

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

The version of record is available at <https://doi.org/10.1109/IC3D53758.2021.9687182>

PERFORMANCE EVALUATION OF HDR IMAGE RECONSTRUCTION TECHNIQUES ON LIGHT FIELD IMAGES

Mary Guindy^{a, b}, Adhikarla V. Kiran^b, Peter A. Kara^{c, d}, Tibor Balogh^a, Aniko Simon^e

^aHolografika, Budapest, Hungary

Email: {m.guindy, t.balogh}@holografika.com

^bPazmany Peter Catholic University, Budapest, Hungary

Email: {guindy.mary.mohsen.messak, adhikarla.vamsi.kiran}@itk.ppke.hu

^cBudapest University of Technology and Economics, Budapest, Hungary

Email: kara@hit.bme.hu

^dKingston University, London, UK

Email: p.kara@kingston.ac.uk

^eSigma Technology, Budapest, Hungary

Email: aniko.simon@sigmatechnology.se

ABSTRACT

The reconstruction of high dynamic range (HDR) images from low dynamic range (LDR) images is a challenging task. Multiple algorithms are implemented to perform the reconstruction process for HDR images and videos. These techniques include, but are not limited to reverse tone mapping, computational photography and convolutional neural networks (CNNs). From the aforementioned techniques, CNNs have proven to be the most efficient when tested against conventional 2D images and videos. However, at the time of this paper, applying such CNNs to light field contents have not yet been performed. Light field images impose more challenges and difficulties to the proposed CNNs, as there are multiple images for the creation of a single light field scene. In this paper, we test some of the existing HDR CNNs (*ExpandNet*, *HDR-DeepCNN* and *DeepHDRVideo*) on the *Teddy* light field image dataset and evaluate their performance using PSNR, SSIM and HDR-VDP 2.2.1. Our work addresses both image and video reconstruction techniques in the context of light field imaging. The results indicate that further modifications to the state-of-the-art reconstruction techniques are required to efficiently accommodate the spatial coherence in light field images.

Index Terms— HDR image, HDR video, light field imaging, convolutional neural network.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813170. Also received funding by 2018-2.1.3-EUREKA-2018-00007 and KFI 16-1-2017-0015, NRD Fund, Hungary.

1. INTRODUCTION

High dynamic range (HDR) images are created with more luminance levels than conventional low dynamic range (LDR) images. In other words, HDR images represent the bright and the dark regions of the content with more realism [1]. HDR images have proven their importance in numerous applications, including but not limited to global illumination in physical based rendering, digital photography, remote sensing and image editing [2].

HDR imaging can be classified into two main categories: single-camera and multi-camera techniques. For single-camera techniques, capturing is done by means of one camera as the name implies. Sequential capturing is carried out for the exposure stack, which is the main disadvantage of the single-camera technique. This drawback is significant for time-critical tasks. Following the capture process, the inverse camera response function is then reconstructed in order to approximately evaluate the radiance mapping to the scene. Finally, a tone mapper is applied. On the other hand, the exposure stack is captured simultaneously, using the multi-camera technique. Although this technique is efficient in terms of capture time, there still remains the issue of stereo view correspondence since the images resulting from the different cameras have different luminance values [1].

For both methods, however, multiple LDR images are required, from which the HDR image is recovered. Accordingly, for single-camera approaches, a tripod is required to capture the same static scene multiple times with different exposures. Hence, these methods will not work with hand-

held cameras. Neither will they work with dynamic scenes, and therefore, ghosting effects may appear in the final HDR image [3].

Since using multiple cameras to capture a single HDR image is more expensive and requires more calibration, therefore, single-camera techniques appear to be better. Moreover, to overcome the drawback concerning the time limitation previously introduced, as well as fixing the camera position, single-shot capturing can be applied instead of taking multiple shots. However, this comes with an adjustment of multiplexing different exposure patterns on the sensor. In order to reconstruct the HDR image from the LDR single-shot images, many methods have been devised, including reverse tone mapping, computational photography and convolutional neural networks (CNNs). Among those approaches, CNNs have proven to be the best.

