Artifact-free High Dynamic Range Imaging

Orazio Gallo*, Natasha Gelfand‡, Wei-Chao Chen‡, Marius Tico ‡, and Kari Pulli.‡

*orazio@soe.ucsc.edu, University of California, Santa Cruz.

‡{natasha.gelfand, wei-chao.chen, marius.tico, kari.pulli}@nokia.com,

Nokia Research Center, Palo Alto.

Single best exposure tone-mapped | Standard HDR | Our result

Figure 1. An exposure stack taken at a sculpture garden. In the first row from left to right: single best exposure tone-mapped, standard HDR, and our result. Our result captures all the range of the scene while being free of ghosting. The second row shows the original exposures in the stack.

Abstract

The contrast in real world scenes is often beyond what consumer cameras can capture. For these situations, High Dynamic Range (HDR) images can be generated by taking multiple exposures of the same scene. When fusing information from different images, however, the slightest change in the scene can generate artifacts which dramatically limit the potential of this solution. We present a technique capable of dealing with a large amount of movement in the scene: we find, in all the available exposures, patches consistent with a reference image previously selected from the stack. We generate the HDR image by averaging the radiance estimates of all such regions and we compensate for camera calibration errors by removing potential seams. We show that our method works even in cases when many moving objects cover large regions of the scene.

1. Introduction

Progress in Computational Photography has provided amateur photographers with tools to edit pictures in ways that, only a decade ago, were exclusive territory of professional photographers. For example, capturing and representing a large portion of the radiance of a real world scene—potentially several orders of magnitude larger than what standard sensors can handle—can be achieved today with consumer cameras and popular software, such as Adobe Photoshop, by combining differently exposed pictures.

Several methods have been proposed that can create a High Dynamic Range (HDR) image from a set of Low Dynamic Range (LDR) ones (*e.g.* [4]). However, for most of these approaches, it is essential that the scene be completely static in order to avoid introducing various artifacts (for a description of such artifacts refer to Section 2). This drastically limits the applicability of HDR imaging as most

scenes of interest do change, be it because people move or for situations as common as branches blown by the wind.

Additionally, misalignment between the different exposures and noise in the data, as well as in the estimated camera response function, can complicate the problem of detecting and correcting these artifacts.

We propose a method to generate an artifact-free HDR image. Given a stack of images, we select as a reference the image with the most useful information about the scene. In each of the other exposures in the stack, we detect regions that do not cause artifacts when combined with the reference image. We then create a ghost-free HDR image averaging only the information from these regions which are, by construction, consistent. Finally, we remove potential boundary discontinuities around areas generated with different sets of exposures.

Our algorithm produces an image that captures as much dynamic range of the scene as possible and guarantees that no artifacts are generated. We show that the proposed method performs extremely well even when the scene is affected by substantial changes, as shown in Figure 1. The proposed approach naturally extends to other applications that require combining multiple pictures, such as noise reduction.

2. Related Work

Hardware solutions for capturing HDR images are at an adolescent stage at best (Reinhard *et al*. provide a comprehensive review [14]). Because they capture the whole range in a single shot, these devices are not affected by scene changes, unless the exposure time is too long. Unfortunately, HDR camera prototypes usually suffer from limited resolution and are not available to consumer users. Moreover, because of the large availability of point and shoot cameras, approaches that only require capturing a set of standard LDR pictures have a greater potential to impact everyday photography.

2.1. HDR Generation

HDR images are usually generated from an LDR image stack by computing a weighted average of the aligned input images [12, 4, 13]. Whenever an object moves, the averaging process washes it out. This artifact is usually referred to as ghosting because of its faint appearance. The visibility of a ghost generally grows with the difference between the background and the moving object. However, because the weighting of input images is designed to disregard pixels that are over- or under-exposed, ghosts of extremely dark or bright objects are usually difficult to notice.

A way of avoiding ghosting altogether is to find the best exposure for each region in the scene. The final HDR image can then be generated as a collage of all these patches [3]. This approach can generate a different type of artifact: an object can be deformed or duplicated if it moves across the boundary of one such region. Eden *et al*. use a similar strategy in that they select only one exposure per region [5]. Because they first pick the middle exposure as the reference frame to paste regions into, scene consistency is encouraged; however, objects moving in or across regions that are over- or under-saturated in the reference frame can still be duplicated or deformed. The main limitation of this class of methods is that, at each pixel location, the information is extracted from a single image.

2.2. Artifact Removal

Ghosting can be tackled through motion estimation. Bogoni, after global registration, uses optical flow estimation as a means of per-pixel registration so that the radiance values from the images in the stack can be correctly combined [2]. Kang *et al*. take a similar approach and extend it to produce HDR video sequences [9]. These methods are limited by the quality of motion estimation as any mismatch can generate ghosting.

