

High quality video in high dynamic range scenes from interlaced dual-ISO footage

Raquel Gil Rodríguez and Marcelo Bertalmío; Department of Information and Communication Technologies; Universitat Pompeu Fabra, Barcelona, Spain

Abstract

In this paper we present a simple and affordable method to generate high quality video from a high dynamic range scene. It is performed without utilizing extra lighting, neither alternating exposures, nor operating with dual-camera set-ups. Our input is an interlaced video alternating row pairs with different ISO values, as some DSLR camera models can provide. The proposed algorithm involves two main steps: first the computation of two single-ISO full-frame images (one for each ISO value) using an inpainting-based deinterlacing method, followed by their combination into a single frame by a weighted average. This results in a high dynamic range frame, containing all the details in bright and dark areas at the same time, that is finally tone-mapped into a low dynamic range frame for display purposes. Current results show this is a practical and cost-effective method that produces outputs free of ghosting artifacts and with very little noise.

Introduction

The human visual system is able to adjust to world scenes where the light intensity values cover a very wide range, capturing details in dark and bright areas simultaneously. Dynamic range is defined as the ratio between the brightest and the darkest intensity levels, and while in common situations the light coming from a scene is of high dynamic range (HDR), the vast majority of camera sensors (and displays) are of low dynamic range (LDR). There is a vast literature on methods for creating HDR images using regular, LDR sensors, that started with the seminal approaches of [17, 7]; in those works, several LDR pictures of the same HDR scene are taken with varying exposure time, so that the short exposures capture details in the bright regions, long exposures capture details in the dark regions, and finally all these images are combined into a single, HDR image with overall detail visibility. In order to be able to show this image on a regular, LDR display, it has to be transformed through a process called tone-mapping that compresses the dynamic range of the image while trying to maintain its details and natural appearance.

In the movie industry there is a growing interest in HDR imaging, but the challenge of shooting HDR scenes using LDR equipment exists since the inception of cinema: the way to address it is to add artificial lights, whose effect is to raise the intensity levels of the darkest parts of the image hence reducing the dynamic range of the scene, fitting it into the reduced range of the capture medium (film or digital). This is a cumbersome, expensive procedure requiring very significant human and material resources that greatly affect the cost of the production. Some alternatives exist but they are not fully practical: some digital cinema camera models are able to alternate exposure times on consecutive frames, creating pairs of different-exposure images that

are then fused following the approach of [7], but camera and/or object motion produces ghosting artifacts on the fusion results. A recent possibility is to use a dual-camera set-up [9], with two synchronized, perfectly registered cameras on an orthogonal rig so that a semi-transparent mirror sends most of the light intensity to one of the cameras, and the rest to the other camera. These images can be fused without problem because they are fully aligned, so there is no risk of ghosting, but the dual-camera process has limitations: cost and practicality considerations stemming from the use of two cameras, image problems caused by imperfections in the mirror, the need to perform tone-mapping to the output. For more details we refer the reader to [3].

More recently, we find a very small number of works that perform HDR reconstruction from a single interlaced image. Gu *et al.* [11] combine rows taken with different exposures times, and since the rows are not captured simultaneously this method produces ghosting artifacts as well, which need to be reduced by estimating and compensating for the motion-blur. The camera software Magic Lantern (ML) [16] allows some camera models to capture image/video with dual-ISO values that alternate between consecutive image line pairs. It provides an implementation to interpolate a full-frame low-ISO image, containing less noise on shadow areas. It does not claim to compute an HDR image, though the final picture is the result of combining the information from both the low-ISO and the high-ISO full-frame images. The method follows a chain of steps: separate the two ISO frames, interpolate the missing lines to get the full images, and combine information from both interpolated frames to highly reduce the noise in dark regions. Hajisharif *et al.* [12] perform at the same time demosaicing, denoising, re-sampling and HDR-reconstruction, starting from the interlaced input provided by the ML software [16]. This method requires a previous radiometric calibration process, therefore it cannot be used when the camera is not available. A similar idea was developed by Heide *et al.* [13], who propose a single optimization step using image priors and regularizers of the different stages going on in the camera color pipeline (denoising, demosaicing, etc). The selection of image priors is crucial for the optimization process, and the values are highly dependent on the set of images selected for learning the best weights. These latter two methods have the advantage of working directly with the RAW data without following a staged pipeline, therefore no cumulative errors are carried out from one process to the next; nevertheless, this integration makes it difficult to further extend the processes involved, since they are not independent, plus it's also challenging to locate and rectify errors in the pipeline.

