



Non Local Exposure Fusion

Cristian Felipe Ocampo-Blandon, Yann Gousseau

► To cite this version:

Cristian Felipe Ocampo-Blandon, Yann Gousseau. Non Local Exposure Fusion. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Nov 2016, Lima, Peru. hal-01334028v2

HAL Id: hal-01334028

<https://hal.science/hal-01334028v2>

Submitted on 15 Jul 2019

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.

Non-local Exposure Fusion

Cristian Ocampo-Blandon and Yann Gousseau

LTCL, CNRS, Telecom ParisTech, Universite Paris-Saclay
75013, Paris, France
{ocampobl,yann.gousseau}@telecom-paristech.fr

Abstract. Exposure fusion is an efficient method to obtain a well exposed and detailed image from a scene with high dynamic range. However, this method fails when there is camera shake and/or object motions. In this work, we tackle this issue by replacing the pixel-based fusion by a fusion between pixels having similar neighborhood (patches) in images with different exposure settings. In order to achieve this, we compare patches in the luminance domain. We show through several experiments that this procedure yield comparable or better results than the state of the art, at a reasonable computing time.

Keywords: exposure fusion, image fusion, patches, non-local methods, high dynamic range.

1 Introduction and Previous Works

Natural scenes often exhibit luminance ranges that are beyond the capacity of most imaging sensors. In such cases, a single standard photograph cannot capture details in dark regions without producing saturated values in the brightest regions. A classical way to create a high dynamic range (HDR) image faithfully representing the luminance of a scene is to combine photographs acquired with different exposure times [8]. Once the response function of the camera is known (or when working with linear RAW images) and when the scene is static, this combination task is essentially a statistical problem [4]. In order to visualize such an image on a classical display, contrast is then compressed by using a so-called *tone mapping* operator [11, 10]. An alternative to this two-step procedure (HDR image creation followed by tone mapping) was proposed in [15] and is called *exposure fusion*. The main idea is to bypass the HDR creation step and to directly create a low dynamic range image (typically made of 8 bits per color channel) by fusing the input images corresponding to different acquisition times. Specific weights are used to ensure that the final result is contrasted enough, has vivid colors and avoid over and under-exposure. Using classical ideas from computer graphics [16], the images are fused in a multi-scale framework, enabling one to blend images seamlessly.

This procedure is very efficient and has yielded numerous softwares and plug-ins. It has also triggered an abundant research literature proposing variants on the original method: involving image decomposition into several components at

the patch level [13], formulating the problem as a MAP estimation [21] or involving the bilateral filter [20]. Nevertheless, such approaches have a common drawback: they fail when there is either camera shake or moving objects in the scene. Indeed, they all perform the fusion at pixel level and assume that images are registered and that objects are still. Of course, this is a serious limitation in practice, and several works have tackled this issue. Anti-ghosting algorithms [5, 17, 22] propose to perform image alignment and possibly to explicitly detect moving objects to prevent using them in the fusion. Recently, the methods from [12, 19] proposed to solve the problem by an iterative optimization procedure relying on correspondences between patches. In particular, the authors of [12] propose to create, from a reference image and a set of differently exposed images, a set of images that are all aligned and radiometrically coherent with the reference, and whose content is automatically modified (in the case of moving objects) to match the reference. This is achieved thanks to contrast prescription and patch-based content comparison between the images. As illustrated in [12], classical exposure fusion can then be applied to the aligned set to yield a satisfactory final image, even in the case of camera shake and object motion.

In this paper, we propose an exposure fusion method that also deals both with camera shake and object motion. The basic idea is very simple and illustrated by Figure 1: instead of fusing values taken by pixels at the same spatial position, we fuse values of pixels having similar neighborhood (thereafter called a patch) in the different images. The method is therefore in the spirit of non-local restoration methods such as the non-local means [7], and also share similarities with the non-local method introduced in [3] for the creation of HDR images. Such a simple approach has the ability to deal with camera motion (although a previous global image registration may be useful to accelerate the process or to enhance the similarity between patches) and with moving objects. By construction, the method also boils down to the original exposure fusion algorithm in the absence of camera and object motions. The rest of the paper is organized as follows. In Section 2, the original exposure fusion algorithm is first briefly recalled and a detailed presentation of the proposed method is provided. In Section 3, we show experimental results and comparisons with the recent state-of-the-art algorithm from [12].

