Towards reconstructing HDR light fields by combining 2D and 3D CNN architectures

Mary Guindy[a,b], Adhikarla V. Kiran[b], Peter A. Kara[c,d], Tibor Balogh[a], and Aniko Simon[e]

[a]Holografika, Budapest, Hungary
[b]Pazmany Peter Catholic University, Budapest, Hungary
[c]Budapest University of Technology and Economics, Budapest, Hungary
[d]Kingston University, London, UK
[e]Sigma Technology, Budapest, Hungary

ABSTRACT

High dynamic range imaging has become a technological trend in the past couple of decades, particularly through its integration into many applications. Numerous attempts were made to reconstruct HDR images from low-dynamic-range data. Such reconstruction techniques can be classified into single-camera and multi-camera approaches. Single-camera setups are less expensive, yet multi-camera setups are more efficient. At the time of this paper, there is already a great number of algorithms for single-camera HDR image reconstruction, but there are only a few for HDR video reconstruction. The latter takes into account the temporal coherence between consecutive video frames, leading to better results. For light field images, this remains a challenging open issue, as the HDR video reconstruction methods do not work as efficiently for light field images as HDR image reconstruction algorithms do. However, analogously to 2D videos, where consecutive frames have temporal coherence, many similarities can be found between the adjacent views of light field contents. In this paper, we investigate the theoretical possibilities of combining CNN architectures utilized for HDR images and videos, in order to enhance the outputs of HDR light field image reconstruction. The concept of our work is to exploit the similarities between light field images since they all visualize the same scene from different angular perspectives.

Keywords: CNN architecture, light field imaging, light field reconstruction, HDR

1. INTRODUCTION

Over the past century, the practical implementations of the concept of light field (LF) have developed and become recently popular in both digital and optical technologies, through which the human visual system (HVS) is better understood. The physical world is represented by LFs via light rays occupying the 3D space under representation.[1] As a means of representing LFs, the plenoptic function was firstly introduced by Adelson and Bergen in 1991.[2] The plenoptic function is a 7D function with 7 parameters, including the position of the eye (V_x, V_y, V_z), and the orientation (θ, ϕ) and wavelength of the incoming light (λ) captured at a given time (t). Although this function provides a full representation of LFs, it has a high complexity and an expensive computational load due to its high dimensionality. Accordingly, the function was furtherly reduced to a 4D representation in free spaces by means of a two-plane parametrization. In other words, two sets of coordinates are used to represent the light ray captured by the camera, where u, v denote the intersection with the camera lens plane and s, t denote the intersection with the sensor plane.[3] As a means of LF visualization, light field displays (LFDs) have been developed to act as a window to the visualized 3D world without the additional need of 3D viewing gears. Hence, LF rendering needs to occur in accordance with LFD properties. In order to

---

Further author information: (Send correspondence to Mary Guindy)
Mary Guindy: E-mail: m.guindy@holografika.com / guindy.mary.mohsen.messak@ppke.hu
Adhikarla V. Kiran: E-mail: adhikarla.vamsi.kiran@itk.ppke.hu
Peter A. Kara: E-mail: kara@hit.bme.hu
Tibor Balogh: E-mail: t.balogh@holografika.com
Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

generate this content most efficiently, gathering multiple images to create the corresponding LFs is performed.[3] In other words, multiple images are needed to represent and render one scene on an LFD.

In addition to LFs, another emerging technology is high dynamic range (HDR) imaging. The term "dynamic range" defines the ratio between the brightest and the darkest pixels. Accordingly, HDR denotes a high ratio value, hence, representing a wider range of colors and brightness compared to low dynamic range (LDR) images. HDR images are also called "radiance maps" due to their ability to store the scene representation in the form of a range of intensities that corresponds to the actual scene[4].[5] Thus, HDR images have higher fidelity and realism compared to LDR images. As a consequence, HDR is useful in many applications, including – but not limited to – image editing, cinematography, virtual reality and computer graphics.