A second class of algorithms seeks to modify the weighted average to account for some meaningful property of the pixel, together with its exposure. Khan *et al*., for instance, suggest to estimate the probability of a pixel belonging to a moving object [10]; instead of employing this information to track and realign the pixels, they incorporate it in the weighting function to strongly attenuate the contribution of moving pixels. The results of this approach look very promising. However, apart from some faint ghosts being still visible in some examples, the generated scene is not guaranteed to be consistent because the weighting is done at a pixel level and, therefore, objects might be duplicated. Finally, it is not clear how it would perform when most of the scene changes, as in the case of tree branches moving in the wind, because of the underlying assumption that neighborhoods around pixels predominantly represent background.

A different approach consists in determining how likely a certain pixel is to generate ghosting. Reinhard *et al*. use the variance across different exposures [14]; we found that, due to slight misalignments and errors in camera calibration, the variance at pixel level is often too noisy. Jacobs *et al*. have recently proposed an entropy-based method that does not even require camera calibration [8]. The common denominator to these techniques is the replacement of entire regions with a single exposure; as the size of the ghosted region increases, the dynamic range of the scene can be compromised. Moreover, Jacobs *et al*., in order to avoid visible seams at the border of pasted regions, still average with the potentially ghosted HDR image.

3. Deghosting Algorithm

We seek to create an HDR image where neither duplication, nor ghosting artifacts are present. Duplication can be avoided by using a single image from the stack as a reference. Additionally, if its dynamic range is extended exclusively with consistent regions from the rest of the stack, ghosting artifacts will not be introduced. We assume that the input images are aligned and that we have an estimate of the camera response function. For the latter we use the approach described by Debevec and Malik [4].

3.1. Reference Image Selection

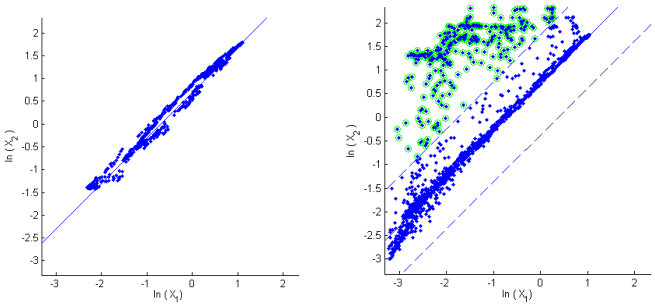
A crucial stage of our algorithm is that of picking a reference image from the stack so we can determine and omit inconsistencies of the rest of the exposures with respect to it. Consistency with the rest of the exposures means absence of ghosting artifacts; additionally, because the reference is a single image rather than a combination, it is guaranteed to be self-consistent, thus no duplication artifacts are possible. On the other hand, even with a moderate amount of motion in the scene, it is virtually impossible to automatically generate a self-consistent reference frame by pasting regions from different images in the stack, for the reasons described in Section 2.1. Because the final result is an HDR version of the reference frame, it is often useful to let the user select it. In this way, undesired objects can be removed, provided that there is at least one exposure where they do not appear, as in the case of Figure 5. If every exposure in the stack is acceptable to the user, the reference frame should be chosen carefully for it strongly impacts the final result. Note that the picture with the overall best exposure, typically the middle one, is not necessarily the optimal choice: an image that is globally over- or under-exposed in which, however, texture is completely preserved, should be preferred to one that is perfectly exposed, apart from one or more completely saturated regions. Regions that are over- or under-saturated, in fact, do not provide any valuable information to avoid ghosting.

To suggest a good reference frame, therefore, we find the saturated pixels in each image of the stack; we then remove small saturated regions with morphological operators (erosion followed by dilation) because, for such areas, the neighborhood usually contains enough information to avoid artifacts. Finally we pick the exposure with the fewest remaining saturated pixels.

3.2. Extending the Dynamic Range

The reciprocity assumption states that, if the radiance of the scene does not change, the exposure time X and the irradiance E are linearly related through the exposure time Δt:

$$X = E \cdot \Delta t. \tag{1}$$



(a) (b)

Figure 2. The values at the pixels in one patch are plotted versus those of the corresponding pixels in an image which is one stop brighter. For matching patches, (a), the data points should lie on the line $y = x + ln(ev_{ij})$. We can measure the *ghosting value* of two patches by computing the percentage of points that are more distant than a given threshold from the line. In (b) such points are indicated by green circles and the threshold by the dashed line.