Our main contribution in this work is to propose a simple and effective method to shoot high quality video in HDR scenar-

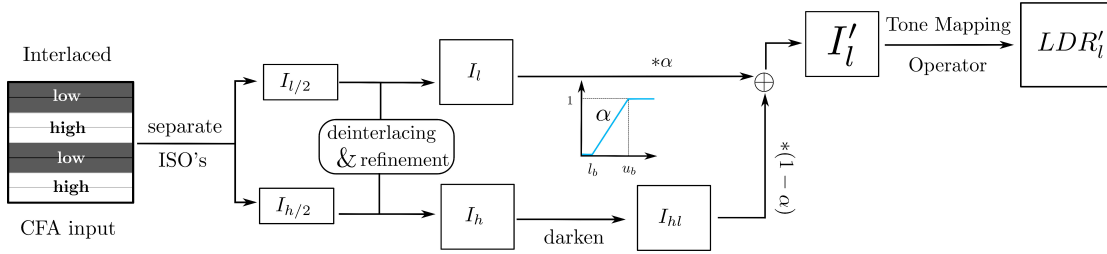


Figure 1. Schematic of the proposed method. First the dual-ISO input is split into two half-size images $I_{l/2}$ and $I_{h/2}$, each one with the rows corresponding to a single ISO value. Next a deinterlacing method is used to generate full-frame images I_l and I_h from $I_{l/2}$ and $I_{h/2}$ respectively. These full-frame images are linearly combined and tone-mapped to produce the final output.

ios using a single camera capable of recording interlaced dual-ISO footage. The interlaced input guarantees that the result will be free of ghosting artifacts, because the low and the high ISO lines are recorded simultaneously. No access to the camera used to record the input is required. Our implementation pipeline incorporates a set of stages: calculating the full-frame single-ISO images applying a deinterlacing algorithm, combining these full-frame images into a single HDR picture, and finally applying a tone-mapping operator to produce a LDR output. For these stages we adapt to our setting state-of-the-art algorithms which produce high-quality results. Tests and comparisons show that our method outperforms other approaches both quantitatively (in terms of PSNR) and qualitatively (no spurious colors, better edge preservation, less noise).

Methodology

The input to our algorithm will be a video sequence in RAW format, where each frame alternates row pairs with different ISO values, and the output will be a LDR video sequence with simultaneous detail visibility in the dark and bright zones of the picture, see scheme in Figure 1. All the stages of our method are applied on a frame-by-frame basis except for the final tone-mapping, which imposes temporal consistency on the output by considering several input frames at the same time. Let's describe our proposed method in detail.

Generation of single-ISO full-frame images

The first stage consists in the computation of two single-ISO full-frame images from an input dual-ISO frame, using an inpainting-based deinterlacing method. In order to obtain both these full-frame pictures I_l (for the low ISO value) and I_h (for the high ISO value) we proceed as follows:

- Split the dual-ISO input into two half-size images $I_{l/2}$ and $I_{h/2}$, each one with the rows corresponding to a single ISO value, see Figure 2.

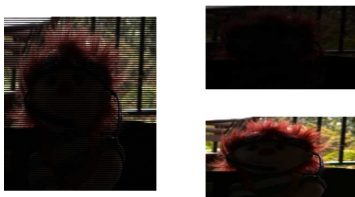


Figure 2. The dual-ISO input frame (left), which is split into two half-size images $I_{l/2}$ and $I_{h/2}$ (right), each one with the rows corresponding to a single ISO value.