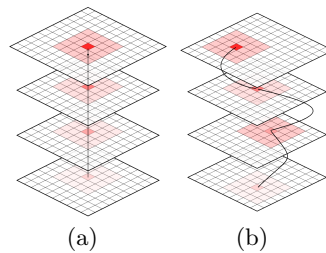


Fig. 1. Instead of fusing values at the same position (a), we use a non-local fusion (b) with values from different positions, to account for object motions and camera shake.

2 Non-local Exposure Fusion

2.1 Classical Exposure Fusion

The method introduced in [15] proposes to fuse a series of images $I_1, \dots, I_N$ acquired with different exposure settings (the knowledge of these is not needed). For each pixel x and index i a weight $W_i(x)$ is defined by taking into account the quality of contrast, color saturation and well-exposedness at this pixel. The idea is then to fuse the values $I_i(x)$ according to the weights. The most straightforward approach would be to define the resulting image R as

$$R(x) = \sum_{i=1}^N W_i(x) I_i(x) , \quad (1)$$

but this yields incoherences in flat regions and visible seams at slow transitions. In order to achieve seamless fusion, the blending is performed in a multi-scale framework. For each level l of a Laplacian pyramid [16], the Laplacian pyramid of the resulting image R is computed as

$$\mathcal{L}_l(R) = \sum_{i=1}^N \mathcal{G}_l(W_i) \mathcal{L}_l(I_i) ,$$

where $\mathcal{L}_l(I_i)$ is the Laplacian pyramid at level l of image i and $\mathcal{G}_l(W_i)$ is the Gaussian pyramid at level l of the weight map W_i . The final image R is then reconstructed from its Laplacian pyramid. As explained in the introduction, this (otherwise very efficient) method fails when there is camera shake or object motions.

2.2 Non-local Fusion

In order to deal with dynamic scenes we propose, for each pixel x , to replace the fusion of values at this pixel with the fusion of values at well chosen pixels $f_1(x), \dots, f_N(x)$ in the series of images. In the static case these pixels would simply be all equal to x , and in the general case, they will be defined in order to compensate motions, as we will now see. In a way similar to the HDR creation method from [3], we define $f_i(x)$ to be the pixel having its neighborhood in I_i the closest to the one of x in I_{ref} (where $ref \in (1, \dots, N)$) a reference image to be defined later. More precisely, we define $P_i(x)$, a patch at pixel x , to be the collection of values in I_i in a $(2s+1) \times (2s+1)$ square window centered at x . We assume that we have a distance d between patches (to be defined in the next section) and then define the pixel most similar to x in I_i to be

$$f_i(x) = \arg \min_{y \in x + \mathcal{N}} (d(P_{ref}(x), P_i(y))) , \quad (2)$$

where $\mathcal{N}$ is a search window (bigger than the patch). In the (unlikely) case where the minimum is achieved for several values, one is chosen at random. If the minimum is achieved for only one value, we have in particular that $f_{ref}(x) = x$. Let us now assume for a moment that we perform the fusion at a single scale (that is, as in Formula (1)). Then the resulting image could be defined as

$$R(x) = \sum_{i=1}^N W_i(f_i(x)) I_i(f_i(x)) .$$

Of course, as in the classical exposure fusion, defining the resulting image R this way would yield incoherences and seams. Therefore, the non-local fusion is defined, for each level l of a multi-scale pyramid, as:

$$\mathcal{L}_l(R) = \sum_{i=1}^N \mathcal{G}_l(\tilde{W}_i) \mathcal{L}_l(\tilde{I}_i) , \quad (3)$$

where $\tilde{I}_i$ is defined as $\tilde{I}_i(x) = I_i(f_i(x))$. The weights $\tilde{W}_i$ are obtained by multiplying the original weights w_i^1 from [15] (accounting for well-exposedness, color saturation and contrast) at position $f_i(x)$ by a weight asserting the similarity between x and $f_i(x)$. Similarly to what is done in the NL-means algorithm [7] we define this similarity weight to be

$$w_i^2(x) = \exp \left(- \frac{d(P_{ref}(x), P_i(f_i(x)))^2}{\sigma^2} \right) , \quad (4)$$

where σ is a parameter. The resulting weights are $\tilde{W}_i(x) = w_i^1(f_i(x)) \times w_i^2(x)$.

