

Consistent Multi- and Single-View HDR-Image Reconstruction from Single Exposures

A. Mohan¹ and J. Zhang^{1,2} and R. Cozot² and C. Loscos¹

¹LICIIS laboratory, University of Reims Champagne-Ardenne, France

²LISIC laboratory, University of Littoral Côte d'Opale, France,



Figure 1: Comparing our HDR Reconstructed outputs with those of two other algorithms. From Left to Right: Input LDR Image, Ground Truth Image, HDRCNN Output, MaskCNN Output, Our algorithm.

Abstract

We propose a CNN-based approach for reconstructing HDR images from just a single exposure. It predicts the saturated areas of LDR images and then blends the linearized input with the predicted outputs. Two loss functions are used: the Mean Absolute Error and the Multi-Scale Structural Similarity Index. The choice of these loss functions allows us to outperform previous algorithms in the reconstructed dynamic range. Once the network trained, we input multi-view images to it to output multi-view coherent images.

1. Introduction

Recently, there have been attempts to obtain high-dynamic range (HDR) images from single exposures and efforts to reconstruct multi-view HDR images using multiple input exposures. However, there have not been any attempts to reconstruct multi-view HDR images from multi-view Single Exposures to the best of our knowledge. We present a two-step methodology to obtain color consistent multi-view HDR reconstructions from single-exposure multi-view low-dynamic-range (LDR) Images. We define a new combination of the Mean Absolute Error and Multi-Scale Structural Similarity Index loss functions to train a network to reconstruct an HDR image from an LDR one. Once trained we use this network to multi-view input.

2. CNN-Based Single-View HDR Reconstruction

In order to train our neural network, we collected 3948 HDR images on the website according to the source list in [MYK], then

used a virtual camera and patch cropping to generate training data. The CNN architecture used is similar to the one used in the paper by Eilertsen et al. [EKD*17]. It contains an autoencoder with skip connections in between. The encoder converts an LDR input to a latent feature representation, and the decoder reconstructs this into an HDR image in the log domain while the skip connection is used to transfer each level of the encoder to the corresponding level on the decoder side.

Zhao et al. [ZGFK17] suggests a combination of the Multi-Scale SSIM (MS-SSIM) and the Mean Absolute Error (MAE) loss functions in image restoration. Since these loss functions are used to improve the quality of reconstructions, we adopted a similar approach in our research. We combine the MAE and MS-SSIM [WSB] to formulate our loss function, which is given by:

$$L(y, \hat{y}) = \lambda_1(MAE) + \lambda_2(1 - MS-SSIM) \quad (1)$$

The pixel-wise function MAE accurately models the colors and lu-

minance while the perceptual loss MS-SSIM preserves the contrast of high-frequency regions - both of which are essential to reconstructing HDR textures. Finally, we combine both of these concepts to formulate our loss function, which is given by:

$$L(y, \hat{y}) = \lambda_1(MAE) + \lambda_2(1 - MS-SSIM) \quad (2)$$

λ_1 and λ_2 were experimentally chosen as 0.5 each. Using a combination of MAE and MS-SSIM as a cost function for the reconstruction produces satisfactory results because they perform complementary tasks used to preserve important aspects of the image. The pixel-wise function MAE accurately models the colors and luminance while the perceptual loss MS-SSIM preserves the contrast of high-frequency regions - both of which are essential to reconstructing HDR textures. The outputs of the Neural Network containing the predictions in the saturated area are mixed with the linearized input using a blending equation.

Extending to multi-view HDR image reconstruction We extend our network to reconstruct multi-view camera grid images so that the adjacent images are color consistent. Passing each of the single views of the multi-view images into the network would not be optimal because (i) predicting views separately would lead to some redundant computations and (ii) it will not take into account neighboring image information, thereby rendering the set of images inconsistent with respect to color among different views. Therefore, we rely on passing image grids to the network to reconstruct the saturated regions. However, predicting the image grid as a whole can cause problems since (i) predicting information for such a large image would require enormous computational power and resources, and (ii) the down scaling carried out in the network internally, may lead to loss of some resolution. As a trade-off between these two aspects, we divide the grid into sub-grids - generally consisting of four images, to carry out the prediction. Due to the network's down-scaling, convolutional and pooling capabilities, the pixel-wise prediction of the downscaled images are affected by its neighbouring pixels, thereby improving coherence among multiple views.