In addition to HDR image reconstruction, HDR video reconstruction has also emerged. Since videos are composed of multiple frames, challenges arise in their process of HDR reconstruction. However, due to the temporal coherence between the consecutive frames, common information can be used for reconstructing HDR videos to boost the results compared to single HDR images. In this paper, we address the different techniques used to reconstruct HDR images from LDR images, as well as HDR video reconstruction [4].

Although research on the LDR to HDR single-shot image reconstruction is still ongoing, the feasibility of the implemented techniques are not yet tested for light field (LF) images. The term “light field” was first introduced by Gershun in 1936 [5]. Light fields define the world as light rays filling up the 3D space under representation. In order to display light fields, light field displays (LFDs) were invented, providing spectators 3D experience without the need for additional visual gears. We can classify two main types of LFDs: horizontal-only parallax (HOP) and full-parallax (FP) displays [6]. In order to create the 3D visual experience, multiple light field images are required for a single scene. In this paper, we introduce the concept of LDR to HDR light field image reconstruction. Although no research in this area has been done so far, similarities can be detected with HDR video reconstruction on 2D displays. HDR video reconstruction exploit the temporal coherence between the neighboring frames. This is analogous to LF images where similarities exist between the adjacent views, since the scene is depicted via multiple LF images in order to achieve 3D visualization. As one of the primary scientific contributions of this paper, we test some of the HDR reconstruction CNNs – for images and videos – on an LF dataset and measure the quality of the achieved results against a set of metrics.

The paper is structured as follows: Section 2 discusses the related work on the different CNNs used for image and video HDR reconstruction. Section 3 elaborates the experimental setup for LDR to HDR light field image reconstruction, including the dataset and the metrics required for quality mea-

surements. The achieved results are presented in Section 4. Finally, the work is concluded in Section 5, providing potential future continuations of this work.

2. RELATED WORK

2.1. HDR image reconstruction

Much work has been carried out concerning HDR image reconstruction via CNNs. In this section, we discuss some of those works. For single-shot single-camera approaches, multiplexing the exposure for the sensor is carried out in order to hold more information about the HDR image. However, this is not always the case. The work of Eilertsen *et al.* [7] provides a method to reconstruct HDR images from single-exposure LDRs. However, the approach is to recover information concerning the saturated pixels, and not those pixels in the lower part of the dynamic range. The idea is based on a CNN acting as an autoencoder with a hybrid dynamic range. The primary concept is to convert the input LDR image by means of an LDR encoder into a set of spatial feature representations to be used later by the HDR decoder in log domain, resulting a recovered HDR image. In addition to the encoder-decoder pipeline and the aim to efficiently exploit the high-resolution details of the HDR image, skip connections are available all the way between both the LDR encoder and the HDR decoder.

Other HDR reconstruction techniques are based on reversing the camera pipeline used to create LDR images [8]. The camera pipeline for LDR image formation is composed of three main steps:

1. Clipping of the dynamic range: First, the values of the HDR image created by the camera are clipped to a limited range. This leads to information loss in the over-exposed (i.e., very bright) regions.
2. Non-linear mapping: This step performs non-linear mapping to the camera response function (CRF) in order to adjust the contrast. Thus, matching the human visual system (HVS) of scenes.
3. Quantization: As the RGB channels in an LDR image have 8 bits each, this step ensures that the pixels are quantized to 8 bits. Quantization, however, leads to the loss of information in the under-exposed (i.e., very dark) regions of the image, as well as in the gradient-smooth regions.

Looking at the LDR image formation pipeline, the idea is to create a pipeline for HDR image reconstruction, reversing that of the LDR. Hence, the HDR pipeline is composed of three steps as well, each reversing its corresponding step in the LDR pipeline. For each of these phases, a CNN network with a specific training is built. The three steps composing the HDR image reconstruction pipeline are the following:

