

ANALYSIS OF THE CONSEQUENCES OF DATA QUALITY AND CALIBRATION ON 3D HDR IMAGE GENERATION

Jennifer Bonnard, Gilles Valette, Jean-Michel Nourrit, Céline Loscos

CReSTIC SIC, Université de Reims Champagne-Ardenne, 51100 REIMS, France

ABSTRACT

We propose to analyze consequences of input data quality on 3D HDR image generation. Input data are images from different viewpoints and different exposures. The ease and precision of 3D HDR images merging depends on how input data are created or acquired. We study the benefits and drawbacks of using an inbuilt multiview camera against a single camera with a simulation on computer generated images. This work builds on a previously published 3D HDR method based on disparity to guide HDR matching. In this paper, we outline the errors that occur when too little precaution is taken, coming on the one hand from poor pixel quality and on the other hand from poor geometrical setup.

Index Terms— 3D HDR videos, camera noises

1. INTRODUCTION

For the last two decades researchers have been interested in improving the dynamic range of images acquired from traditional cameras to reproduce all the intensity shades visible by human eyes. In 1995, Mann and Picard [1] first introduce High Dynamic Range (HDR) images, in opposition to traditional images called Low Dynamic Range (LDR) images. Their method to obtain these images consists of a weighted average of differently exposed images acquired from the same viewpoint. The book of Reinhard et al. [2] gives a detailed overview of existing approaches. In parallel, 3D was democratized by the Avatar movie (2009) allowing immersion and a better feeling of the action with the acquisition or synthesis of sequences from at least two viewpoints to recreate the binocular disparity used by human visual system.

The combination of 3D and HDR requires new sets of methods for their acquisition, manipulation, rendering and visualization. In Bonnard et al. [3], we proposed a method to acquire 3D HDR images which proved to be very sensitive to input data. The study presented here highlights input consequences on the final results. A test bed considered perfect is generated virtually for several viewpoints aligned along the axis supporting their optical centers. These data are altered to represent imprecisions and inaccuracies that we observed during our acquisitions, that were both geometric and colorimetric. We correlate these results with data acquired from

two different devices: an 8-objective video camera [4] (OctoCam) and a rail-mounted digital camera. In this paper, 3D HDR images correspond to the fusion of 8 images acquired from several viewpoints and for which visualization can be done on autostereoscopic screens (without glasses).

The case of 3D HDR reconstruction from input coming from more than two stereoscopic views brings out considerations already made for 2D HDR reconstruction, but for which solutions are more difficult to find. The HDR process requires to register pixels in terms of position and color. Altered position or color affect the registration process and impact on the quality of the HDR reconstructed values. For the generation of 2D HDR images, solutions were proposed both for alignment and color correction. Alignment is easier for rigid motion and linear transformations (translation and rotation) which find satisfactory solutions with Median Threshold Bitmap [5] or Normalized Cross Correlation [6]. Camera response curve functions have helped to linearize the color data [7] but can be estimated mostly for a unique viewpoint from one camera and a static scene. In our case, parallax occurs and these types of algorithms are not sufficient. Even optical flow [8] is not adequate because apparent motion of foreground objects can be large. Other methods were also proposed for stereo and multiview HDR generation, the interested readers may refer to [9] for explanations on each of them.

One solution in our case is to apply multiview camera calibration methods [10]. A chessboard can be used for geometric calibration and a color checker for color calibration. However, this requires a preprocessing step. In our experiments, we observed that systematic calibration before any acquisition helped but is not precise enough for HDR reconstruction when run on eight views.

In this paper, we study the geometric and colorimetric impact of approximations of the input data in the 3D HDR reconstruction. The goal is to identify factors that are difficult to compensate during calibration and registration in order to guide the design of new algorithms that will apply corrections during the HDR reconstruction phase. While analysis is done only on 3D HDR images produced with our approach, we believe that the conclusions of our study could benefit to the use of any other method. To guide our evaluation, we focus on parts of HDR images when errors are visually detectable and

we compute the associated PSNR (Peak Signal to Noise Ratio). Others methods can also be used for the evaluation such as the HDR-VDP-2 [11] but the PSNR is enough discriminating.

This article is organized as follows. Section 2 presents the pipeline used to generate 3D HDR images. Section 3 details precautions which must be taken on the geometrical setup and impacts on the 3D HDR process and section 4 is dedicated to the pixel color. In section 5 we present errors produced when the quality of input images is affected by both the geometrical and the color differences before concluding in section 6.