- Using an adapted version of the inpainting-based deinterlacing method [2], generate full-frame images I_l and I_h from $I_{l/2}$ and $I_{h/2}$ respectively.
- Perform demosaicing using [20], and apply a refinement step to improve the interpolated results, see Figure 3.



Figure 3. Generated full-frame images I_l (left) and I_h (right).

Row interpolation by deinterlacing

In order to interpolate the missing rows and generate I_l from $I_{l/2}$ and I_h from $I_{h/2}$, we adapt the deinterlacing method of Ballester *et al.* [2]. This is a state-of-the-art technique that follows the dense stereo matching approach of Cox *et al.* [5] and fills-in a missing line $\mathcal{L}_0$ between two given lines $\mathcal{L}_-$ and $\mathcal{L}_+$ by (see Fig. 4):

- First, performing a global matching between lines $\mathcal{L}_-$ and $\mathcal{L}_+$ by computing the correlation matrix between their image values and finding the matches (optimal path) through dynamic programming. In practice only a band around the diagonal is considered for the search for the optimal path, and the matches are estimated assuming a certain noise variance for the image values.
- Second, each matching pair of pixels determines a segment that crosses the missing line $\mathcal{L}_0$ at a pixel location that is filled-in with the average of the matching pair. For those points that were not matched, their values are computed by bilinear interpolation from the neighboring correspondences.

The images $I_{l/2}$ and I_h are color filter array (CFA) RAW pictures with the 2×2 Bayer pattern 'RGGB'; with the R and B values we create half-size channels that can be deinterlaced directly with [2], while for the G channel we first demosaic the values using [20] and then apply the deinterlacing method [2] but with the modification of filling-in two consecutive lines simultaneously given the upper and lower neighboring lines of the pair; see Figure 5.

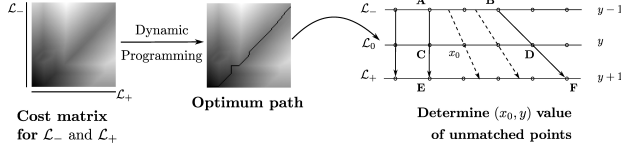


Figure 4. Inpainting-based deinterlacing algorithm. Left: global matching between two given lines L_- and L_+ . Right: each matching pair of pixels determines a segment that crosses the missing line L_0 at a pixel location that is filled-in with the average of the matching pair. Figure adapted from [2].

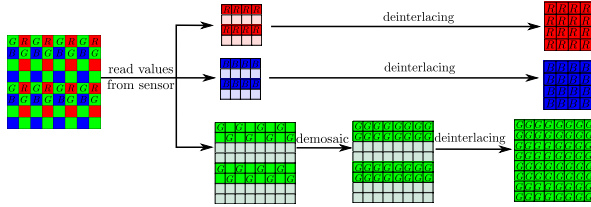


Figure 5. Extended deinterlacing algorithm. It is shown how RGB values are read from the CFA image before applying deinterlacing, and the final interpolated RGB planes.

Refinement step

In addition, a refinement step is required to improve the interpolated values in both the I_l and I_h full-frame images, see Figure 6. We describe the process for I_l , since it will be analogous for I_h . First, we demosaic the frame applying [20]. Second, we perform deinterlacing once again using the same method [2] as before. For simplification we focus on a small row neighborhood defined by four rows r_{s_1} , r_{d_1} , r_{d_2} and r_{s_2} . The subscript s denotes a row with values given by the sensor, and d denotes a row calculated from deinterlacing in the previous step:

1. Consider r_{s_1} , r_{d_1} , r_{d_2} and apply deinterlacing assuming that r_{d_1} is unknown. Store the new computed row in r_{d_1} .
2. Consider r_{d_1} , r_{d_2} , r_{s_2} and apply deinterlacing assuming that r_{d_2} is unknown. Store the new computed row in r_{d_2} .

As a result, we obtain the final full-frame images I_l and I_h . This refinement step produces more accurate and smoother edges, as Figure 6 shows.