1. **Dequantization:** The dequantization process aims to remove the contouring artefacts and the noise generated in the smooth regions due to the quantization process in the LDR image generation pipeline. Training for the dequantization CNN is carried out such that the loss is minimized between the dequantized image (\hat{I}_{deq}) and its respective ground truth image (I_n), where $I_n = F(C(H))$. In other words, I_n is the image resulting from the clipping of the dynamic range, followed by non-linear mapping with CRF. The loss – which is required to be minimized – can be described as follows: $L_{deq} = \|\hat{I}_{deq} - I_n\|_2^2$.
2. **Linearization:** This step aims to convert the LDR non-linear image into a linear radiance via CRF. The CRF is estimated in this step with certain constraints: monotonically increasing function and mapping of the minimum and maximum values of the function to their corresponding minimum and maximum values in the output. These constraints are taken into account when designing the CNN network for the linearization process.
3. **Hallucination:** This step aims to recover the information lost in the over-exposed regions due to the dynamic range clipping step. Training for the CNN hallucination step is done by minimizing the log loss: $L_{hal} = \|\log(\hat{H}) - \log(H)\|_2^2$ where \hat{H} is the image resulting from the hallucination step and H is the HDR ground truth image. Since hallucination is used to limit the artefacts arising from the clipping dynamic range step, therefore, the loss is measured in the log domain since the over-exposed regions have extremely high values that may easily produce errors in the linear domain.

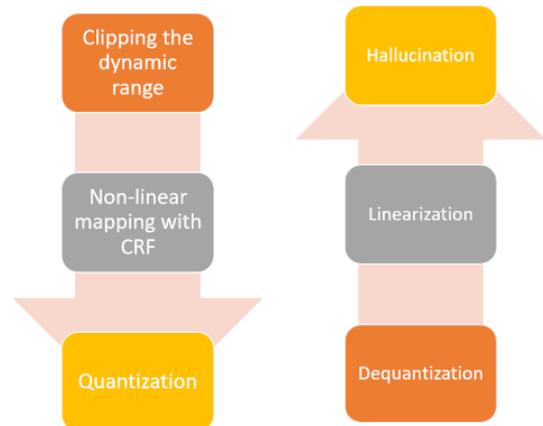
Figure 1 shows both the LDR and the HDR pipelines with elaboration on each step and their corresponding ones in the other pipeline.

Another suggested method for HDR image reconstruction is the *ExpandNet* CNN [9]. This CNN takes an LDR image and propagates it through three branches simultaneously: (i) local branch, (ii) dilation branch and (iii) global branch. Each of the branches handles a respective level of detail (low, medium and high details, respectively). For the first two branches, the LDR image is passed without any sampling – unlike the global branch, where the image is down-sampled. Finally, the outputs from all branches are convoluted, resulting an estimated HDR image.

2.2. HDR video reconstruction

Unlike the previous methods, where the HDR image reconstruction approaches are applied for single images, the works of Kalantari and Ramamoorthi [3] [10] tackle the problem of HDR video reconstruction from multi-exposure frame sequences. Considering videos, HDR image reconstruction is usually carried out in the following two steps:

LDR image formation pipeline



HDR image reconstruction pipeline

Fig. 1: Camera pipelines for LDR and HDR image reconstruction.

1. The first step is to align consecutive frames with various exposures to the current frame. Frames need to be temporally coherent. Therefore, reconstructing frame Z_i is done via its neighboring frames Z_{i-1} and Z_{i+1} . The optical flow method proposed by Liu *et al.* [11] is used for optical flow prediction [3]. This method is carried out by aligning the images with extreme exposures (low and high) to that with medium exposure. A later work proposes a CNN to estimate the optical flow in order to minimize the resulting error between the estimated HDR image and the ground truth image [10].
2. The second step is to fuse the aligned frames for HDR image generation. The proposed CNN [3] [10] is utilized to estimate the fusion weights used in the merging process, hence improving the quality of the resulting images.

Although this HDR video reconstruction method [3] [10] achieves success, ghosting artefacts arise. This is because of the noise and the missing information in the under- and over-exposed regions, respectively. Accordingly, accurate image alignment and fusion is not feasible, leading to ghosting [12]. In order to overcome the aforementioned issue, a coarse-to-fine CNN was proposed for a more accurate image alignment and HDR fusion. The proposed algorithm [12] consists of two main steps:

