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

A Mohan, Jing Zhang, Rémi Cozot, C Loscos

Eurographics Workshop on Intelligent Cinematography and Editing (2022)
R. Ronfard and H.-Y. Wu (Editors)

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

A. Mohan[1,3] and J. Zhang[1,2] and R. Cozot[2] and C. Loscos[1]

[1]LICIIS laboratory, [2] LISIC laboratory, University of Littoral Côte d'Opale, France,
[3] BITS Pilani K.K. Birla Goa Campus, India

Abstract

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. When tested on single images, the outputs achieve competitive results with the state-of-the-art. Quantitative and qualitative metrics applied to our results and to the state-of-the-art show that our HDR expansion is better than others while maintaining similar qualitative reconstruction results. We also demonstrate that applying this network on multi-view images ensures coherence throughout the generated grid of HDR images.

CCS Concepts

• ***Computing methodologies*** → ***Computational photography; Machine learning; 3D imaging;***

1. Introduction

In the last two decades, creative industries and researchers proposed significant advances in media content acquisition systems in three main directions: increase of resolution and image quality with the new ultra-high-definition (UHD) standard that uses 3840x2160 pixels resolution (also called 4K resolution); stereo capture for 3D content (depth information); and high-dynamic range (HDR) imaging raising the dynamic range of the image to at least 16-fstops. These recent advances addressed the full media production pipeline: acquisition, image data enhancement, and display, with the development of 3D and grid cameras, HDR imaging, UHD resolution, autostereoscopic displays, immersive VR headsets, HDR displays. These new technologies raise incontestable enthusiasm by both professionals and end users, but are currently limited by low creative content potential. For instance, todays offered 360° panoramic image for VR immersive visualization would not be convincing for a natural light outdoor landscape. The user would be perceptually limited in the range of intensity and restricted to rotating navigation.

1.1. Project context

The ANR ReVeRY project is a collaborative project addressing, among other objectives, solutions to enable user perception of high intensity ranges as well as free navigation inside the scene in an embedded distributed media adaptive to the diversity of nowadays displays. In other words, there should be no capability difference when virtually visualizing real or synthetic scenes.

The ReVeRY project conducts fundamental research to address the full pipeline from acquisition to display. Our team focus on the acquisition step, in order to deliver a depth+HDR point cloud. In this paper, we concentrate on the acquisition step targeting high-resolution, HDR values. Key limitations addressed are:

1. Higher resolution is still requested (for example, to increase definition of small objects in large views);
2. Stereo streams are shot under fixed geometrical choices (i.e., rig-to-scene distance, inter-ocular distance, and baseline orientation) restraining or focusing their use according to specific display geometries (movie theater vs. 3DTV) and limiting the postproduction reframing possibilities (i.e., frame rotation destroys depth effect as image rows are no longer parallel to the rig baseline);
3. HDR imaging technology, whose deployment to HDR-enabled home TVs is imminent, is still nowadays limited by too few HDR video acquisition solutions;
4. Capturing different images with varying exposure times remains challenging in the case of dynamic scenes with movement or light changes.

DOI: 10.2312/wiced.20221050

Figure 1: *The ReVeRY camera built by our partner, XD Productions. It is a 16-image synchronised camera system. It is a wide-baseline camera array, with cameras apart of 20cm.*

The ReVeRY project aim is to overcome the following current limitations of current media to the creative industries:

1. The traditional acquisition pipeline uses and outputs video streams, either UHD, stereo, or HDR. After shooting, possible changes on the media (i.e., viewpoint, framing, aperture, lighting) are limited, which may force a film director to shoot the scene again if any changes are necessary;
2. For now, a pre-shooting choice is made according to the nature of the media to capture (2D, stereo, or HDR) depending on the foreseen use. However, these choices are currently never jointly offered.
3. Advanced acquisition systems for specific displays (autostereoscopic, HDR, VR headset) are often at the stage of prototype. Quality delivery often implies the industries to limit themselves to traditional content converted with heuristics to specific outputs, with the risk of visible artifacts and user discomfort.