2. 3D HDR: MATERIALS & METHODS

In this section, we briefly present the 3D HDR method [3] used in this work.

Acquisition: The acquisition step is done by using a multiview camera [12] called OctoCam, equipped with eight horizontally aligned and synchronized objectives designed to deliver 3D content for auto-stereoscopic displays. This camera is based on a simplified epipolar geometry that permits strong assumptions on 3D stereovision algorithms [4] and horizontally align epipolar lines. Each of its sensors allows the acquisition of 10 bits per color channel. A neutral density filter is fixed on each objective; consequently, a different percentage of the light reaches the sensor for each view, hence acquired images represent different exposures. In this paper, we only use computer generated images as input. We reproduce the geometry of the camera to render eight images from aligned viewpoints of a synthetic scene using the POV-Ray ray-tracer.

Registration: The next step consists of the application of a pixel matching algorithm [13] to aggregate in a unique group called match, pixels which represent the same 3D point in the scene. The search for the corresponding pixels is only done on one line with the same ordinate thanks to the simplified epipolar geometry induced by the OctoCam, which increases the pixel matching algorithm speed.

HDR reconstruction: We adapted the Debevec and Malik's HDR method [14] for the 3D case. The HDR computation is not done on pixels occupying position in each image but on those belonging to the same match. Thereby all these pixels obtain the same HDR value called radiance:

$$\forall m \in \mathcal{M}, \quad \forall p = (i, j, k) \in m, \\ L_c(p) = \frac{\sum_{q \in m} w(Z_c(q)) Z_c(q) / \Delta t(q)}{\sum_{q \in m} w(Z_c(q))} \quad (1)$$

where m is a match, $p = (i, j, k)$ is a pixel of coordinates (i, j) in image k , $Z_c(p)$ is the value of the pixel p belonging to the match m on color channel $c \in \{R, G, B\}$, $\Delta t(p)$ is the exposure time of the image k in which the pixel p is and w is the weighting function.

We present in Fig. 1 a PSNR based comparison between two specific parts of the whole image (from view 0). In Fig. 2,

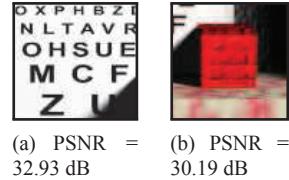


Fig. 1. Parts of the best tone mapped HDR images obtained with the 3D HDR process on synthetic images.

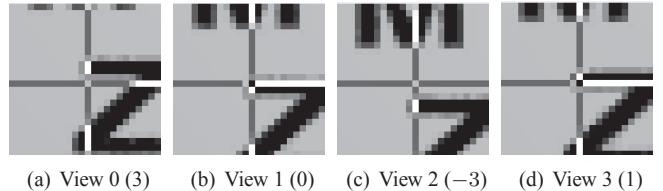


Fig. 3. Representation of a part of the misaligned synthetic images taken into account in the 3D HDR process. Here only four of the eight pixels belonging to a match representing the same 3D point in the scene are shown for better visibility. Digits in brackets corresponds to the displacement along vertical axis based on values found with the OctoCam.

are shown four of the eight synthetic images generated by POV-Ray ray-tracer (first line) and tone mapped HDR results obtained with the described 3D HDR process (second line). This example shows the best results our method [3] currently generates. Reference HDR images that represent ground truth in this paper are generated separately for each view using four simulated exposures per view and Debevec and Malik's method [14].

3. GEOMETRICAL SETUP

Most HDR reconstruction methods assume that input images are perfectly aligned so that each position (i, j) corresponds to the same 3D point. In our camera setup objectives are aligned so that only i is looked for during registration: a same 3D point lays at the same j coordinate in all images. Like for any HDR method, misalignment impacts on the 3D HDR process. The use of technology implies imperfect alignment even if a geometrical calibration is done to correct it. Fig. 3 shows an example of vertical misalignment. In that example, we focus on the Monoyer in the synthetic images to demonstrate misalignment. Pixels at the intersection of vertical and horizontal grey lines (added for better visibility) correspond to pixels belonging to the same match. In that case pixels detected as corresponding by our registration method are erroneous. In order to guide future correcting algorithms, we tested different cases and tried to estimate the impact of misalignment, varying the amount of vertical misalignment and the position of the image in which it occurs.