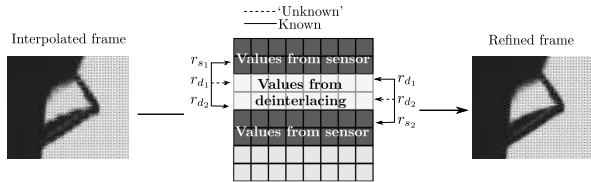


Figure 6. The diagram shows the rows considered for the refinement step: r_{s_1} and r_{s_2} from sensor readout, r_{d_1} and r_{d_2} from the previous deinterlacing step. The images show a crop of an I_l frame, before and after applying the refinement step (in each channel separately). Notice the smoothness of the resulting contours.

LDR video creation

In order for the final result to have simultaneous detail visibility in the dark and bright zones of the picture, we need to combine the information from I_l and I_h . We do this by first bringing

the high-ISO image I_h to the range of I_l (simply dividing the values of I_h by the ratio between ISO values), obtaining an image we call I_{hl} , and then generating an HDR low-ISO image I'_l by a linear combination of I_l and I_{hl} : $I'_l(p) = (1 - \alpha)I_l(p) + \alpha I_{hl}(p)$, where p denotes pixel location and the weighting parameter α depends on the value of $I_{hl}(p)$. Let us note that, since the reconstructed HDR I'_l is a CFA RAW image, we convert it into an RGB image following the main stages in the color processing pipeline described in [4], without applying any gamma correction nor non-linearity: we perform denoising on CFA images [1], then apply demosaicing [20], then multiply by the corresponding color encoding and color characterization matrices. The final LDR image LDR'_l is obtained using the state-of-the-art tone mapping operator (TMO) [6], which can be applied to single pictures, as well as video, see Figure 7.



Figure 7. Left: dual-ISO input. Right: tone-mapped output from our method.

Results and Evaluation

The only parameters of our method are the noise standard deviation σ and the band size B (both parameters required by the deinterlacing step), plus the shape of the weighting function α . For all our tests, carried out on 12-bit input images, we use as default parameter values $\sigma = 100$, $B = 50$, and for the function α : $\alpha(x) = 0 \forall x < l_b$; $\alpha(x) = 1 \forall x > u_b$; $\alpha(x) = \frac{x - l_b}{u_b - l_b} \forall x \mid l_b \leq x \leq u_b$, with $l_b = 2048$ and $u_b = 0.8 * \max(I_{hl})$ (see plot of α in Figure 1).

We start by showing the effectiveness of our proposed approach. We take image data from the HDR survey of M. Fairchild [8] and generate our dual-ISO input pictures by interlacing pairs of lines coming from images with a 4-stop difference in exposure (the scenes are static, therefore different exposure times are equivalent to different ISO values). Applying our method to these pictures produces the results that appear in Figure 8, with simultaneous visibility of details in both dark and bright zones, and absence of noise or visual artifacts like spurious colors or halos.

Quantitative evaluation

For a quantitative evaluation we take our results and those obtained with the camera software Magic Lantern (ML) [16] and compare against the ground truth which can be obtained from the Fairchild database. We choose ML because its code is openly available, it seems to be the most widely used method for this application, and it provides intermediate outputs equivalent to our images I_l and I'_l . For the ground truth we select a sample of 20 scenes from Fairchild's HDR survey so that it includes night and daytime images, outdoor and indoor environments, varying from very high to low dynamic ranges: 507, Amikeus-BeaverDamPM1, BenAndJerrys, CanadianFalls, DevilsBathtub, Flamingo, GoldenGate1, HancockKitchenOutside, JesseBrown-Cabin, LabBooth, LetchworthTeaTable1, MammothHotSprings,



Figure 8. Four scenes from Fairchild dataset are shown: SmokyTunnel, WillyDesk, 507 and BenAndJerrys. The first row represents the simulated dual-ISO input, and the second row shows the final tone-mapped output of our method.