The ReVeRY project ambitions are to propose the ReVeRY media, a joint embedded UHD, HDR, and 3D information in a dedicated format, along with one or more suitable representations, to develop a demonstrative prototype of a dedicated acquisition system, and to prove their benefits to the media creative industry. The main project objective is to manage and process specifically designed multi-video data, in order to compute, manipulate, and deliver an innovative, rich high quality (HDR, UHD) video+depth media.

In this project, a new camera-grid system prototype was built for acquiring a multiview/multiexposed video. The 16-camera system is illustrated in FIGURE 1. It is a cost-efficient, camera-grid system to acquire at once several viewpoints under several exposures. One objective of the reVeRY project is to provide new algorithms dedicated to HDR+depth reconstruction from multiview/multiexposed raw data, and converting this reconstruction output into an easy-to-use HDR UHD video+depth media.

1.2. Research statement

Multi-view images have been used for many areas in Computer Graphics and Computer Vision, with 3D content reconstruction and lightfield imaging being the most extensively used applications. A major targeted breakthrough of the project was to resolve jointly HDR and Depth reconstruction. 3D reconstruction by stereovision relates to the automatic depth extraction of a 3D scene structure from different viewpoints (2 to n) acquired at the same time. It comes down to match all homologous pixels from the 3D point projections on the n images. The ReVeRY project considers simplified multi-epipolar geometry, reached either by using directly a capture configuration in parallel geometry or applying a pre-processing step of rectification on each image [HZ04], leading to epipolar lines parallel to image columns or rows. A matching scheme defines data similarity (or dissimilarity) within a given neighborhood between potential homologous pixels, rendered difficult by lack of information in images (such as occlusions) or ambiguous information (such as homogeneous/repetitive area or luminosity variations). Two classes of methods have been proposed: hybrid and multiscopic matching. Hybrid methods combine a pixel-per-pixel matching with structured primitives such as regions obtained by per-image segmentations the quality of depth estimation. However, this is almost dedicated to the binocular case and the risk of inconsistency between regions in different images increases with the number of views. Multiscopic methods usually gain robustness with information redundancy by computing simultaneously the n depth maps and a minimum of four views is needed [NPR10a] [NPR10b]. The ReVeRY camera array system, a regular rectangular grid, provides simplified multi-epipolar geometry and enough redundancy.

Combining HDR with stereovision enables high-quality depth perception reproduction of real-world scenes, but few contributions have been made in this domain. Lately, solutions were proposed for 3D HDR images with stereo cameras [LC09] [SMW10] or multi-stereo cameras [BLV*12] following stereovision-based procedures, with remaining inaccuracies in under- or over-exposed areas [BVL19]. This can be improved using patch-matching along the epipolar line [OLMA13] but spatial coherence is lost.

There were several solutions envisaged. The first solution was to acquire in a single shot different exposures per view and to solve for HDR and depth in a same algorithm. However, this kind of algorithm was overruled at the mid-point of the project. First, it is difficult to synchronize camera which have a different exposure time. Second, solving depth with varied exposures has proven difficult [BVL19] [OLMA13]. With the raise of machine learning algorithms, it was soon realized that we could investigate towards other solutions. We decided to work separately on two types of algorithms, one to solve depth and the other one to solve HDR imaging. With the upgrade of sensor dynamic range, we bet on the fact that there could be sufficient information to solve for depth. As we will analyse in the previous work SECTION (see SECTION 2, there are now possibilities to use machine learning to estimate HDR values from a single exposure image.

1.3. Contributions

Our work focuses on reconstructing *consistent* multi-view high-dynamic range (HDR) images in addition to reconstructing single view images to feed a multi-view camera system with HDR content. It is very likely that using multi-view HDR images would en-

hance the 3D reconstruction, as long as we provide calibrated and consistent inputs.

We propose a CNN-based approach for reconstructing HDR images from just a single LDR exposure per view, described in SECTION 3. This prevents difficulties linked to multi-exposure acquisition like synchronizing video input. 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 is trained, we input multi-view images to it to output multi-view consistent images. Our results are described in SECTION 4.