3. Results

Single-view evaluation

We compare our algorithm with two other state-of-the-art algorithms for HDR reconstruction using single exposure: HDRCNN [EKD*17] and MaskCNN [SRK20] (see figure 1). Table 1 shows the metrics computed over 40 different images. We have observed that our algorithm performs well in reconstructing the intensities of the HDR Samples. We also measured the dynamic range retention, i.e., on how estimated dynamic ranges are comparable to those of ground truth (base-2 logarithmic value of the absolute difference between the maximum valued and minimum valued pixels in the image). The last line of Table 1 shows that our algorithm achieves a lower error than the HDRCNN and MaskCNN algorithms, thereby indicating that the images obtain a dynamic range which is close to the ground truth one.

Multi-view Consistency Assessment

We ran our consistency test on a set of real images captured using a camera array at 2 different exposures to build an HDR ground truth image. We ran the algorithm on a set of four LDR images as described in section 2. Our original training algorithm inputs

Metric	HDRCNN	MaskCNN	Our
HDR-VDP	33.75	34.02	34.43
PSNR	57.59	57.54	58.60
Harmonic-IQA	0.314	0.315	0.310
SSIM	0.29	0.29	0.29
Dynamic Range Error	1.051	1.026	0.921

Table 1: Scores averaged over 40 images of different exposures and scenes.

images of 2500×2100 . We adapt it to fit a size of 5000×4200 of input LDR; this upper limit is set by the memory buffer size of our testing GPU, a NVIDIA Tesla P100 GPU. Four of these GPUs were also used for training. Results are shown in Table 2. As seen in the table, the metrics of Sum of Absolute Difference and Normalized Cross Correlation metrics, which are used for coherence evaluation, score better in prediction through a grid view.

Metric	Independent Views	Grid Views
SAD	3831483.42	1388318.70
NCC	0.014	0.22

Table 2: Results of the consistency evaluation.

4. Conclusion

We present a new methodology for a two-stage CNN network applied to multi-view HDR reconstruction and evaluated its efficiency with respect to state-of-the-art algorithms. The loss functions we have incorporated provide an alternative method to obtain the HDR reconstructions with a better dynamic range retention when compared to existing algorithms. We demonstrate that such a network is adapted to multi-view imaging, allowing to achieve consistency among the multiple views. With this approach, we successfully provide a pipeline to multi-view HDR Images tailored for a wide range of applications - including 3D reconstruction.

Acknowledgements

This work was funded by ANR-ReVeRY (ANR-17-CE23-0020) and used URCA computing facilities of *Centre Image* and *ROMEIO*.

References

- [EKD*17] EILERTSEN G., KRONANDER J., DENES G., MANTIUK R. K., UNGER J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics* 36, 6 (2017), 1–15. doi:10.1145/3130800.3130816. 1, 2
- [MYK] MORIWAKI K., YOSHIHASHI R., KAWAKAMI R.: Hybrid Loss for Learning Single-Image-based HDR Reconstruction. 21. 1
- [SRK20] SANTOS M. S., REN T. I., KALANTARI N. K.: Single image hdr reconstruction using a cnn with masked features and perceptual loss. *ACM Transactions on Graphics* 39, 4 (2020). doi:10.1145/3386569.3392403. 2
- [WSB] WANG Z., SIMONCELLI E., BOVIK A.: Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, IEEE, pp. 1398–1402. doi:10.1109/ACSSC.2003.1292216. 1
- [ZGFK17] ZHAO H., GALLO O., FROSIO I., KAUTZ J.: Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging* 3, 1 (2017), 47–57. doi:10.1109/tci.2016.2644865. 1