In short, for a given pixel, the fusion is performed by using the most similar pixels in each image of the sequence and by modulating the original weights of the fusion by a similarity score.

2.3 Computing Pixel Similarity

Here, we detail how to compute the distance between patches. For this, images acquired with different exposure settings (e.g. different exposure times) should be comparable. In this paper, we assume that we have access to the linear RAW images of the captured scene. This is not a real limitation in the common case of embedded processing. When these RAW images are not available, other scenarios are possible (such as matching all image histograms to the one of the reference image, in a way similar to [12]), but these will not be investigated in this paper.

In addition to the processed images $I_1, \dots, I_N$ (typically JPEG images after demosaicking, white balance and gamma transform), we consider the corresponding RAW (linear) images $R_1, \dots, R_N$. We assume that these have been acquired with exposure times $t_1, \dots, t_N$. The luminance images can then be computed as $L_i = g t_i^{-1} (R_i - O)$, where O is the offset of the camera and g the gain (see e.g. [2]). Without loss of generality, we will assume that $g = 1$ and that the offset is known. The comparison of pixels is then carried out in the luminance domain, so that the distance d between two patches $P_i(x)$ and $P_j(y)$ is defined as the Sum of Squared Differences (SSD) between the values of the patches in the respective luminance images L_i and L_j , that is,

$$d(P_i(x), P_j(y)) = \sum_{k=-s}^s (L_i(x+k) - L_j(y+k))^2 . \quad (5)$$

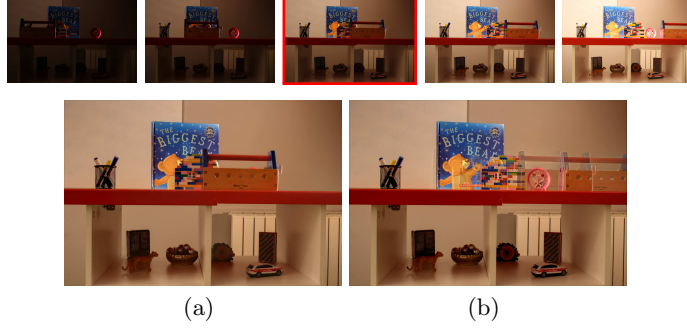


Fig. 2. Image fusion: (a) non-local exposure fusion (NLEF), (b) exposure fusion [15].

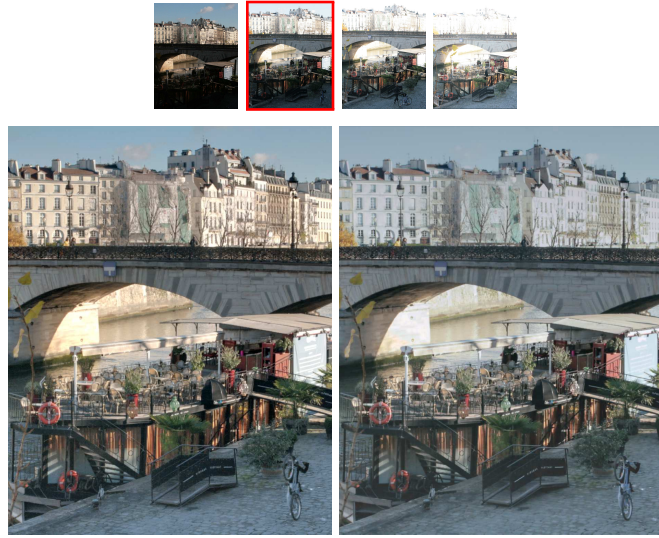
If the camera offset is unknown, or in order to achieve results that are more robust to spatial variability of the gain (photo response non-uniformity, PRNU) or imprecision in the exposure time, the L^2 distance can be replaced by a similarity measure that is invariant to affine transforms of the luminance. Although this aspect will not be developed here, we have observed that the affine invariant distance proposed in [9] is very robust to errors in the luminance conversion.