2. Related work

The current reconstruction of Single View High Dynamic Range Images primarily rely on techniques using either inverse tone mapping operators (iTMO), generating multi-exposure stack and by directly passing the single LDR Image to a function which reconstructs the HDR equivalent. Multi-View reconstructions often rely on Multi Exposure LDR Inputs which could possible cause difficulties in relation to the lack of availability of multiple exposures or pixel misalignment across multiple views and exposures. We provide a study of the current techniques in relation to our proposed methodology.

2.1. Single-View HDR Reconstructions

Using inverse tone mapping operators such as introduced in Banterle et al., 2007 [BLDC06] is one way to retrieve HDR Content. Tone mapping operators are generally used to retain the contrast of the image while compressing the range of illuminance in order to display HDR Images on typical display devices. Inverse tone mapping operators aim to achieve the opposite, to produce an HDR Image from an LDR one, like using the median cut algorithm by Debevec et al. [Deb08].

Another method to generate HDR Content arises from Debevec's [DM08] method to merge multiple exposures of LDR images. This is exemplified in the research proposed by Lee et al [LAK18]. They propose the use of Generative Adversial Networks loss coupled with the L1 loss to generate multiple exposures of an image in order to merge them and form the HDR image. Endo et al. [KK19] synthesizes bracketed LDR Images by learning the relative change of each pixels when the exposure is increased/decreased.

Yet another method is to generate HDR Images directly from a single exposure LDR Image. This is exhibited in the work by Eilerstsen et al. [EKD*17] which uses a U-NET based architecture to predict the saturated regions and uses a blending function to combine the original image with the predicted one. They optimize the illuminance and reflectance cost functions to achieve their results. Santos et al [SRK20a] 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, VGG and the Style loss to optimize their network. Our results would be primarily displayed and evaluated against the two algorithms mentioned above.

Our work approach most relates to [EKD*17] and [SRK20a] because it uses neural networks to predict saturated areas from single exposures. However, we propose different loss functions, and extend to multi-view data.

2.2. Multi-View HDR Reconstructions

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 we would tackle in our research. One of the reasons why reconstructing Multi-View Stereo HDR using multi exposure is not very practical is because they require the conversion of pixel values to relative scene radiance values to accurately model the saturated regions. For this conversion, highly accurate pixel alignment is required across the multiple exposures which is not always a guarantee in the case of Multi View Stereo systems.

Troccoli et. al [TKS06] explains how there is a lack of HDR Texture 3D Reconstructions due to the possible violation of the brightness constancy constraint among multiple exposures and proceeds to introduce an exposure invariant matching technique to create 3D Reconstructions. However, the reliance on multi-exposure images is still a problem since the availability of different exposures of the same image is not always guaranteed, and poses acquisition issues.

3. CNN-Based Single-View HDR Reconstruction

We propose a CNN-based approach to obtain consistent multi-view HDR outputs. Our pipeline is shown in FIGURE 2. An initialisation step predicts the saturated regions of an input LDR image. In a first step, it feeds a learning network to obtain HDR values. In a second step, this trained network is then run on a multi-view LDR map to infer coherent HDR images of a multi-view camera grid.

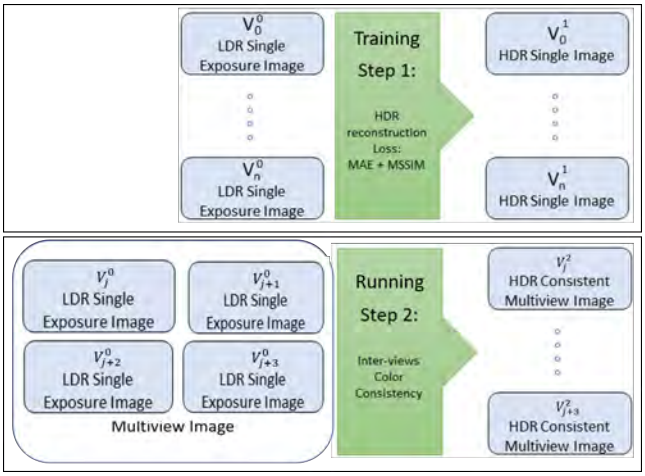

Figure 2: *Our two-stage pipeline for consistent multi-view HDR reconstructions. First step: training HDR reconstruction for a single view. Second step: ensuring color consistency between views by predicting through a map containing all views. Here, $V^i{}_j$ implies the j_{th} view of the results of the i_{th} step.*

