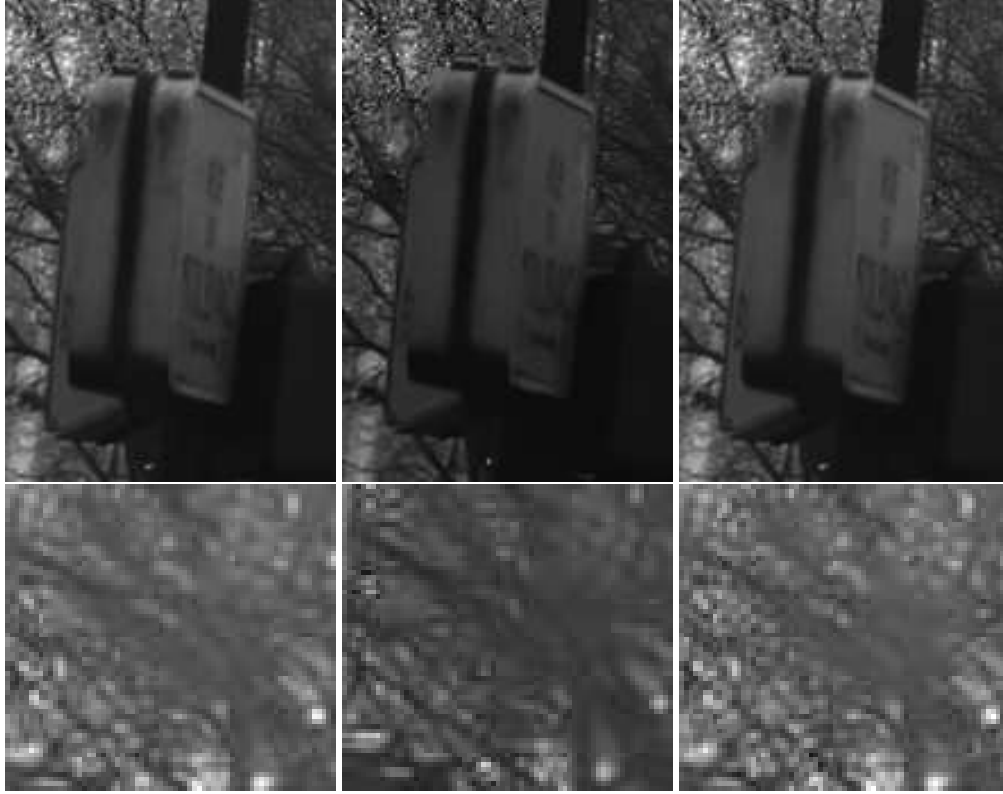


Figure 9: **Real data.** **Left:** Extracts from the NL-MEANS denoising of the normalized reference frame. **Center:** NL-MEANS denoising of the result by Sen *et al.* **Right:** Proposed non local approach. The proposed approach manages to better denoise the bright regions (see the street panel) while better preserving details on dark regions (see the tree branches). On the contrary, this is not the case for the two other examples. The denoising technique manages to remove part of the noise on the bright regions but at the cost of blurring the dark zones. Only the green channel irradiance is displayed in order to avoid contrast changes introduced by the tone mapping techniques (needed to display HDR color images) and better visualize noise level differences.

- [21] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3):68:1–68:10, Aug. 2008.
- [22] T. Min, R. Park, and S. Chang. Histogram based ghost removal in high dynamic range images. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 530–533. IEEE, 2009.
- [23] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *CVPR*, pages 1374–1380, 1999.
- [24] F. Pece and J. Kautz. Bitmap movement detection: HDR for dynamic scenes. In *Visual Media Production (CVMP), 2010 Conference on*, pages 1–8, 2010.
- [25] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers, SIGGRAPH '03*, pages 313–318, New York, NY, USA, 2003. ACM.
- [26] 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.
- [27] D. Sidibe, W. Puech, and O. Strauss. Ghost detection and removal in high dynamic range images. In *European Signal Processing Conference*, 2009.
- [28] A. Srikantha and D. Sidibé. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 2012.
- [29] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):463–476, 2007.
- [30] H. Zimmer, A. Bruhn, and J. Weickert. Freehand hdr imaging of moving scenes with simultaneous resolution enhancement. *Comput. Graph. Forum*, 30(2):405–414, 2011.