2.4 Algorithm Details

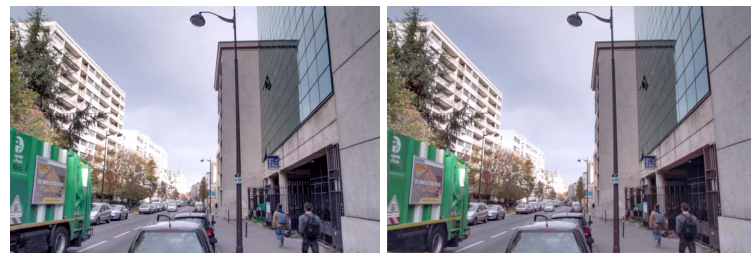
First, the images are aligned through a global SIFT-based homography estimation and bicubic interpolation. Although not necessary, we observed that this step yields more reliable patch comparisons. A reference image is picked as in [12] by choosing the image having the least number of saturated pixels. Then, saturated regions in the reference are removed (see below). Second, and for each pixel in the reference, its best match in image i is chosen using Formula (2) and the similarity measure defined by (5). In order to accelerate the algorithm, the best match is found using the PatchMatch algorithm [6], yielding, for each patch of the reference image, an approximate nearest neighbor (ANN) in each image L_i . The PatchMatch algorithm is run with its default parameters, which in particular implies that the search window $\mathcal{N}$ is the entire image.

The corresponding value of the final image is obtained by fusing the values of the best matches using Formula (3). In this formula, the weights $\tilde{W}_i$ are obtained by multiplying the original weight of [15] by the similarity weight defined by Formula (4). The original weights from [15] depend on three exponents that we all choose to be equal to 1. The parameter σ of the similarity measure (4) is chosen as the quantile of level 5% of the distribution of the distances between patches (separately for each couple L_{ref}, L_i). We empirically found this choice to provide a reasonable compromise in order to simultaneously obtain a result similar to the original algorithm from [15] and efficiently discard moving objects.

A last point that we did not explain is what to do with saturated pixels. We could have chosen a strategy similar to the one of [12]: use only the non-saturated pixels when comparing patches. However, this fails as soon as the saturated region is wider than the patch, which is very common in practice. In order to get more robust results, we use the same strategy as in [3]. For each connected



(a) Our Method - NLEF, (69 sec). (b) Fusion with [12], (143 sec).



(c) Our Method - NLEF, (84 sec). (d) Fusion with [12], (116 sec).



(e) Our Method - NLEF, (184 sec). (f) Fusion with [12], (394 sec).

Fig. 3. Sets of bracketed exposure images and their fusion. (Third image set is courtesy of EMPA Media Technology.)

region Ω of saturated pixels, we detect the region from the other images that is the most similar to Ω on its immediate neighborhood (a ring outside Ω). This region is then blended into the reference using Poisson editing [18].

3 Experiments

In all experiments (see Figure 3), we use a search window equal to the full image. The patch size is set to $s = 1$ (that is 3×3 patches).

First, as a sanity check, we compare the result of our method, NLEF (non-local exposure fusion), with that of the original exposure fusion [15] on a static scene with a moving object, from the database [1]. As we can see in Figure 2, the results are very similar on static parts (everything except one object on the table), but no ghost is produced with our method for the moving object.

Then, we provide three comparisons between our method and the recent method from [12], using the code kindly provided by the authors. Since exposure fusion does not aim at producing a real HDR image (a faithful representation of the luminance values), we only provide visual comparisons. A full user study would be the best way to evaluate the quality of the results [14] but is beyond the scope of this paper. Both methods rely on PatchMatch to get patch correspondences, and we use the same parameters for this algorithm in all experiments and for the two approaches (full search over the complete image and patch size of $s = 1$). The examples corresponding to Figures 3(a) and 3(c), are acquired with a Canon 400D and Canon 7D camera, respectively, for which the offsets are 256 and 2046. Images of these experiments, as well as other comparisons, are included on the supplementary material.

In Figures 3(a) and 3(e), one can see that our method produces more accurate colors. However, the color errors from [12] could probably be solved by taking into account the luminance computed from linear images. Second, one observes that both methods produce accurate details and no ghosts, with occasional artifacts at different positions. However, our approach is both faster and much simpler, and in particular involves no complex optimization procedure.

4 Conclusions

In this work, we have introduced a non-local exposure fusion yielding good results in the presence of camera shake and objects' motions. There are several ways this work could be continued. First, the saturated parts are handled using a Poisson editing procedure, but this step could probably be integrated in the PatchMatch algorithm, at the propagation step. Second, the method may produce noisy results, in particular when the reference image is noisy. We plan to tackle this issue by using more patches for the fusion, in a way similar to the NL-means algorithm. Last, similar non-local approaches could be studied in different fusion schemes, for instance to perform focus stacking.