1. The first step is to align and blend images. This is done via *CoarseNet* CNN. This CNN has the same structure as the CNN of Kalantari and Ramamoorthi [10], as it estimates the optical flow (using the *flow network*) and the blending weights (using the *weight network*). As

the name implies, this CNN results in coarse HDR reconstruction, since it uses a smaller number of feature channels compared to the CNN of Kalantari and Ramamoorthi [10]. Calculation of the loss function used in network training is done via the computation of the tonemapping loss in HDR space, using the μ -law function:

$$T_i^c = \frac{\log(1 + \mu H_i^c)}{\log(1 + \mu)} \quad (1)$$

where μ is the parameter used in controlling the level of compression and it is set to 5000. T_i^c is the HDR image resulting from the tonemapping process. Accordingly, the loss in the *CoarseNet* is calculated against the ground truth HDR image (\tilde{T}_i) as $\|T_i^c - \tilde{T}_i\|_1$. This CNN succeeds at recovering some of the missing information in the over-exposed regions, as well as removing some noise from the under-exposed regions.

- The second step is the alignment and fusion in feature space. This is done via *RefineNet* CNN, which is applied in feature space while performing frame alignment and fusion. *RefineNet* starts by taking as input three coarse HDR images, denoted as $H_{i-1}^c, H_i^c, H_{i+1}^c$ and producing the corresponding 64-channel feature outputs $F_{i-1}^c, F_i^c, F_{i+1}^c$. Deformable convolution [13] is then applied to perform feature alignment, resulting in $\tilde{F}_{i-1}^c, \tilde{F}_i^c, \tilde{F}_{i+1}^c$. These features are then convoluted into the center frame.

Finally, at the end of the pipeline, the reconstruction branch applies regression to the input-fused feature image, resulting in H_i^r , which is used to compute the final estimated HDR image (H_i) as follows:

$$H_i = M_i \odot H_i^c + (1 - M_i) \odot H_i^r \quad (2)$$

where the element-wise product is denoted by \odot and M_i is a mask used to define the well-exposed areas for reference frame i . The following equations show how M_i is defined for low- and high-exposure reference images L_i , respectively.

$$M_i(\text{low}) = \begin{cases} 1, & \text{if } L_i \geq 0.15 \\ (L_i/0.15)^2, & \text{if } L_i < 0.15 \end{cases} \quad (3)$$

$$M_i(\text{high}) = \begin{cases} 1, & \text{if } L_i \leq 0.9 \\ (-\frac{L_i}{0.1} + 10)^2, & \text{if } L_i > 0.9 \end{cases} \quad (4)$$

3. EXPERIMENTAL SETUP

Whereas the different HDR reconstruction CNNs were applied for either conventional images or videos, applying the same CNNs for LF images still remains an open question. In this section, we introduce the experimental setup of the tests that addressed some of the implemented HDR image and video reconstruction CNNs on LF images.

3.1. Dataset

Although LF has been an ongoing research area for some time now, the availability of databases for LF images is sparse. In the work of Gul *et al.* [14], Fraunhofer has managed – by means of a high-quality digital camera – to capture a dataset for LF images in HDR (named *Teddy*), in addition to ensuring large spatial resolution. The dataset was captured by using the exposure bracketing technique, resulting in four static LF images. Accordingly, we used this dataset in our work for LF HDR images reconstruction.

3.2. HDR Reconstruction for LF images

From Section 2, three CNN architectures are used for experimenting with HDR image generation: *ExpandNet* [9], *HDR-DeepCNN* [7] and *DeepHDRVideo* [12]. We have considered real-world HDR LF dataset *Teddy* [14] that contains geometry and color-calibrated HDR LF images. This dataset consists of 50×50 LF images with horizontal and vertical parallax. From the original 50×50 LF images, we generated 36 non-overlapping subsets of LFs, each containing 8×8 images. For each algorithm, the average performance results are reported over all the 36 LF sets.

For testing the HDR image reconstruction algorithms (*ExpandNet* and *HDR-DeepCNN*), we simulated constant-exposure LDR images from the *Teddy* HDR images and fed each network one image at a time. The performance is measured for one image at a time and then averaged over a given LF subset. For testing the *DeepHDRVideo* method, we considered three alternating exposures version of this trained algorithm. Precisely, given a LF subset, we extracted three consecutive HDR images at a time and generated three LDR images with varying exposures. The overview of the procedure is provided in Section 2.2, in accordance with the work of Chen *et al.* [12]. Then, these multiple-exposure LF images are fed to this network, three images at a time, for reconstructing HDR LFs. After generating the corresponding HDR LF subset – similarly to HDR image reconstruction methods – we measure the performance for one image at a time and then average it over a give LF subset.