Displaying HDR content on LFDs can be rather powerful, since HDR provides more sense of realism, in addition to LFDs adding a sense of immersion to spectators. However, reconstructing HDR content from LDR LF images can be challenging, and at the same time, generates better outputs since the information of the scene is encoded in multiple images.

In this paper, we discuss the possible solutions to the problem of HDR LF images reconstruction. The remainder of the paper is structured as follows: Section 2 discusses the previous work done in the investigated area. In Section 3, we present different techniques and ideas that could be used in the reconstruction process. Finally, the work is concluded in Section 4, with possible suggestions for future work.

2. RELATED WORK

2.1 HDR image reconstruction

LDR-to-HDR image reconstruction is a long, on-going research that has developed over the past years. Many techniques have been devised in the field of HDR image reconstruction, including modifying the hardware components of the capture systems or implementing algorithms used in the reconstruction process. Very few attempts though were targeted for the HDR reconstruction of LF images.

2.1.1 Hardware approach for HDR image reconstruction

Hardware approaches aim at reconstructing the HDR images by modifying the hardware of the capture systems. Hence, these setups are usually expensive, as they require specific, custom-made sensors or optical systems to be installed in cameras. Single-shot techniques include modifying the camera hardware to capture the image while splitting the incoming light to multiple sensors via a beam-splitter, usage of coded per-pixel exposure or modulus images. Addressing the same HDR reconstruction problem in LFs can be challenging due to their angularly selective nature. Considering narrow-baseline LFs, few attempts have been made in the HDR reconstruction process. To begin with, Lumsdaine and Gerogiev[6] proposed a focused plenoptic camera that uses a lenset array. This camera has the ability to produce HDR LF images with acceptable parallax and higher spatial resolution, while reducing the angular density. As an extension to their work, Georgiev *et al.*[7] used their focused plenoptic camera for rich image capture. To achieve their goal, two alternative upgrades were introduced to the focused plenoptic cameras. This is done by interleaving the camera filters at the main camera lens or at the microlens array in order to multiplex the captured plenoptic function.

Afterwards, Wang *et al.*[8] designed a hybrid system for HDR LF capture with two cameras: (i) an LF camera and (ii) a single-lens digital reflex camera with high resolution. These cameras have a beam splitter while sharing the same optical path. While the usage of a single plenoptic camera can produce LF images, acquiring more plenoptic cameras with different exposures will result in HDR LF images. Two attempts were made in this area. The first one was suggested by the work of Li *et al.*,[9] where the merging process of the different LF images was done using the method suggested by Debevec and Malik[10] on 4D LF images. Although this method has plausible results, two main drawbacks arise: (i) additional computational load is required since large data is needed for this method and (ii) the method has some requirements that should be taken into account during the capture process, as the images captured with multiple exposures need to have some overlaps in the well-exposed areas. This furtherly complicates the acquisition procedure. As a second attempt to create HDR LFs by means of multiple plenoptic cameras with different exposures, Le *et al.*[11] proposed using the RAW data to achieve soft detection for saturated pixels, which will be used later in the HDR reconstruction process. The multiple sub-aperture images captured at different exposures are then organized in a matrix, that is later completed using the Weighted Low Rank Approximation (WLRA) to make use of the redundancies between the different views for better reconstruction.

Although, these few attempts work on the HDR LF capture system, they modify the hardware setups for the narrow-baseline plenoptic cameras, resulting in narrow-baseline LFs. Considering the wide-baseline setups, the HDR reconstruction process can be more challenging, as the camera setup consists of a one- or two-dimensional array of cameras, whereas the display systems are relatively big in size. Thus, for wide-baseline capture systems, applying the same methods is not feasible. Moreover, if a modification is needed for one camera to allow for higher resolution capture, the same modification must be applied for all the cameras in the 2D array setup, hence it is computationally expensive and and exceptionally difficult to perform.