In other words, given the value at the pixels $\mathbf{p}_i$ in the i-th image of the stack, the exposure at all corresponding pixels in the j-th image with relative exposure ev_{ij} should satisfy

$$X(\mathbf{p}_j) = X(\mathbf{p}_i) \cdot ev_{ij}. \tag{2}$$

Aside from over- and under-saturated pixels, Eq. 2 should only break when the scene changes, and can therefore be used to decide if the irradiance at a given pixel in the reference frame can be combined with that of the corresponding pixel in another image in the stack. In practice, however, a small misalignment or imprecise estimation of the camera response function can produce large deviations from this behavior. To obviate this lack of robustness at the pixel level, we compare patches instead.

Figure 2 shows the log-exposure of each pixel in one patch of one image, plotted versus the log-exposure of the corresponding pixels in an image one stop brighter. Each dot in a plot corresponds to one color channel for a pixel, thus an RGB pixel produces three dots. Figure 2 (a) shows a result when the scene within a patch aligns well. Taking the logarithm of Eq. (2), we get $ln(X(\mathbf{p}_2)) = ln(X(\mathbf{p}_1)) + ln(ev_{12})$, that is each log-exposure value in a patch should be offset by a constant value. The ideal transfer function is marked by the straight $45°$ line, and the corresponding pixels cluster very close to the line. However, when the scene changes within the patch, the exposure values at the same pixel coordinates do not follow this simple relation, as shown in Figure 2 (b). Based on this observation, we define the *ghosting value*, a measure of the deviation of the exposure in a patch from the model predicted from another patch. We first detect the set of outliers, samples that are farther than a threshold from the expected line (see circled samples in Figure 2 (b)). The ghosting value for the pair of patches is the maximum, over the three color channels,

of the number of outliers over the total number of pixels in the patch. In other words, the ghosting value measures the percentage of pixels that can cause ghosting in a patch. We then determine if the patch pair is consistent by using a second threshold, this time on the ghosting value. Both thresholds can be selected by the user to adjust the sensitivity to ghosting: while more conservative thresholds may be required in some situations to remove subtle ghosting, they potentially limit the dynamic range of the final image. In the results presented in this paper, because the different exposures were at least one stop apart, the first threshold was set to 0.75, and the patch was accepted when containing less than 0.5% of outliers.

With this approach, we can find, for a given patch in the reference frame, all the consistent patches in the other images in the stack. Repeating this process for all the patches in the reference image, we can generate a ghost-free HDR image combining all the irradiance information available. For instance, if an object moves only in one frame (other than the reference), all the remaining pictures from the stack are used. This provides a significant improvement over other approaches (*e.g.* [8], [5]) which use only one exposure for regions where inconsistency is detected, even when the potential source of ghosting is a single frame in the stack.

All the results shown in the paper are obtained by dividing the reference image in a 40×40 grid of patches; smaller patches can allow for a larger dynamic range close to areas with motion, but result in an increased sensitivity of the ghosting value to noise. In principle, the patches can be of arbitrarily complex shape—they can, for instance, be determined by segmenting the reference image—but we found that simple, rectangular patches work well.

3.3. Blending

The log-irradiance $L_i(\mathbf{p})$ of each input image i at pixel $\mathbf{p}$ can be estimated, given the camera response function g, as

$$L_i(\mathbf{p}) = g(I_i(\mathbf{p})) - ln(t_i), \tag{3}$$

where I_i denotes the i-th input image and t_i is the correspondent exposure time.

Multiple exposures for each pixel $\mathbf{p}$ can be combined with a weighted average as Debevec and Malik did [4] to get the HDR irradiance map L_H. As described in the previous section, for each patch we find the largest set $\mathcal{I}$ of input images that are consistent with the reference image; two neighboring patches should therefore merge seamlessly, even if they are computed from different subsets of exposures, L being a property of the scene.

In practice, significant artifacts are often visible at the boundaries of blocks averaged from different sets of input images because of inaccuracies in the camera response function estimation (see Fig. 3).

To compute the final radiance image with no visible boundary, we adapt the method proposed by Fattal *et al.* [6]. First, the gradient $\mathbf{G}(\mathbf{p})$ of the log-irradiance image is estimated at each pixel in a block as

$$\mathbf{G}(\mathbf{p}) = \nabla L_H(\mathbf{p}), \tag{4}$$

where $L_H(\mathbf{p})$ is computed from the exposures in $\mathcal{I}$ only, and $\mathbf{G}(\mathbf{p}) = [G_x(\mathbf{p}),\ G_y(\mathbf{p})]$ is the numerically estimated gradient field of the sought log-irradiance image inside the block.