3.3. Metrics used for evaluation

In order to test the efficiency of the produced results, quantitative analysis is carried out, where the generated HDR LF images are compared against the ground truth images. In this subsection, we discuss the metrics used in the comparison process including:

- PSNR (Peak-Signal-to-Noise-Ratio): PSNR is calculated between two gray-scale images f and g , given that their size is $N \times M$ as follows [15]:

$$PSNR(f, g) = 10 \log_{10}(255^2 / MSE(f, g)) \quad (5)$$

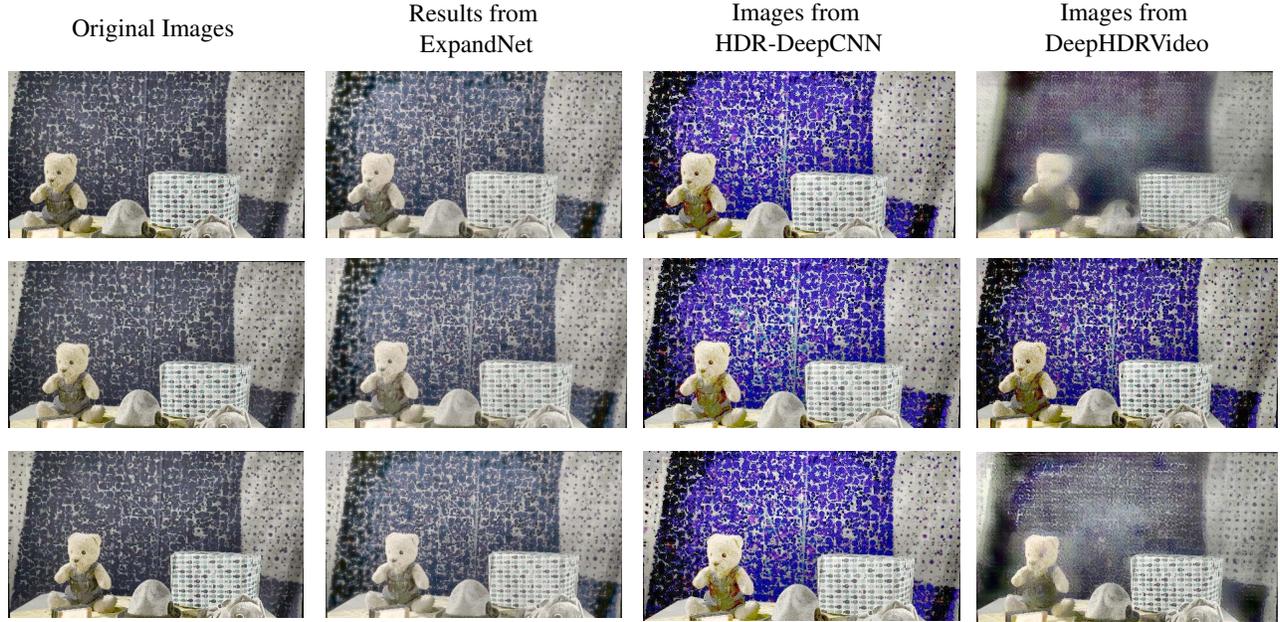


Fig. 2: Original and predicted tonemapped HDR images from the considered algorithms for experiment.