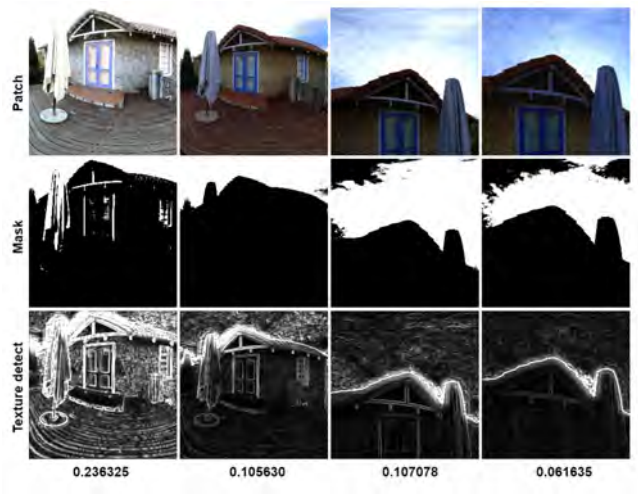

Figure 3: *Patch sampling result, the top row shows patch cropping result in a same HDR image, the second row is the related mask for overexposed area, the third row is detected texture with patch sampling algorithm [SRK20b], and the number in the bottom is the texture score for each patch, from left to right, the smaller the score, the fewer textures it contains.*

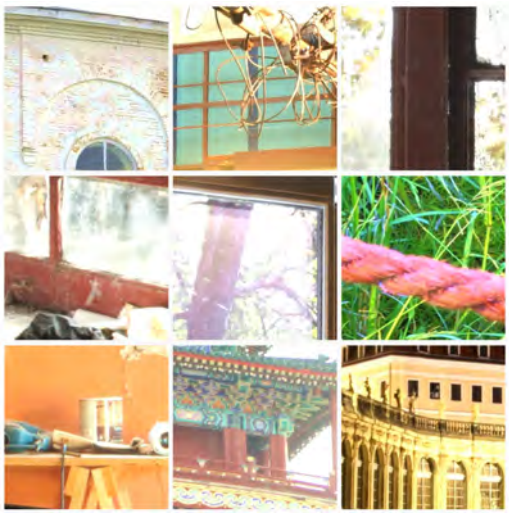

Figure 4: *Samples of selected training data.*

3.1. Dataset

In order to train our neural network, we collected 3948 HDR images on the website according to the source list in [MYK18], then used a virtual camera [EKD*17] and patch cropping [SRK20b] to generate training data. In a picture layout, the principal component usually contains the saturated area in the middle, and the edges contain the background information.

Due to the randomness of cropping, there is no guarantee that each generated training data can contain enough texture. Therefore, we apply a patch sampling algorithm [SRK20b] to detect texture information, this algorithm returns a gradient score for each patch, the higher score stands to include more texture as shown in FIGURE 3, the patch whose score is more than 0.1 will be selected as training data. Finally, 281895 image pairs are selected as training data, example shows in FIGURE 4.

3.2. Architecture

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.

3.3. Loss Functions

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.

Loss functions will be applied to specific areas of the images. We want to find the loss between images which give a substantial amount of intensity information which would be where the HDR and LDR images differ. To achieve this, we obtain the intensity of the image by subtracting the RG Chromaticity (Normalized RGB image) from the image. The RG Chromaticity when used alone provides information about the apparent colors of the scene rather than the intensity information. Therefore, subtracting this from the original image would give us more intensity information to apply our loss function on. The RG Chromaticity is represented by the proportion of the red, green and blue in the color. The difference between the original and the transformed color space to which we would be applying the loss function to is shown in FIGURE 5.