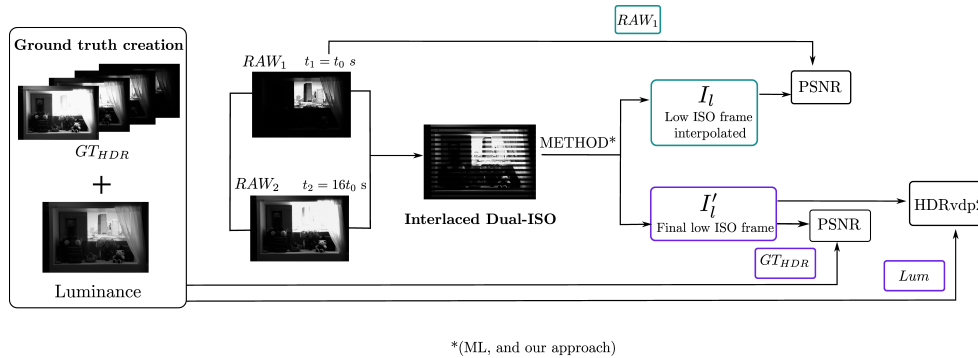


Figure 9. Evaluation diagram. For each scene we select two images RAW_1 and RAW_2 of 4-stops difference in exposure time, and interlace them to create the dual-ISO input. We compute as well the ground truth HDR image GT_{HDR} from the set of all the differently-exposed pictures that comprise the scene; Lum is the luminance of GT_{HDR} . We compute: PSNR between I_l and RAW_1 , PSNR between I'_l and GT_{HDR} , HDR-VDP-2 between I'_l and Lum .

MirrorLake, OtterPoint, SmokyTunnel, SunsetPoint1, TheNarrows2, WaffleHouse, WillyDesk, and Zentrum. All scenes consist of 9 RAW pictures taken with different exposure times. For each scene we generate our dual-ISO input by interlacing pairs of lines coming from pictures RAW_1 and RAW_2 that have a 4-stop difference in exposure. We apply ML and our method to these interlaced inputs, each algorithm providing intermediate outputs I_l (interpolated full-frame low-ISO image) and I'_l (that combines I_l and I_h). We have three ground truth images: the low-ISO (actually short-exposure) LDR image RAW_1 , the HDR image GT_{HDR} (obtained combining the 9 differently-exposed pictures with the method [10]), and the HDR image Lum which is the luminance of GT_{HDR} . The quantitative comparisons are the following: PSNR between I_l and RAW_1 , PSNR between I'_l and GT_{HDR} , and the HDR visibility difference metric HDR-VDP-2 [19] between I'_l and Lum . See Figure 9.

Table 1 shows the results, averaged over the 20 images of our sample of Fairchild's database; we can see that our approach outperforms ML for all three measures considered.

Qualitative evaluation

Figure 10 compares our (tone-mapped) result to those of Hajisharif *et al.* [12] and ML [16] on an actual dual-ISO picture (ISO values 100 and 1600) taken with a *Canon EOS 7D Mark III* camera (access to the camera was required for the application of the

	PSNR		HDR-VDP-2
	I_l vs RAW_1	I'_l vs GT_{HDR}	I'_l vs Lum
ML	47.71	29.89	54.17
Ours	48.79	30.58	57.43

Quantitative evaluation. Results for proposed method and for ML [16].

method of Hajisharif *et al.* [12], so those authors kindly provided the original image and their result for comparison). The first row uses as a tone-mapping operator our default method [6], while for the second row we have used another state-of-the-art TMO, that of Mantiuk *et al.* [18]. We can see that ML produces a result with severe color problems, due to the fact that the output of ML is not really an HDR image so the application of a TMO increases the overall brightness at the prize of creating unrealistic colors. On the other hand, the result of Hajisharif *et al.* [12] is of very good quality, but zooming-in on a textured area we can notice some color artifacts.

Figure 11 compares ML and our method for the Fairchild database scenes *HancockKitchenOutside* and *JesseBrownsCabin*. In this case the tone-mapping is not problematic for ML, but we can appreciate in the zoomed-in details (bottom row) how the results of ML are quite noisy.