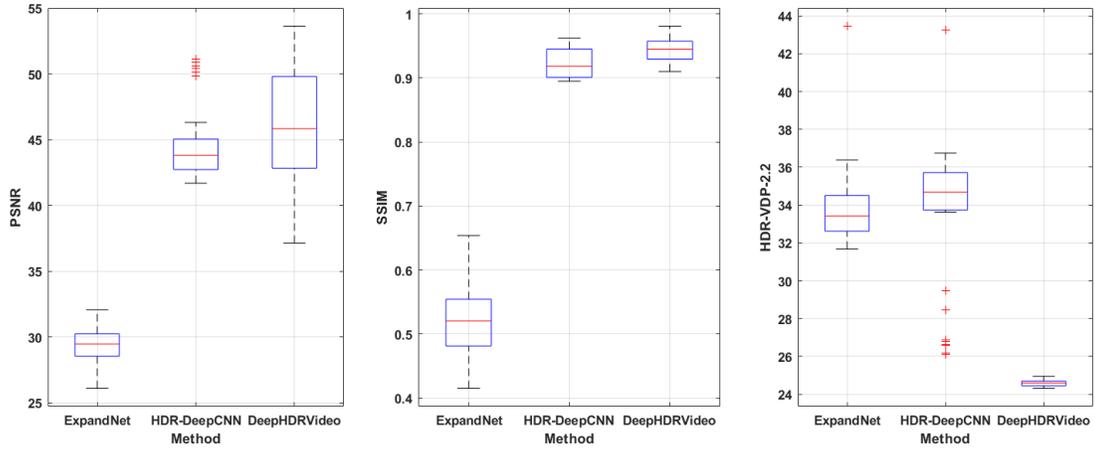


Fig. 3: Boxplots showing the performance of the considered methods for experiment on the HDR light field dataset *Teddy*. Left: box plot using the PSNR metric. Middle: box plot using the SSIM metric. Right: box plot using the HDR-VDP metric.

where

$$MSE(f, g) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2 \quad (6)$$

- SSIM (Structured Similarity Index Measure):

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \quad (7)$$

where

$$\begin{aligned} l(f, g) &= \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1}, \\ c(f, g) &= \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2}, \\ s(f, g) &= \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3} \end{aligned} \quad (8)$$

where l , c and s are the luminance, contrast and structure comparison functions, respectively. The terms μ_f and μ_g are the luminance means for images f and g . The standard deviation is denoted by σ , whereas σ_{fg} is the correlation coefficient between both images (f and g). Finally, C_1 , C_2 and C_3 are positive constants used to ensure that the denominator is not null.

- **HDR-VDP-2.2.1 (High Dynamic Range Visible Difference Predictor):** This metric is an upgrade for the HDR-VDP metric, which in turn is a modification to the VDP metric. It takes two images as input: the original and the distorted one. Both input images are transformed to their luminance values used in the comparison process. The metric results in a probability map, indicating the differences between the input images [16]. The HDR-VDP-2 [17] enhances the visibility metric, specially for low conditions with regards to luminance.

4. RESULTS

Figure 2 shows the results of our experiments on the considered HDR LF dataset *Teddy*.

According to Figure 2, *ExpandNet* produces plausible results close to the ground truth images. However, visually ghosting artifacts arise in the background on the images. Hence, learning concatenated features from different branches – local, dilation and global – seems to be a good direction for generalizing to other datasets. On the other hand, in the *HDR-DeepCNN*, we see inconsistencies in the colours in the reconstructed textures. These inconsistencies could be the result of using skip connections that include domain transformation from LDR display values to logarithmic HDR. However, the method should be thoroughly investigated further with more LF datasets. Finally, in the *DeepHDRVideo* reconstruction method, inconsistencies are quite visible in the shape and texture of the reconstructed images. The method uses sub network that aligns the input images to a common frame and then does HDR reconstruction using the aligned features. The artifacts seen in the reconstructed images show that there are errors in the optical flow reconstruction due to complex patterns in the scene which are carried forward to the next stage where HDR reconstruction is done. Without a reliable method for calculating accurate optical flow, such a direction for HDR reconstruction does not seem valid for LF images.

As mentioned earlier, objective performance of the considered methods is measured using three traditional IQMs (Image Quality Metrics): PSNR, SSIM and a perceptually-guided HDR image quality metric HDR-VDP-2.2.1. Results are interpreted by using box plots (see Figure 3), which show the min and max scores together with the inter-quartile range (shown in boxes). A larger inter-quartile range translates to more global inconsistencies in the performance.