Next, $\mathbf{G}$ is extended over the entire image by pasting the gradient fields of all image blocks. The value of the gradient at the boundary of the patches needs to be hallucinated, for example by replicating its last row and column. In order to avoid additional artifacts, however, before detecting potential ghosts as described in Section 3.2, we extend the patches by one row and one column. In addition to providing the real gradient at each location of the patch, this strategy also benefits the consistency of the gradient between neighboring patches, as it causes nearby blocks, which are already consistent with the reference image, to overlap in the log-irradiance domain.

The final log-irradiance image L_H^* can be estimated by integrating the gradient field $\mathbf{G}(\mathbf{p})$ over the image domain. To do this, we aim to estimate an image whose gradient is closest to $\mathbf{G}$, in the mean squared error sense. Formally, the solution L_H^* must minimize

$$\iint ||\nabla L_H^*(\mathbf{p}) - \mathbf{G}(\mathbf{p})||^2, \tag{5}$$

where the integration is done over the entire image. According to variational analysis theory, the solution to Eq. (5) must satisfy the following Euler-Lagrange equation at each pixel location $\mathbf{p}$:

$$\Delta L_H^*(\mathbf{p}) = div\, \mathbf{G}(\mathbf{p}), \tag{6}$$

where Δ and div stand for Laplace and divergence operators, respectively.

Equation (6) is known as Poisson equation, for which various numerical solutions have been described [7]. In our work we imposed Neumann boundary conditions for solving the differential equation Eq. (6). In other words, the gradient of L_H^* is assumed zero at the boundary of the image along the boundary normal.

Because we solve Eq. (6) for each color channel separately, we also need to correct the color balance of the output. After converting L_H^* to the estimated irradiance E, we pick a pixel $\hat{\mathbf{p}}$ that is not affected by ghosting when averaged over all the exposures and calculate its irradiance $\hat{E}$. We then scale E so that $E(\hat{\mathbf{p}}) = \hat{E}$. This is similar to the approach taken by Agarwala *et al.* [1].

Figure 3. First row from left to right: simple HDR image with ghosting, the output of our algorithm before blending, and our result after blending. Particularly in smooth areas, abrupt changes at the boundary of different patches are visible before blending. The second row shows the original exposures in the stack (note that the scene changes substantially with some objects moving and some disappearing).

Figure 4. The top left image is the simple HDR image with ghosting from a sequence of 4 pictures taken in a forest. A person walking in the scene causes ghosting on the right side of the left image. Moreover, the branches present strong artifacts due to the wind blowing. The top right image is our result. A detail from standard HDR and our result can be seen in the bottom row (right), together with two images from the stack (left).

Standard HDR

Our result

Figure 5. Our algorithm can also be used to remove unwanted objects from the HDR image. The second row shows the original exposures in the stack. According to our measure, the exposure to be used as the reference was the 4^{th}; however, because we aimed to also remove the people walking, the 5^{th} picture was manually selected from the stack. Note that some regions of the 4^{th} frame that did not cause ghosting were still combined in the final HDR (see Figure 7).

4. Results

We tested our algorithm on a variety of scenes to evaluate its performance in the presence of different sources of artifacts. All the stacks were taken by manually bracketing the exposure time, all other settings of the camera being left unvaried. The camera response function was estimated only once with the approach described in [4]. All the HDR pictures shown in the paper have been tone-mapped with the algorithm proposed by Lischinski *et al*. [11].

Figure 1 shows a common case: a high dynamic range scene characterized by the presence of people walking around. The result shown is obtained by automatically selecting the reference frame because the exposures in the stack were semantically equivalent. A different situation is shown in Figure 5. The frame automatically selected as a reference was the fourth from the left on the bottom row, however, given the presence of people walking, we manually picked the fifth, the choice being only based on the scene content. Note that our result captures most of the range, part of which is still extracted from regions of the fourth frame, while preserving complete consistency with the reference frame. For both Figure 1 and Figure 5, we also show how many exposures were used to generate the single patches (Figure 7). Note that it is infrequent that only one exposure is selected; even in regions where some motion occurs, all the consistent images are combined. Figure 4 shows another typical situation that our algorithm can handle; the wind is causing both the branches and their shadows to move.

Figure 6 shows another application of the same principle. A dark scene is captured with a high ISO sensitivity, with obvious repercussions on the level of sensor noise. An effective strategy for removing noise is to take a stack of pictures and average them [15]; however, when movement occurs, ghosts are introduced, as shown in the picture in the middle. Our de-ghosting approach can be applied to average only patches that are consistent with a reference frame; in other words, we can treat the set of pictures as an exposure stack where all the exposure values are the same. The results are identical to those of a simple average where no movement occurs, while in the dynamic regions ghosting is traded for some noise.