Figure 12 presents the result of applying our method to a

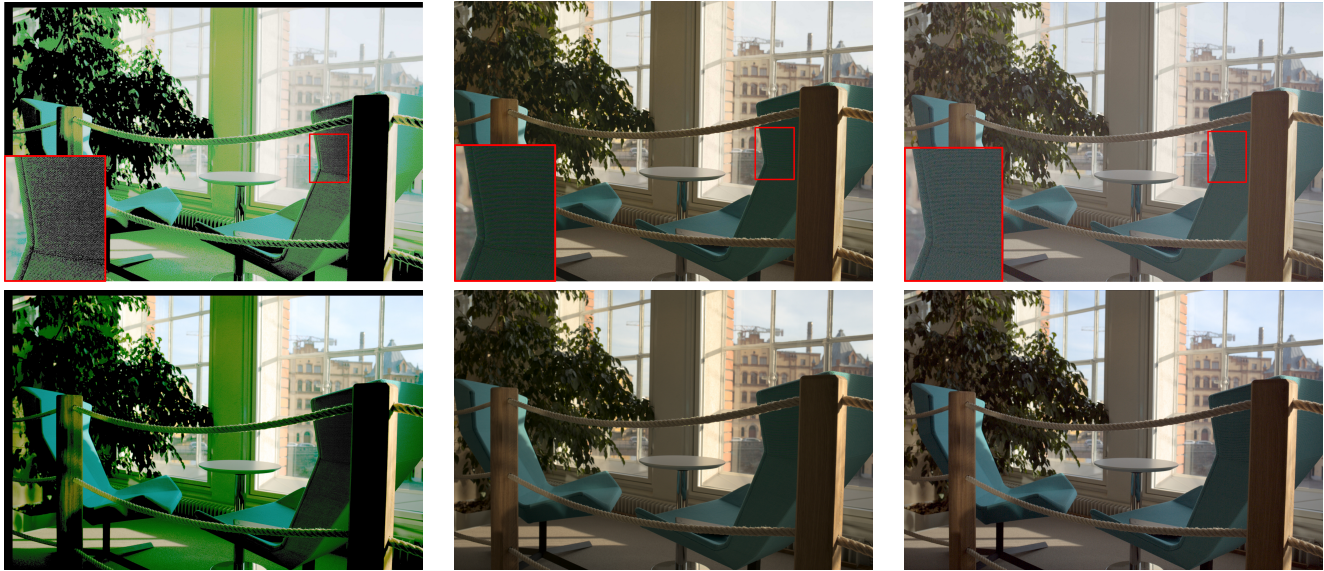


Figure 10. Left: tone-mapped result of ML [16]. Middle: tone-mapped result of Hajisharif et al. [12]. Right: our tone-mapped result. Top row: results tone-mapped with [6] (the default in our method). Bottom row: results tone-mapped with [18].



Figure 11. Final tone-mapped images of HancockKitchenOutside and JesseBrownsCabin scenes from Fairchild data. For each pair of images we show the output of ML (left), and of our approach (right). Notice how the results of ML are quite noisy.

dual-ISO video sequence taken with a *Canon EOS 5D Mark II* camera using ISO values 100 and 1600. The top row shows some frames from the interlaced dual-ISO input and the bottom row the corresponding LDR images obtained with our approach. This video has several moving objects, but none of the resulting frames shows noticeable visual artifacts of any kind, and the output LDR video doesn't suffer from temporal artifacts either.

Figure 13 compares ML and our method for two video sequences captured with a *Canon EOS 5D Mark II* camera using dual-ISO values 100 and 1600. Again, as it was the case in the static scenes of Figure 11, we can see that ML produces results that are rather noisy, as it can be appreciated on the shadowed parts of the images.

Finally, the top row of Figure 14 shows that while the default choice of parameter values produces good overall results, in some cases there can be some mismatches on the interpolated missing lines. The bottom row shows that these problems can

be significantly reduced with a better, image-dependent choice of parameters (in this case: $\sigma = 20$, $B = 100$).

Conclusion and future work

We have presented a practical approach for creating LDR video of HDR scenes from single-camera interlaced dual-ISO footage. Our method is affordable, since the input is obtained from a DSLR camera; it doesn't suffer from ghosting artifacts due to the dual-ISO information in each frame, instead of alternating exposures in consecutive frames; it minimizes spatial artifacts, thanks to the application of a high quality deinterlacing method; and it provides results with visible details in dark and bright areas, consistent with human perception, thanks to the tone mapping method used.