Mean Absolute Error: The first loss function taken into consideration is the Mean Absolute Error (MAE), or the *L1* loss function - a pixel-wise loss function which gives the absolute distance between two corresponding pixels of two images and gives relatively higher quality images. The *L1* loss is given by:

$$\mathrm{MAE} = \frac{\sum_{c,i} \left| \widehat{y_{c,i}} - y_{c,i} \right|}{3 \cdot n} \tag{1}$$

Here, $\widehat{y_{c,i}}$ is the i^{th} pixel of the c^{th} color channel of the predicted image and $y_{c,i}$ is the i^{th} pixel of the c^{th} color channel of the ground truth image.

Multi-Scale Structural Similarity Index: The Multi-Scale SSIM (MS-SSIM) [WSB] was introduced to tackle a dependency that the Structural Similarity Index (SSIM)s metric was bound to - that SSIM was dependent on the scale of the image, and therefore, changes in viewing distance from the image to observer would have

Figure 5: *From Left to Right: Input Low Dynamic Range Image, Intensity Image on which the loss functions are computed, Image Mask after extracting the saturated regions of the image.*

an impact on the score. MS-SSIM, however, downsamples the images through the application of low-pass filters. The equation, in terms of luminance, contrast, and structure is given as:

$$MS{-}SSIM(x,y) = \left[l_{M(x,y)}\right]^{\alpha_M} \cdot \prod_{j=1}^{M} \left[c_j(x,y)\right]^{\beta_j} \left[s_j(x,y)\right]^{\gamma_j} \quad (2)$$

Here, M is the number of scales over which we iterate the function. We chose this particular metric as the perceptual loss because, given its scale invariance, the metric is proven to perform equally or better than the other evaluation criteria.

Finally, we combine both of these concepts to formulate our loss function, which is given by:

$$L(y,\hat{y}) = \lambda_1(MAE) + \lambda_2(1 - MS{-}SSIM) \quad (3)$$

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. The saturated regions are extracted through a mask which we define as *msk*, where

$$msk = \frac{max(0, x_{in} - thr)}{1 - thr} \quad (4)$$

The threshold value *thr* was chosen as 0.95 to give the best saturated area mask. *msk* is involved in the prediction of the final HDR Image through equation 5:

$$y_{\text{out}} = (1 - msk)x_{lin} + msk(y_{predict}) \quad (5)$$

Input linearization is approximately evaluated by the equation: $x_{lin} = x^2$ where x is the input and x_{lin} is the linearized output.

3.4. 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, as shown in 7 to carry out the prediction. Due to the network's downscaling, convolutional and pooling capabilities, the pixel-wise prediction of the downscaled images are affected by its neighbouring pixels, thereby improving coherence among multiple views.

4. Results

To evaluate our algorithm that predicts HDR images for single-view inputs, we make use of the SSIM, PSNR, HDR-VDP [MKRH11] metrics which usually quantify the HDR quality. Additionally, we use the Harmonic_HDR_IQA metric by Rousselot et al. [RDMC19] - a full reference quality metric developed which is tailored to compare HDR content.

4.1. Comparison with state-of-the-art

We compare our algorithm with two other state-of-the-art algorithms for HDR reconstruction using single exposure: HDRCNN [EKD*17] and MaskCNN [SRK20a]. 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. Example of results are show in FIGURE 6. Our algorithm turns out to be more robust in terms of quality metrics and produces a better score using the HDR-VDP metric as compared to the other algorithms. This is illustrated in FIGURE 7 where errors are encoded from red (biggest) to blue (least).