2.1.2 Software approach for HDR image reconstruction

Unlike the previous attempts of modifying the hardware of the LF capture systems to allow HDR reconstruction, an alternative would be to apply HDR reconstruction algorithms to the already existing LDR LF images. Considering conventional HDR images, the reconstruction techniques can be categorized into one of the following: (i) reverse tone mapping, (ii) computational photography and (iii) convolutional neural networks (CNNs).[12] Among those techniques, CNNs proved to be the most efficient. Considering LF imaging, applying the same CNNs used for conventional images can result in rather suboptimal outputs for LF visualization. In our previous work,[13] we applied different CNNs to the LF images to test their efficiency in the HDR reconstruction process. Although it was expected that the video reconstruction CNNs would have better results due to the exploitation of the temporal coherence between the video frames – which somewhat resembles the spatial coherence between the different LF images representing a scene – this was not the case. As for the image-based CNNs, plausible results were obtained for LF images but could be furtherly improved. Accordingly, LF images impose more challenges in the HDR reconstruction process.

2.2 Datasets

Considering conventional HDR reconstruction, many datasets are available that can be used further in training and testing to improve the design of the CNNs. However, when working with HDR LF images, one of the main challenges is the lack of the availability of datasets for training and testing. At the time of writing this paper, and to the best knowledge of the authors, only a single work has been done to create (i.e., capture) an HDR LF dataset, where a high-quality digital camera fixed on two linear axes is used. Moreover, as a means of capture per every viewpoint, exposure bracketing is used. The result is a dataset composed of six static LFs.[14]

3. HDR LF RECONSTRUCTION

As previously stated, the LDR-to-HDR reconstruction process for LF images is more challenging compared to the conventional ones. This is due to the fact that the process is applied to multiple LF images with different angular information. However, since spatial information is shared across the different LF images constructing, constituting an LF scene, better HDR reconstruction results can be achieved by exploiting the shared information.

Let us consider the simple case of a scene with an object placed in front of an open window, allowing the rays of the sunlight to pass through. For conventional LDR imaging, the object will appear almost black. However, having LF images construct the scene, the middle image may have a black object but the other side images – with different angular perspectives – will contain more information about the object color and details. Hence, applying an HDR CNN to multiple LF images simultaneously can better enhance the results.

In the remainder of this section, we discuss in detail the different approaches that could be used in the HDR LF reconstruction process.

3.1 Single-image-based HDR LF image reconstruction

Single-image-based HDR LF image reconstruction denotes a one-to-one relation between the input LDR and the output HDR image, where an HDR reconstruction CNN is applied to each LF image independently. Hence, the CNN is applied sequentially to the LF images, representing the scene. This can be achieved by applying one of the CNNs used for conventional HDR image reconstruction.

Considering LF images, applying the CNNs for each image individually can produce acceptable results at the expense of time, since the time required for reconstructing an HDR LF scene is equivalent to the time required for a single LF image multiplied by the number of LF images representing the scene. In addition to the computational cost, the reconstructed images will not provide the best results due to the misutilization of the shared information across the different images, since this method does not take into account the spatial similarity between the neighboring images, that could result in better reconstruction. Therefore, deploying single-image-based methods for LF images does not have any added value compared to the conventional ones.

3.2 Collective-image-based HDR LF images reconstruction

As previously stated in Section 2, modifications are inhibited in the hardware of the plenoptic cameras in order to capture narrow-baseline HDR LF images. Although this technique is likely to work, it has some disadvantages, among which is the cost of the used hardware. Moreover, due to the aforementioned reasons, these modifications cannot be applied to wide-baseline capture systems.

An alternative is to apply an HDR reconstruction CNN to multiple LF images at once, similar to the idea of applying video CNNs on multiple frames. Accordingly, the reconstruction process is carried out simultaneously while the information is being shared across the different images. Although this technique could be useful in the HDR reconstruction process for LF images by means of sharing the spatial information across them, the angular information could be somewhat tricky, as it differs slightly from one image to the other. On the other hand, the difference in the angular perspective can be used to encode information about the scene that could not be interpreted otherwise, as in the case of specular lighting. Hence, using CNNs for LF image reconstruction is challenging and requires much work.