5. Limitations and Future Work

Occasionally, some barely visible seams at the boundary of different patches can survive the Poisson Blending step. To remove this artifact, the shape of the patches can be allowed to vary to match that of objects in the reference image. Additionally a multi-scale approach can be designed so as to preserve the robustness offered by patches while allowing for smaller patches where needed.

It is also noteworthy that, when every image in the stack contains a large region that is over- or under-saturated, the risk exists that the correctly exposed picture will be detected as generating ghosting by our algorithm, thus being discarded from the HDR computation. In practice we found that, because we use patches, the small amount of remaining texture is generally enough to prevent this from happening. Further work can be done on defining a smart collage of the different images from the stack to address this problem in a way that does not create duplication artifacts.

6. Conclusions

We presented an algorithm capable of capturing as much dynamic range of a scene as possible, without introducing ghosting. When the scene is static, our approach is equivalent to standard HDR techniques; otherwise it successfully determines which regions can be combined with the reference image from other exposures in the stack so that consistency is preserved. Our algorithm proved successful on a variety of different scenarios, even when the motion affects a substantial part of the scene. In addition to preventing artifacts, it also allows the user to select the reference frame

Figure 6. Stack consisting of 10 pictures affected by ISO noise, one of which is shown on the left. The noise can be attenuated by averaging the pixel values across the stack, but moving objects can cause ghosting as shown in the middle picture. Alternatively, a reference picture can be selected and only pixels consistent with it can be averaged together as shown in the picture on the right. Note that, where movement occurs, noise may be not removed (see inset on the puppet).

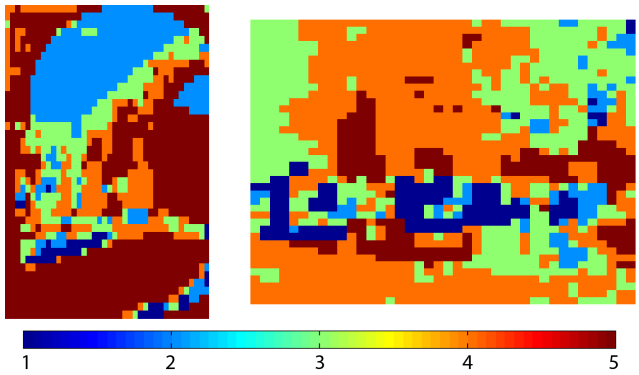


Figure 7. Our algorithm combines as much dynamic range of a scene as possible without introducing ghosting. In these pictures, the color of a patch is proportional to the number of images combined. On the left is the map for Figure 5, on the right that for Figure 1. Note that, even when a region changes significantly in all the exposures, as in the case of the people walking in Figure 1, our algorithm still finds multiple exposures to combine.

so as to remove potentially unwanted objects. The idea can be extended to other applications requiring a stack of pictures, as we showed by de-noising images affected by ISO noise.

References

[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Trans. Graph*, 23:294–302, 2004. 4

[2] L. Bogoni. Extending dynamic range of monochrome and color images through fusion. *International Conference on Pattern Recognition (ICPR)*, 3, 2000. 2

[3] W.-H. Cho and K.-S. Hong. Extending dynamic range of two color images under different exposures. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, 4:853–856 Vol.4, Aug. 2004. 2

[4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 369–378, 1997. 1, 2, 3, 4, 6

[5] A. Eden, M. Uyttendaele, and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:2498–2505, 2006. 2, 4

[6] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 249–256. ACM, 2002. 4

[7] J. D. Hoffman. *Numerical Methods for Engineers and Scientists*. CRC Press, 2001. 4

[8] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28(2):84–93, 2008. 2, 4

[9] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 319–325. ACM, 2003. 2

[10] E. Khan, A. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. *2006 IEEE International Conference on Image Processing*, pages 2005–2008, Oct. 2006. 2

[11] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski. Interactive local adjustment of tonal values. *ACM Trans. Graph.*, 25(3):646–653, 2006. 6

[12] S. Mann and R. Picard. Being undigital with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 323, M.I.T. Media Lab Perceptual Computing Section, 1994. Also appears, IS&Ts 48th annual conference, Cambridge, Massachusetts, May 1995. 2

[13] T. Mitsunaga and S. Nayar. Radiometric Self Calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 374–380, Jun 1999. 2

[14] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann Publishers, December 2005. 2

[15] M. Tico and M. Vehvilainen. Robust image fusion for image stabilization. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 565–568, 2007. 6