Figure 6: *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. Results are tonemapped for display purpose using [RD05].*

Figure 7: *The input grid image consisting of 16 views, used to evaluate our algorithm. The black box includes the sub-grid passed through the network for prediction.*

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

Table 1: *Scores averaged over 40 images of different exposures and scenes. HDR-VDP, the most commonly used metric, gives a slightly better score for our algorithm.*

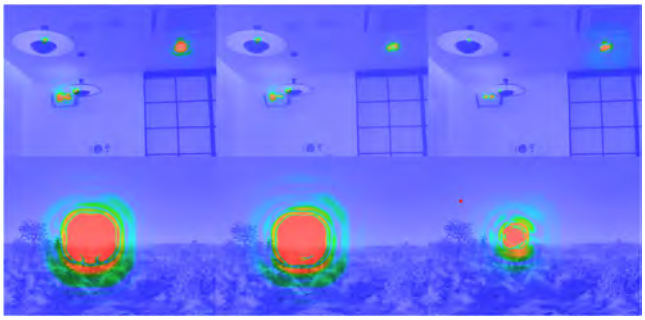

Figure 8: *HDR-VDP visual results when compared to the reference image for images of FIGURE 6. From Left to Right: HDRCNN Output, MaskCNN Output, Our algorithm.*

Our algorithm outperforms the state-of-the-art in single-view HDR Reconstruction when it comes to dynamic range retention. Retention here refers to obtaining dynamic ranges comparable to those of the ground truth HDR Images. The dynamic range of an image is generally calculated by taking the base-2 logarithmic value of the absolute difference between the maximum valued and minimum valued pixels in the image. The results, summarized in Table 2, show 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.

Metric	HDRCNN	MaskCNN	Our
Dynamic Range Error	1.051	1.026	0.921

Table 2: *Average of the absolute errors between the Dynamic Range of the ground truth images and the Dynamic Range of the images predicted through our algorithm.*

4.2. Multi-view Consistency Assessment

We ran our consistency test on a set of real images captured using the ReVeRY camera array at 2 different exposures to build an HDR ground truth image [DM08]. We ran the algorithm on a set of four LDR images as described in SECTION 3.4. Our original training algorithm inputs 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 FIGURE 9 (toned map pictures) and Table 3. 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.

Figure 9: *From top to bottom: LDR Inputs, multi-view outputs of single view inputs, multi-view outputs of grid inputs, ground truth HDR images. All of the above images are operated on views 1,6, 11 and 16 of the camera array.*

Metric	Independent Views	Grid Views
SAD	3831483.42	1388318.70
NCC	0.014	0.22

Table 3: *Results of the consistency evaluation.*

5. Conclusion

We presented a new methodology for a three-stage 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.

5.1. Future work

A new machine-learning-based depth reconstruction approach was proposed in the ReVeRY project [BGPC22]. It is tailored for wide-baseline camera systems, like the ReVeRY camera (each camera is distant of 20cm). The next step is now to adapt this depth-reconstruction to a the HDR multiview input proposed in this paper. This will allow us to study how Barrios et al. [BGPC22] performs on HDR images, but also whether our HDR reconstruction is of sufficient quality for depth reconstruction. An excellent outcome would be that using HDR images as input improves depth reconstruction accuracy.

Another interesting approach would be to use the recovered depth to guide the HDR multiview consistency step. Typically, based on ReVeRY camera configuration, depth values are associated to disparity values which represent the offset in number of pixels between adjacent images like it was done in Bonnard et al. [BLV*12] [BVL19]. One drawback of these previous approaches is that disparity was computed on differently exposed images, which would be overcome with the current approach. However, this would not directly solve for overexposed areas. An additional final step could refine iteratively both depth and HDR values.

6. Acknowledgments

The work was funded by the ANR ReVeRY project (ANR-17-CE23-0020). Facilities of the Centre Image and the computing centre ROMEO were used for this project. We thank our project collaborators for data and fruitful discussions.