Due to the aforementioned reasons, new CNNs for LF HDR image reconstruction need to be designed, or at least some modifications to the already existing CNNs need to be applied in order to better fit for LF images. In this portion of the section, we discuss some of the key approaches that could be used in a CNN to better reconstruct HDR LF images. These include the following:

- Depth map: Unlike the problematic case of generating depth maps for single images, depth maps for LF images are easily generated due to the parallax effect (horizontal-only parallax or full parallax). Since depth maps convey information about the scene, they can be integrated in the CNNs for better LF image reconstruction. Although wide-baseline systems impose more challenges compared to the narrow-baseline ones, they provide better accuracy in the reconstruction process of the captured LFs on the LFD, as the reconstruction accuracy is directly proportional to the baseline between the viewpoints.[15] Moreover, better depth estimation occurs with wider baseline setups.[16]
- Keypoint detection: This is done by means of feature descriptors. Using the keypoints across the different LF images can better enhance the HDR reconstruction process. Although many techniques have been devised over the years to achieve this – including SIFT, SURF, BRIEF and others – they could generate better results by adapting them to LFs. Several attempts have been done in this area. The work of Tovsic *et al.*[17] proposes a method for keypoint extraction by means of scale-depth space analysis. First, a scale-space for LF images is computed using Ray-Gaussian kernel. Then, for each layer in the computed scale-space, an edge detector is applied, as it is assumed in the paper that edges are more important features than blobs when it comes to LFs. Finally, a depth value is assigned to each detected keypoint, which is the reason why this method effectively detects the 3D keypoints. Another solution was suggested by Ghasemi *et al.*[18] to implement a scale-invariant feature descriptor, specifically for LFs. This is achieved by taking into account that the Epipolar Planar Image (EPI) lines are proportional to the scene's depth. In addition to the proposed methods, Dansereau *et al.* introduced LiFF (Light Field Features in Scale and Depth).[19]

	Single-image-based CNNs	Collective-image-based CNNs
Relation between input and output	One-to-one	Many-to-many
Training time	Less	More
Testing time	More	Less
Mode of application	Sequential	Simultaneous
Advantages	Simple techniques used	Spatial information shared
Disadvantages	Ignore similarities between images	More complex algorithms and the lack of datasets
Results	Worse	Better

Table 1. Differences between the CNNs used for HDR LF image reconstruction.

LiFF is an extension to SIFT (Scale-Invariant Feature Transform) that is applicable to 4D LFs, where it acts as a feature detector/ descriptor. As in SIFT, LiFF is a scale-invariant solution. Additionally, it detects features in 4D LFs, along with their estimated depth, while being robust to perspective changes.

- Global illumination: Compared to the other major points that can further enhance HDR LF reconstruction, global illumination is considered to be a more advanced and complex concept. Global illumination – also known as indirect illumination – aims at generating more realistic lights in 3D scenes by means of specific methods. Among the proposed methods is the "Light field probes" suggested by McGuire *et al.*[20] Light field probes overcome the issues arising in image-based lighting techniques by encoding further information describing the geometry of the scene. In order to illustrate the idea of light field probes, let us consider it as an emitting sphere of light that can be placed anywhere in a 3D scene.[21] Incorporating the concept of light field probes in HDR LF reconstruction CNNs can give more information about the illumination in the scene and thus further help in the reconstruction process.

The differences between the single-image-based and collective-image-based HDR LF CNNs are illustrated in Table 1.

4. CONCLUSION AND FUTURE WORK

In this paper, we presented different concepts for using CNNs on LF images. Although, single-image-based CNNs can provide plausible results, better outputs are achieved by applying CNNs simultaneously on multiple LF images where the spatial coherence is exploited, in addition to the angular information being taken into account. We discussed some of the key ideas and approaches that could be used in collective-image-based CNNs to achieve better outputs.

There are certainly many possibilities for future work, including the creation of more datasets for HDR LF images to be used in training and testing. Moreover, testing different existing CNNs on various LF images and applying the key concepts for better reconstruction, as well as designing new CNNs that can be used solely for LF images may contribute to better HDR LF reconstruction.