To begin with, we tested a misalignment of 1, 3 or 7 ver-

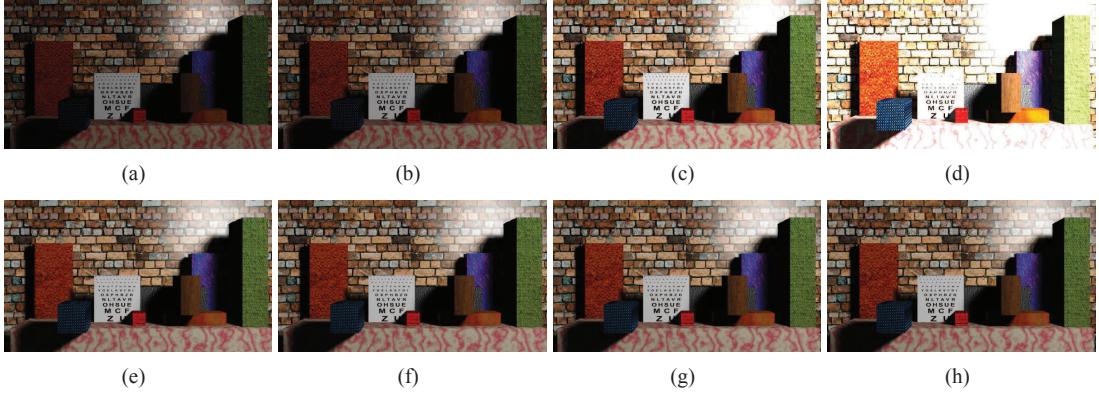


Fig. 2. First four images of a set of eight images by line. First line: input images with different exposures on each view. Second line : tone mapped HDR images generated with the 3D HDR process on each view.

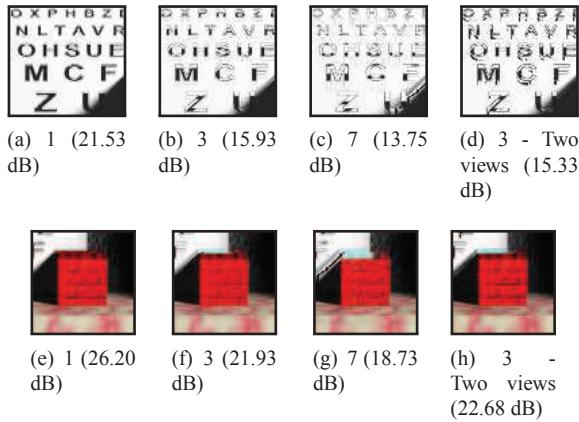


Fig. 4. Parts of tone mapped HDR images obtained on view 0 with different misalignment amplitudes in one view (a)-(c) and (e)-(g) and in two views (d) and (h). Digital values correspond to the shift amplitude.

tical pixels on one view to understand how the amplitude of vertical misalignment can affect the process. Images generated on view 0 when view 2 is misaligned are shown in Fig. 4. We can notice that one misaligned image with a shift of 1 pixel is acceptable. However, a 3-pixel shift already shows errors. We can see a deterioration of the results when increasing the misalignment and it gets slightly worse when on two views (Fig. 4(d) and (h)).

4. PIXEL QUALITY

In this section, we study noise types associated to sensors that can modify color pixel quality. They can be classified into two categories: temporal noises and spatially varying noises. Those involved in the first category change over time. This category contains the Photon Shot Noise (PSN), the

Dark Current Shot Noise (DCSN) and the Readout Noise also called Reset Noise. Spatially varying noises affect pixels according to their position in the acquired image. Photo Response Non Uniformity (PRNU) and Dark Current Non Uniformity (DCNU) also called Fixed Pattern Noise are in this category. The interested reader can find more details in [2, 7].

We focus on the estimation of the Noise Level Function that allows to determine the temporal noise that occurred in the OctoCam and the pixel color quality that is involved in the consistency between the color of pixels acquired on each sensor.

4.1. Noise Level Function

In this paper we use the Noise Level Function proposed by Liu et al. [15] to evaluate temporal noise due to the sensor itself. We acquired one set of 100 images in the same conditions (no change in acquisition parameters and in luminosity around the acquired scene). We estimated it on each sensor of the OctoCam and on each color channel separately because sensors are supposed to be identical but their actual response to noises is different.