As future work, we plan to optimize the code aiming for real-time in-camera implementation, and to automatize the optimal selection of image-dependent or sequence-dependent parameter values.



Figure 12. Top row: frames from a dual-ISO video. Bottom row: results obtained with proposed method. The video does not show temporal artifacts.

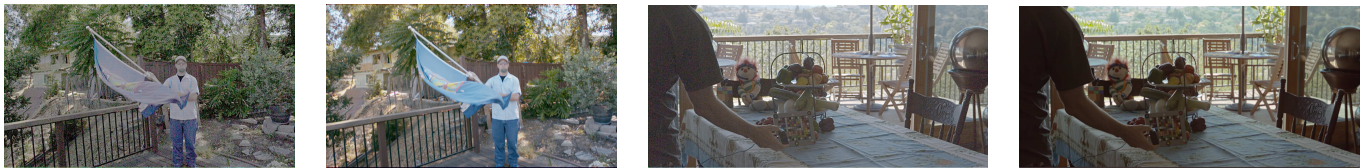


Figure 13. Tone-mapped results from two video sequences. (a) and (c): results of ML. (b) and (d): results of our approach.

Acknowledgments

The authors want to thank Pierre Jasmin from Re:Vision Effects Inc. and S. Hajisharif from Linköping University for providing us with dual-ISO images and videos for testing. This work was supported by the European Research Council, Starting Grant ref. 306337, by the Spanish government, grant ref. TIN2012-38112, and by the Icrea Academia Award.

References

- [1] H. Akiyama, M. Tanaka and M. Okutomi, Pseudo Four-Channel Image Denoising for Noisy CFA Raw Data, Proceedings of IEEE International Conference on Image Processing (ICIP2015), 2015.
- [2] C. Ballester, M. Bertalmío, V. Caselles, L. Garrido, A. Marques, and F. Ranchin, An Inpainting- Based Deinterlacing Method. Image Processing, IEEE Transactions on, 2007.
- [3] M. Bertalmío, Image Processing for Cinema, CRC Press, 2014.
- [4] S. Bianco, A. Bruna, F. Naccari, and R. Schettini, Color space transformations for digital photography exploiting information about the illuminant estimation process, JOSA A, Optical Society of America, 2012.
- [5] I. J. Cox, S. L. Hingorani, S. B. Rao and B. M. Maggs, A maximum likelihood stereo algorithm, Computer Vision and Image Understanding, 1996.
- [6] P. Cyriac, D. Kane and M. Bertalmío, Perceptual dynamic range for in-camera image processing, Proceedings of the British Machine Vision Conference (BMVC), 2015.
- [7] P. E. Debevec and J. Malik, Recovering High Dynamic Range Radiance Maps from Photographs. In Processings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97, 369-378, 1997.
- [8] M. D. Fairchild, The HDR photographic survey. Color and Imaging Conference, 2007 (1):223-238.
- [9] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, Andreas and H. Brendel, Creating Cinematic Wide Gamut HDR-Video