References

[BGPC22] BARRIOS T., GERHARS J., PRÉVOST S., CÉLINE L.: Fast and fine disparity reconstruction for wide-baseline camera arrays with deep neural networks. In *Proceedings of the poster session of the annual Eurographics conference* (2022). 7

[BLDC06] BANTERLE F., LEDDA P., DEBATTISTA K., CHALMERS A.: Inverse tone mapping. *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia - GRAPHITE 06* (2006). doi:10.1145/1174429.1174489. 3

[BLV*12] BONNARD J., LOSCOS C., VALETTE G., NOURRIT J.-M., LUCAS L.: High-dynamic range video acquisition with a multiview camera. In *Optics, Photonics, and Digital Technologies for Multimedia Applications II* (2012), Schelkens P., Ebrahimi T., Cristòbal G., Truchetet F., Saarikko P., (Eds.), vol. 8436, International Society for Optics and Photonics, SPIE, pp. 81 – 91. 2, 7

[BVL19] BONNARD J., VALETTE G., LOSCOS C.: Disparity-based hdr imaging. vol. 1905.07918, arXiv. 2, 7

[Deb08] DEBEVEC P.: A median cut algorithm for light probe sampling. *ACM SIGGRAPH 2008 classes on - SIGGRAPH 08* (2008). doi:10.1145/1401132.1401176. 3

[DM08] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. *ACM SIGGRAPH 2008 classes on - SIGGRAPH 08* (2008). doi:10.1145/1401132.1401174. 3, 7

[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. 3, 4, 5

[HZ04] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, 2 ed. Cambridge University Press, 2004. doi:10.1017/CBO9780511811685. 2

[KK19] KINOSHITA Y., KIYA H.: itm-net: Deep inverse tone mapping using novel loss function based on tone mapping operator. *2019 27th European Signal Processing Conference (EUSIPCO)* (2019). doi:10.23919/eusipco.2019.8902744. 3

[LAK18] LEE S., AN G. H., KANG S.-J.: Deep recursive hdri: Inverse tone mapping using generative adversarial networks. *Computer Vision – ECCV 2018 Lecture Notes in Computer Science* (2018), 613–628. doi:10.1007/978-3-030-01216-8_37. 3

[LC09] LIN H.-Y., CHANG W.-Z.: High dynamic range imaging for stereoscopic scene representation. In *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt* (2009), IEEE, pp. 4305–4308. URL: http://dx.doi.org/10.1109/ICIP.2009.5413665, doi:10.1109/ICIP.2009.5413665. 2

[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. doi:10.1007/978-3-642-19309-5_32. 3

[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: Hdr-vdp-2. *ACM SIGGRAPH 2011 papers on - SIGGRAPH 11* (2011). doi:10.1145/1964921.1964935. 5

[MYK18] MORIWAKI K., YOSHIHASHI R., KAWAKAMI R.: Hybrid Loss for Learning Single-Image-based HDR Reconstruction. *arXiv preprint arXiv:1812.07134* (2018). 4

[NPR10a] NIQUIN C., PRÉVOST S., RÉMION Y.: Depth extraction for auto-stereoscopic sets by means of a mesh reconstruction algorithm. In *IST/SPIE Electronic Imaging, Conference EI101 : Stereoscopic Displays and Application XXI* (01 2010). 2

[NPR10b] NIQUIN C., PRÉVOST S., RÉMION Y.: An occlusion approach with consistency constraint for multiscopic depth extraction. *Int. J. Digital Multimedia Broadcasting 2010* (01 2010). doi:10.1155/2010/857160. 2

[OLMA13] OROZCO R. R., LOSCOS C., MARTIN I., ARTUSI A.: Patch-based registration for auto-stereoscopic HDR content creation. *HDRi2013 - First International Conference and SME Workshop on HDR imaging* (2013), 7. 2

[RD05] REINHARD E., DEVLIN K.: Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1 (2005), 13–24. doi:10.1109/TVCG.2005.9. 6

[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). doi:10.1109/icip.2019.8803635. 5

[SMW10] SUN N., MANSOUR H., WARD R.: Hdr image construction from multi-exposed stereo ldr images. In *Proceedings - International Conference on Image Processing, ICIP* (09 2010), pp. 2973–2976. doi:10.1109/ICIP.2010.5653371. 2

[SRK20a] 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. 3, 5

[SRK20b] 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. 4

[TKS06] TROCCOLI A., KANG S. B., SEITZ S.: Multi-view multi-exposure stereo. *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT06)* (2006). doi:10.1109/3dpvt.2006.98. 3

[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. doi:10.1109/ACSSC.2003.1292216. 4

[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. 4