In Fig. 5, we represent the standard deviation as a function of pixel values. The black curve corresponds to the regression obtained from the point cloud. This regression is an estimate of the researched Noise Level Function of the OctoCam. By using this function, we are able to generate noisy pixel values.

To show the temporal effect of the Noise Level Function, we computed 25 sets of 8 modified images and generated 3D HDR images for each set. We compared the 25 HDR images obtained on each view with the HDR reference by computing PSNR. Results on each view are presented in Table 1. Surprisingly, our measure shows that the Noise Level Function measured on our OctoCam has no impact in our 3D HDR process ($PSNR > 25$ dB). This can be understandable because the most affected pixels are over-exposed and are not taken into

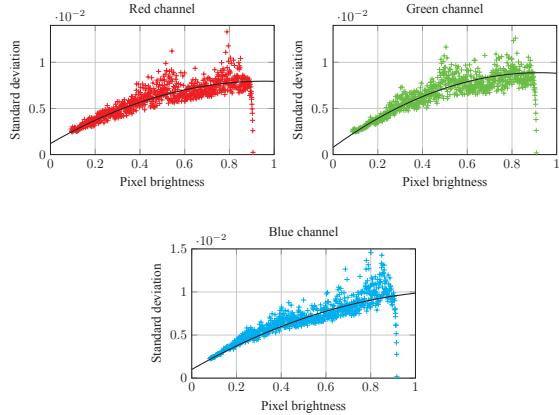


Fig. 5. Noise Level Function estimated using Liu et al. [15] on images acquired on view 4 of the OctoCam. Results are similar on the other views.

View	0	1	2	3
PSNR (dB)	26.7481	29.0453	31.9739	33.8696
View	4	5	6	7
PSNR (dB)	34.0809	32.7847	29.4415	27.0031

Table 1. PSNR values for estimating the impact of temporal noise on the 3D DHR process.

account in the HDR computation by the use of the weighting function. We have established that a unit to each value of the regression curve corresponds to a threshold not-to-exceed to avoid errors in the 3D HDR process.

4.2. Color quality

To simulate color inconsistency between sensors, we added a uniform noise with different amplitudes and then we computed 3D HDR images and compared results with references.

As presented in Fig. 6, even if the amplitude of the color is only of one color unit (for example a pixel of value 48 for a maximum of 1023 can have a value of 47, 48 or 49 in other images), we show that it is enough to produce visible errors on HDR images generated by the 3D HDR process. In this example, all the views are affected by these color differences. We note that different parts of the image are affected by colorimetric noise and results cannot be acceptable.

We tested the color inconsistency on only one of the eight views with an amplitude of one color unit to determine if the number of affected views can alter the results. We changed the position of the modified image in the 8-view sequence to see if it has or not an impact. We notice that only one view affected by color difference of only one color unit is enough to visualize errors independently of its position even if they are smaller than those produced when all views are affected.

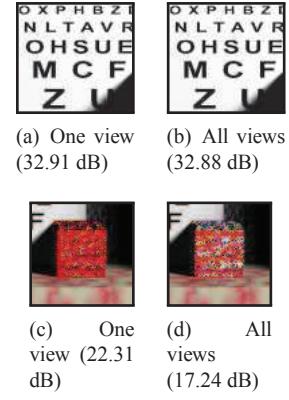


Fig. 6. HDR image generated on view 0 after the addition of a uniform random noise of amplitude 1.

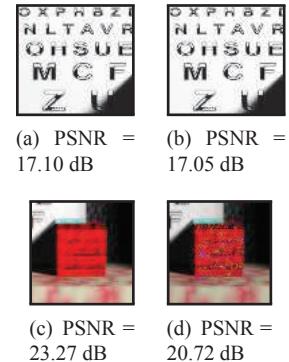


Fig. 7. Parts of tone mapped HDR images obtained with the combination of misalignment and color pixel inconsistency on different views (left) and on the same view (right).

5. GEOMETRY & COLOR

In this part we show the impact when there are errors in both geometry and color. We used conclusions obtained in the two previous parts to orient tests. In section 3, we demonstrated that a misalignment of three pixels shift is not acceptable and in section 4, we showed that a small error (one unit) in color consistency gives bad results. We evaluate in this part if the combination of these two smallest errors can be brought together without decreasing the quality of output results.