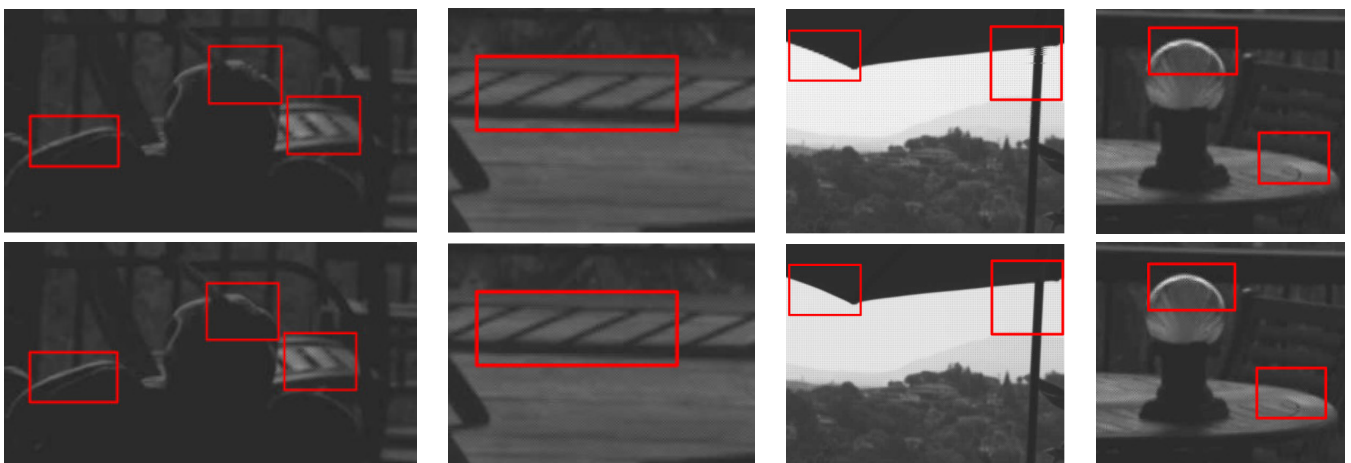


Figure 14. Top row: results of our method using default choice of parameter values, notice some artifacts. Bottom row: results with better, image-dependent parameter values.

- for the Evaluation of Tone Mapping Operators and HDR-Displays. Proc. SPIE, 2014.
- [10] R. Gil Rodríguez, J. Vazquez-Corral and M. Bertalmío, The intrinsic error of exposure fusion for HDR imaging, and a way to reduce it, Proceedings of the British Machine Vision Conference (BMVC), 2015.
 - [11] J. Gu and Y. Hitomi and T. Mitsunaga and S.K. Nayar, Coded Rolling Shutter Photography: Flexible Space-Time Sampling, IEEE International Conference on Computational Photography (ICCP), 2010.
 - [12] S. Hajisharif, J. Kronander, J. Unger, HDR reconstruction for alternating gain (ISO) sensor readout, Eurographics 2014 Short Papers 2014.
 - [13] F. Heide, M. Steinberger, Y. T. Tsai, M. Rouf, D. Paj kak, D. Reddy, G. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, K. Pulli, Flex-ISP: A Flexible Camera Image Processing Framework, ACM Trans. Graph. 2014.
 - [14] N. K. Kalantari, E. Shechtman, C. Barnes and S. Darabi, Dan B Goldman and P. Sen, Patch-based High Dynamic Range Video. ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH Asia 2013).
 - [15] S. B. Kang, M. Uyttendaele, S. Winder and R. Szeliski, High Dynamic Range Video. ACM Trans. Graph., ACM, 2003.
 - [16] Magic Lantern <http://www.magiclantern.fm/index.html>
 - [17] S. Mann and R. W. Picard, On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures, Proceedings of IS&T, 1995.
 - [18] R. Mantiuk, S. Daly, Scott and L. Kerofsky, Display Adaptive Tone Mapping, ACM Trans. Graph., 2008.
 - [19] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions, ACM Trans. Graph., 2011.
 - [20] D. Zhang, and W. Xiaolin, Color demosaicking via directional linear minimum mean square-error estimation, Image Processing, IEEE Transactions on, 2005.

Author Biography

Raquel Gil Rodríguez received her BS in Mathematics from Universitat de Barcelona and her MSc in Advanced and Professional Mathematics from Universitat de Barcelona, Spain and her Erasmus Mundus MSc in Computer Vision and Robotics from Université de Bourgogne, France. She has started her PhD in Image Processing for Enhanced Cinematography at Universitat Pompeu Fabra. Her current research interests are HDR imaging applications.

Marcelo Bertalmío received the Ph.D. degree in electrical and computer engineering from the University of Minnesota in 2001. He is an Associate Professor at University Pompeu Fabra in Barcelona, Spain. His interests are Image Processing and Computer Vision for digital cinema applications, although he prefers the (analog) films of Luis Buñuel and Luis García Berlanga.