Acknowledgments. The authors thank C. Aguerrebere for her help and advice. This work is supported by Colciencias under the grant "Programa doctoral de Becas" call 529 and the French Research Agency (ANR) under project MIRIAM.

References

1. HDR imaging website, international university of sarajevo, <http://projects.ius.edu.ba/computergraphics/hdr/>, 2016.
2. AGUERREBERE, C., DELON, J., GOUSSEAU, Y., AND MUSÉ, P. Study of the camera acquisition process and statistical modeling of the sensor raw data. *Prep. HAL*.
3. AGUERREBERE, C., DELON, J., GOUSSEAU, Y., AND MUSE, P. Simultaneous HDR image reconstruction and denoising for dynamic scenes. In *ICCP, 2013 IEEE International Conference on* (2013), IEEE, pp. 1–11.
4. AGUERREBERE, C., DELON, J., GOUSSEAU, Y., AND MUSÉ, P. Best algorithms for HDR image generation. a study of performance bounds. *SIAM J. Imaging Sci.* 7, 1 (2014), 1–34.
5. AN, J., LEE, S. H., KUK, J. G., AND CHO, N. I. A multi-exposure image fusion algorithm without ghost effect. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (2011), IEEE, pp. 1565–1568.
6. BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics-TOG* 28, 3 (2009), 24.
7. BUADES, A., COLL, B., AND MOREL, J.-M. A non-local algorithm for image denoising. In *IEEE Conf. Comp. Soc. CPVR*. (2005), vol. 2, IEEE, pp. 60–65.
8. DEBEVEC, P. E., AND MALIK, J. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes* (2008), ACM, p. 31.
9. DELON, J., AND DESOLNEUX, A. Stabilization of flicker like effects in image sequences through local contrast correction. *SIAM J. Imaging Sci.* (2010), 703–734.
10. DURAND, F., AND DORSEY, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. on Graphics (TOG)* 21, 3 (2002), 257–266.
11. FATTAL, R., LISCHINSKI, D., AND WERMAN, M. Gradient domain high dynamic range compression. In *ACM TOG* (2002), vol. 21, ACM, pp. 249–256.
12. HU, J., GALLO, O., PULLI, K., AND SUN, X. HDR deghosting: How to deal with saturation? In *2013 IEEE Conference on CVPR* (2013), IEEE, pp. 1163–1170.
13. MA, K., AND WANG, Z. Multi-exposure image fusion: A patch-wise approach. In *IEEE International Conference on Image Processing (ICIP)* (2015.).
14. MA, K., ZENG, K., AND WANG, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* 24, 11 (2015), 3345–3356.
15. MERTENS, T., KAUTZ, J., AND VAN REETH, F. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications, 2007*. (2007), pp. 382–390.
16. OGDEN, J. M., ADELSON, E. H., BERGEN, J. R., AND BURT, P. J. Pyramid-based computer graphics. *RCA Engineer* 30, 5 (1985), 4–15.
17. PECE, F., AND KAUTZ, J. Bitmap movement detection: HDR for dynamic scenes. In *Visual Media Production (CVMP), 2010 Conference on* (2010), IEEE, pp. 1–8.
18. PÉREZ, P., GANGNET, M., AND BLAKE, A. Poisson image editing. In *ACM Transactions on Graphics (TOG)* (2003), vol. 22, ACM, pp. 313–318.
19. QIN, X., SHEN, J., MAO, X., LI, X., AND JIA, Y. Robust match fusion using optimization. *IEEE Transactions on Cybernetics* (2015).
20. RAMAN, S., AND CHAUDHURI, S. Bilateral filter based compositing for variable exposure photography. In *Proceedings of Eurographics* (2009).
21. SONG, M., TAO, D., CHEN, C., BU, J., LUO, J., AND ZHANG, C. Probabilistic exposure fusion. *IEEE Trans. Image Process.* 21, 1 (2012), 341–357.
22. ZHENG, J., LI, Z., ZHU, Z., WU, S., AND RAHARDJA, S. Hybrid patching for a sequence of differently exposed images with moving objects. *IEEE Trans. Image Process.* 22, 12 (2013), 5190–5201.