We constructed images with a misalignment of three pixels and we added the smallest color inconsistency. We moved the erroneous image on the eight views to see if it changed the impact on the 3D HDR image generation. We also tested the impact when one image is affected by misalignment and another one by color inconsistency. As presented in Fig. 7, the addition of these two sources of errors in one image gives poorer results than when on individual sources.

6. CONCLUSION

In this article, we presented impacts of the input data quality on 3D HDR reconstruction process first by studying misalignment, secondly by treating pixel quality with the Noise Level Function and inconsistency, and finally by using a combination of both noises. We notice that a three pixel shift affects the process and the smallest error in color consistency creates wrong results. The combination of errors both in geometry and color causes an increase in defects in the images produced that cannot be accepted. One solution would be to work on new calibration procedures to reduce misalignment and noises when working with 8-view cameras. Another solution is to use the results of the study to guide the HDR reconstruction and take into account the data inaccuracy when combining input values (new weight function). It is this direction that we planned to work on the near future.

ACKNOWLEDGEMENTS

This work is funded by the regional project ESSAIMAGE 3D-HDR. Authors of this paper participate within the COST action HDRI IC1005.

REFERENCES

- [1] 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*, 1995, pp. 442–448.
- [2] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, The Morgan Kaufmann series in Computer Graphics. Elsevier (Morgan Kaufmann), Burlington, MA, 2nd edition, 2010.
- [3] J. Bonnard, C. Loscos, G. Valette, J.-M. Nourrit, and L. Lucas, “High-dynamic range video acquisition with a multiview camera,” *Proc. SPIE*, vol. 8436, pp. 84360A–0 – 84360A–11, 2012.
- [4] J. Prévost, L. Lucas, and Y. Remion, “Shooting and viewing geometries in 3DTV,” in *3D Video: From Capture to Diffusion*, L. Lucas, C. Loscos, and Y. Remion, Eds., pp. 71–90. Wiley-ISTE, 2013.
- [5] G. Ward, “Fast, robust image registration for compositing high dynamic range photographs from handheld exposures,” *JOURNAL OF GRAPHICS TOOLS*, vol. 8, pp. 17–30, 2003.
- [6] R. Ramirez, C. Loscos, I. Martin, and A. Artusi, “Patch-based registration for auto-stereoscopic hdr content creation,” in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.
- [7] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch, “Optimal hdr reconstruction with linear digital cameras,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 215–222.
- [8] S. B. Kang and R. Szeliski, “Extracting view-dependent depth maps from a collection of images,” *Int. J. Comput. Vision*, vol. 58, no. 2, pp. 139–163, July 2004.
- [9] J. Bonnard, G. Valette, C. Loscos, and J.-M. Nourrit, “3D HDR images and videos: Acquisition and restitution,” in *3D Video: From Capture to Diffusion*, L. Lucas, C. Loscos, and Y. Remion, Eds., pp. 369–383. ISTE Ltd., 2013.
- [10] V. Nozick and J.-B. Thomas, “Camera calibration: geometric and colorimetric correction,” in *3D Video: From Capture to Diffusion*, ISTE Ltd., L. Lucas, C. Loscos, and Y. Remion, Eds., pp. 91–110. Wiley-ISTE, 2013.
- [11] 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.*, vol. 30, no. 4, pp. 40:1–40:14, July 2011.
- [12] J. Prévost, S. Chalençon-Piotin, D. Debons, L. Lucas, and Y. Remion, “Multi-view shooting geometry for multiscopic rendering with controlled distortion,” *International Journal of Digital Multimedia Broadcasting (IJDMB), special issue Advances in 3DTV: Theory and Practice*, vol. 2010, pp. 1–11, Mar. 2010.
- [13] C. Niquin, S. Prévost, and Y. Remion, “An occlusion approach with consistency constraint for multiscopic depth extraction,” *International Journal of Digital Multimedia Broadcasting (IJDMB), special issue Advances in 3DTV: Theory and Practice*, vol. 2010, pp. 1–11, Feb. 2010.
- [14] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *Proceedings of SIGGRAPH ’97, Computer Graphics Proceedings, Annual Conference Series*, 1997, pp. 369–378.
- [15] C. Liu, R. Szeliski, S.-B. Kang, C.L. Zitnick, and W.T. Freeman, “Automatic estimation and removal of noise from a single image,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 299–314, Feb 2008.