Our experimental results show that according to metrics PSNR and SSIM, *DeepHDRVideo* method performs better than other two. It is important to note that the inter-quartile range in obtained PSNR values for the *DeepHDRVideo* method is higher. Also the difference between observed minimum and maximum PSNR values for this method is higher than other two methods showing that there are greater deviations in the performance of this method. In contrast, the SSIM results show no such fluctuations in the performance of this method. HDR-VDP scores show that *HDR-DeepCNN* performs better.

Assuming that HDR-VDP better correlates to the HVS than the other two metrics, our experiments show that the *HDR-DeepCNN* method achieves more consistent results than the other two methods, and also achieves better visual quality of the reconstructed HDR images. Although, HDR video reconstruction methods involve retrieving more scene information than HDR image reconstruction methods, our results indicate that given a properly color-calibrated set of single-exposure LDR LF images of a scene – such as scene *Teddy* – architectures like *HDR-DeepCNN* are capable of reconstructing more globally consistent HDR LFs. This direction is particularly desirable since it only involves capturing single-exposure LDR images, and therefore, also supports faster processing times for HDR reconstruction than methods involving multiple exposures. To further develop such architectures, there is a demand for producing more LF HDR datasets for fine tuning, which is currently lacking in the scientific literature. Furthermore, irregularities in the reported quality scores of various metrics show that there is a great need for novel quality metrics that are more suitable for LFs.

5. CONCLUSION AND FUTURE WORK

In conclusion, we have elaborated the different CNNs used for LDR to HDR image and video reconstruction. Although much work has already been done in this area, the application of these algorithms to LF images has not yet been investigated. In this paper, we have tested different HDR CNNs: (i) *ExpandNet*, (ii) *HDR-DeepCNN* and (iii) *DeepHDRVideo* for LF images. The first two networks are used for images, whereas the third one is used for videos. Results were evaluated with various metrics. While expectations were in favor of the video reconstruction techniques due to exploiting the temporal coherence – which is in a way similar to the spatial coherence in the LF images – results for *HDR-DeepCNN* proved to be more plausible in terms of quality. Moreover, it had the best value when evaluated against HDR-VDP.

In the future, we plan to generate a synthetic LF HDR dataset and train/fine-tune the *HDR-DeepCNN* architecture for experimenting HDR LF reconstruction. We also plan to explore systems and methods to capture real-world HDR LF content to further support such experiments.

6. REFERENCES

- [1] P.-J. Wu, K.-T. Shih, and H. Chen, "Dual-camera HDR synthesis guided by long-exposure image," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2016, pp. 1–4.
- [2] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.
- [3] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, vol. 36, no. 4, 2017.
- [4] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1375–1385.
- [5] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1-4, pp. 51–151, 1939. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm193918151>
- [6] N. Balram and I. Tošić, "Light-field imaging and display systems," *Information Display*, vol. 32, no. 4, pp. 6–13, 2016.
- [7] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM transactions on graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.
- [8] Y.-L. Liu, W.-S. Lai, Y.-S. Chen, Y.-L. Kao, M.-H. Yang, Y.-Y. Chuang, and J.-B. Huang, "Single-image HDR reconstruction by learning to reverse the camera pipeline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1651–1660.
- [9] D. Marnierides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," in *Computer Graphics Forum*, vol. 37, no. 2. Wiley Online Library, 2018, pp. 37–49.
- [10] N. K. Kalantari and R. Ramamoorthi, "Deep HDR video from sequences with alternating exposures," in *Computer Graphics Forum*, vol. 38, no. 2. Wiley Online Library, 2019, pp. 193–205.
- [11] C. Liu *et al.*, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- [12] G. Chen, C. Chen, S. Guo, Z. Liang, K.-Y. K. Wong, and L. Zhang, "Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark dataset," 2021.
- [13] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 764–773.
- [14] M. S. K. Gul, T. Wolf, M. Bätz, M. Ziegler, and J. Keinert, "A high-resolution high dynamic range light-field dataset with an application to view synthesis and tone-mapping," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2020, pp. 1–6.
- [15] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.
- [16] R. Mantiuk, S. J. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images: model and its calibration," in *Human Vision and Electronic Imaging X*, vol. 5666. International Society for Optics and Photonics, 2005, pp. 204–214.
- [17] 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 Transactions on graphics (TOG)*, vol. 30, no. 4, pp. 1–14, 2011.