ACKNOWLEDGMENTS

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, NRDI Fund, Hungary. The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

REFERENCES

[1] Balram, N. and Tošić, I., "Light-field imaging and display systems," *Information Display* **32**(4), 6–13 (2016).

[2] Bergen, J. R. and Adelson, E. H., "The plenoptic function and the elements of early vision," *Computational models of visual processing* **1**, 8 (1991).

[3] Levoy, M. and Hanrahan, P., "Light field rendering," in [*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*], 31–42 (1996).

[4] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K., [*High dynamic range imaging: acquisition, display, and image-based lighting*], Morgan Kaufmann (2010).

[5] Mantiuk, R., Krawczyk, G., Zdrojewska, D., Mantiuk, R., Myszkowski, K., and Seidel, H.-P., [*High dynamic range imaging*], na (2015).

[6] Lumsdaine, A. and Georgiev, T., "The focused plenoptic camera," in [*2009 IEEE International Conference on Computational Photography (ICCP)*], 1–8, IEEE (2009).

[7] Georgiev, T., Lumsdaine, A., and Chunev, G., "Using focused plenoptic cameras for rich image capture," *IEEE Computer Graphics and Applications* **31**(1), 62–73 (2010).

[8] Wang, X., Li, L., and Hou, G., "High-resolution light field reconstruction using a hybrid imaging system," *Applied optics* **55**(10), 2580–2593 (2016).

[9] Li, C. and Zhang, X., "High dynamic range and all-focus image from light field," in [*2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*], 7–12, IEEE (2015).

[10] Debevec, P. E. and Malik, J., "Recovering high dynamic range radiance maps from photographs," in [*ACM SIGGRAPH 2008 classes*], 1–10 (2008).

[11] Le Pendu, M., Guillemot, C., and Smolic, A., "High dynamic range light fields via weighted low rank approximation," in [*2018 25th IEEE International Conference on Image Processing (ICIP)*], 1728–1732, IEEE (2018).

[12] Metzler, C. A., Ikoma, H., Peng, Y., and Wetzstein, G., "Deep optics for single-shot high-dynamic-range imaging," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 1375–1385 (2020).

[13] Guindy, M., Kiran, A. V., Kara, P. A., Balogh, T., and Simon, A., "Performance evaluation of HDR image reconstruction techniques on light field images," in [*2021 International Conference on 3D Immersion (IC3D)*], 1–7, IEEE (2021).

[14] Gul, M. S. K., Wolf, T., Bätz, M., Ziegler, M., and Keinert, J., "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)*], 1–6, IEEE (2020).

[15] Leistner, T., Schilling, H., Mackowiak, R., Gumhold, S., and Rother, C., "Learning to think outside the box: Wide-baseline light field depth estimation with epi-shift," in [*2019 International Conference on 3D Vision (3DV)*], 249–257, IEEE (2019).

[16] Alam, M. Z. and Gunturk, B. K., "Hybrid light field imaging for improved spatial resolution and depth range," *Machine Vision and Applications* **29**(1), 11–22 (2018).

[17] Tošić, I. and Berkner, K., "3D keypoint detection by light field scale-depth space analysis," in [*2014 IEEE International Conference on Image Processing (ICIP)*], 1927–1931, IEEE (2014).

[18] Ghasemi, A. and Vetterli, M., "Scale-invariant representation of light field images for object recognition and tracking," in [*Computational Imaging XII*], **9020**, 291–298, SPIE (2014).

[19] Dansereau, D. G., Girod, B., and Wetzstein, G., "LiFF: Light field features in scale and depth," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 8042–8051 (2019).

[20] McGuire, M., Mara, M., Nowrouzezahrai, D., and Luebke, D., "Real-time global illumination using precomputed light field probes," in [*Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*], 1–11 (2017).

[21] Antinozzi, S. A., "An overview of modern global illumination," *Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal* **4**(2), 1 (2017).