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

Aditya Mohan[1] , Jing Zhang[1,2], Rémi Cozot[2] and Céline Loscos[1]
1. LICIIS laboratory, University of Reims Champagne-Ardenne, France
2. LISIC laboratory, University of Littoral Côte d'Opale, France

OVERVIEW

Our **goal** is to reconstruct High-dynamic range (HDR) values for each image of an low-dynamic range (LDR) camera grid.

Input: LDR images of an LDR camera grid

⬇

Output: consistent HDR images

RELATED WORK

CNNs have opened up the possibility of reconstructing HDR images from just single-exposure LDR inputs. Eilerstsen et al. [EKD*17] use a U-NET-based architecture to predict the saturated regions and use a blending function to combine the original image with the predicted one. They optimize the illuminance and reflectance cost functions to achieve their results.

Kalantari et al. [SRK20] propose a feature masking mechanism that reduces the ambiguities caused by invalid information in the saturated areas, and makes use of a combination of L1 loss, VGG-based perceptual loss, and the Style loss function to optimize their network.

A method to obtain multi-view images with a High Dynamic Range Texture by Lu et al. [LJDE11] relies on multi exposure capturing - a constraint that we would tackle in our research.

Acknowledgements

anr® ReVeRY ANR-17-CE23-0020

Centre Image Maison de la Simulation de Champagne-Ardenne

Université de Reims Champagne-Ardenne

LISIC

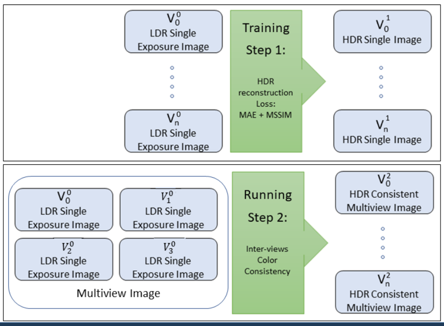
ulco Université Littoral Côte d'Opale

LICIIS DIGIT

EG'22 TheEvent

METHODOLOGY

We propose a two-step CNN-based approach for reconstructing HDR images from just a single exposure. It predicts the saturated areas of LDR images and then blend the linearized input with the predicted outputs. Once the network trained, we input multi-view images to it to output multi-view coherent images.



Training

Training was made on crops of 3948 HDR images collected from the source list in [MYK] and containing saturated areas in its middle.

Step2: multi-view HDR image reconstruction

We extend our network to reconstruct multi-view camera grid images so that the adjacent images are color consistent. We divide the grid of images into sub-grids - generally consisting of four images, to carry out the prediction.

Step1 - CNN-Based Single-View HDR Reconstruction

The **CNN architecture** is an autoencoder with skip connections in between like[EKD*17]. 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.

Loss functions: a combination of the Multi-Scale SSIM (MS-SSIM) and the Mean Absolute Error (MAE) loss functions like [ZGFK17]. For the pixels $\hat{y}$ of the predicted image and y of the ground truth image:

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

The HDR predictions in the saturated area (extracted through a mask) are mixed with the linearized input using a blending equation:

$$y_{out} = (1 - msk)x_{lin} + msk(y_{predict})$$

The **Mean Absolute Error (MAE) Loss function** gives the absolute distance between two corresponding pixels of two images, and accurately models the colors and luminance :

$$MAE = \frac{\sum_{c,i}|\widehat{y_{c,i}} - y_{c,i}|}{3.n}$$

for the i^th pixels $\widehat{y_{c,i}}$ of the predicted image and $y_{c,i}$ of the ground truth image, with c the color channel.

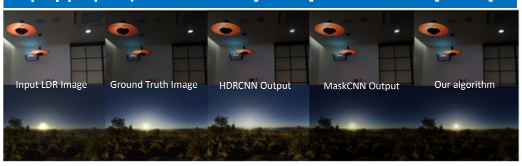
The **Multi-Scale SSIM (MS-SSIM) Loss function** [WSB] preserves the contrast of high-frequency regions. The equation, in terms of luminance, contrast, and structure is given for M scale iteration as:

$$MSSSIM(x,y) = [l_{M(x,y)}]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}$$

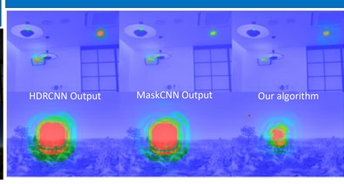
RESULTS

Step1 - CNN-Based Single-View HDR Reconstruction

HDR reconstruction comparison (tone mapped using [RD05] for display purpose) with HDRCNN [EKD17] and MaskCNN [SRK20]



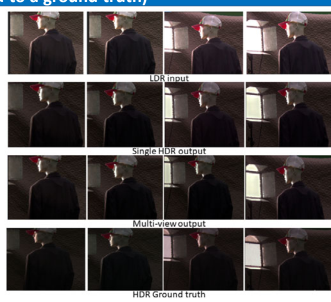
HDR-VDP visual errors (red is high and blue is low)



Our visual results in single-view HDR reconstruction are similar to previous approaches, but closer to the ground truth for high intensity values.

Step2: multi-view HDR image reconstruction

Multi-view consistency results on a real dataset compared to a ground truth)



Our original training algorithm inputs images of 2500x2100. We adapt it to fit a size of 5000x4200 of LDR input; 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.

Metrics

Scores averaged over 40 images of different exposures and scenes

Metric	HDRCNN	MaskCNN	Ours
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

Metrics: SSIM, PSNR, HDR-VDP [MKRH11], Harmonic_HDR_IQA [RDMC19]

Our algorithm outperforms the state-of-the-art on dynamic range retention.

The dynamic range error of an image is the base-2 logarithmic value of the absolute difference between the maximum valued and minimum valued pixels in the image.

Metric	HDRCNN	MaskCNN	Ours
Dynamic Range Error	1.051	1.026	**0.921**

Metric	Independant views	Grid Views
SAD	3831483.42	**1388318.70**
NCC	0.014	**0.22**

The coherence evaluation metrics (Sum of Absolute Difference and Normalized Cross Correlation), score better in prediction through a grid view.

REFERENCES

[EKD17] 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.
[LJDE11] LU F., JI X., DAI Q., ER G.: Multi-view stereo reconstruction with high dynamic range texture. Computer Vision – ACCV 2010 Lecture Notes in Computer Science (2011), 412–425.
[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: Hdr-vdp-2. ACM SIGGRAPH 2011 papers on - SIGGRAPH 11 (2011).
[MYK] MORIWAKI K., YOSHIHASHI R., KAWAKAMI R.: Hybrid Loss for Learning Single-Image-based HDR Reconstruction. arXiv:1812.07134. (2018).
[RD05] REINHARD E. and DEVLIN K. : Dynamic range reduction inspired by photoreceptor physiology, IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 1, pp. 13-24, Jan.-Feb. 2005
[RDMC19] ROUSSELOT M., DUCLOUX X., MEUR O. L., COZOT R.: Quality metric aggregation for hdr/wcg images. 2019 IEEE International Conference on Image Processing (ICIP) (2019).
[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).
[WSB] WANG Z., SIMONCELLI E., BOVIK A.: Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, IEEE, pp. 1398–1402.
[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.