A Versatile 5D Light Field Capturing Array

Kelvin Chelli, Tobias Lange and Thorsten Herfet
Telecommunications Lab,
Saarland Informatics Campus,
D-66123 Saarbrücken,Germany
Email: {chelli, lange, herfet}@cs.uni-saarland.de

Martin Alain and Aljosa Smolic
V-SENSE Project,
School of Computer Science and Statistics,
Trinity College Dublin, Ireland
Email: {alainm, smolica}@tcd.ie

Marek Solony and Pavel Smrz
Faculty of Information Technology,
Brno University of Technology, Czech Republic
Email: {isolony, smrz}@fit.vutbr.cz

Jonas Trottnow and Volker Helzle
Filmakademie Baden-Württemberg,
Ludwigsburg, Germany
Email: {jonas.trottnow, volker.helzle}@filmakademie.de

***Abstract*—In this paper, we describe a versatile light field capturing device able to generate sparse 4D LF, LF Video and 5D LF images. The capturing array has proven to be an ubiquitous tool for the experimental generation of light fields and the development of post processing algorithms to provide so called LF assets that can be re-used and re-purposed in creative environments.**

***Index Terms*—5D light fields, light field assets, light field capturing array.**

I. Introduction

The space around us is filled by a dense array of light rays with different intensities. These were termed as radiant pyramids by Leonardo Da Vinci [1]. Although, the basic concept of a light field is quite old, novel usage scenarios for light fields in content generation as well as consumption along with the necessary hardware and software tools enabling light field processing, have evoked a wide spread interest and is considered to be the next disruptive technology in visual computing. In its essence, a light field is nothing but all the light that goes through a volume of space. The 4D set of rays that constitute a light field is specified by its position as well as its direction. Thus, each light ray is usually parameterized by its intersection with two planes, each with a 2D coordinate system, that constitutes a 4D light field [2] which is denoted in this paper as $L(b, a, y, x)$. The (b, a) and the (y, x) planes denote the camera and focal plane respectively. We also denote a 5D light field[1] as $L(t, b, a, y, x)$. It introduces a temporal dimension to the 4D light field that enables us to capture details in scenes that are otherwise not seen. Capturing light fields is challenging due to the sheer amount of data that is generated, the very precise calibration and rectification required, the various processing algorithms that are needed and the requirement for a flexible architecture that provides deterministic behavior in a studio setting. Moreover, the costs involved must not be prohibitive.

[1]not to be confused with the 5D plenoptic function whose radiance is constant along a line in free space

Besides the existing challenges there have been many attempts at creating LF capture devices. It started with light field images from static scenes using capture gantries and evolved quickly to single sensor cameras with an additional micro-lens array (i.e. Lytro Illum) which accelerated the capturing process immensely. In recent years, cameras capable of capturing light field video appeared. One of the first was the Lytro Immerge 2.0, an array with 95 'normal' cameras, but it disappeared quickly after the company was acquired by Google in 2018. Other examples include Raytrix, a German company manufacturing micro-lens based cameras for fast light field video capture and K-Lens, a company creating lens systems with mirrors which can convert most conventional photo or video cameras into light field capturing devices without permanent hardware modifications. One feature that all those approaches are missing is the full control of the time plane during capturing. They allow to set the frame rate of the sequence, but every subframe of the light field is captured at the exact same time. In later sections we introduce a system with better time plane control and the benefits of being able to capture certain subframes at a different time then the rest.

In this paper, a versatile camera array that is able to be configured to capture, 4D LF, 5D LF and LF video is introduced. The capturing pipeline from the pre-configuration of the camera array all the way up to the creation of an light field asset will be described. The proposed light field capturing array has been an ubiquitous tool in the generation of unique light field assets.

II. The Camera Array

A. Hardware

Camera array rigs allow to capture sparse light fields that have several advantages compared to dense light fields like increased spatial resolution, aperture and flexibility as discussed in [3]. An initial impression of our camera array is seen in figure 1.

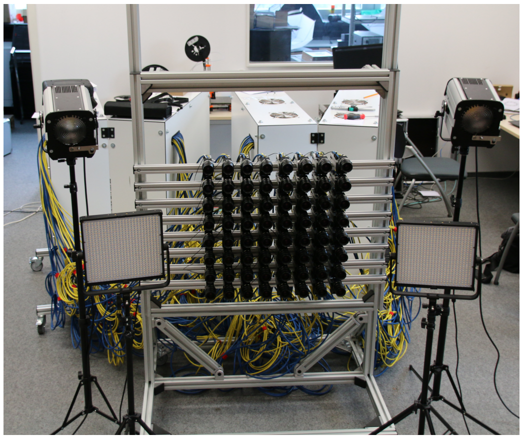

Figure 1. The 5D light field camera array

Our light field capturing array consists of 64 full-HD cameras (1920×1200@40 fps) separated into 4 array modules as shown in figure 2. Each camera is connected to a computing unit which is responsible for caching the captured frames and for most of the pre- and post-processing steps. The camera along with its computing unit forms a camera node. The camera array with a frame rate of 40 fps generates 8.85 GB of raw sensor data per second. In order to store such vast amounts of data, each camera node is networked to the storage infrastructure and the control server via a central switch. The storage infrastructure is realized by a CEPH[2] storage cluster that consists of 4 storage nodes each with a raw capacity of 96 TB. Thus, with replication of the raw data and separate storage of the processed frames, a total of about 115 TB of storage can be provided which is more than adequate to exhibit the capability of the camera array as well as capture various light field content in a production setup. In addition to this, a sync plane has been implemented ensuring that the cameras are gen-locked and run with the same frame-rate. The phase is controlled for every camera in the array individually. This provides the ability to subdivide the frame into a variable number of sub-frames. The requirements for efficient sub-framing patterns are described in [3]. The result is a camera array that is able to capture a wide variety of light fields that can be turned into interesting assets. The architecture is modular and enables the removal or addition of cameras with a minimal overhead. Moreover, the structure on which the cameras are mounted is flexible to rearrange the cameras in a plane so that it can be configured for different camera baselines and layouts, depending on the requirements of the scene.

B. Capturing Pipeline

Our capturing pipeline starts in the camera nodes where the raw images from the cameras are received and stored in a separate SSD cache for every camera. Once the scene is over, the captured frames are transferred to the central storage cluster until further processing is required. The first step of the usual processing chain is the demosaicing to produce full color images. At the moment an AHD-based algorithm is used, but we are currently working on a custom learning-based solution. The second step is a colour alignment step to correct for slight differences in sensor sensitivity and settings, developed by Grogan et al[4]. The last processing step is required because the mechanical alignment of the cameras in the array is never perfect. A camera calibration similar to the algorithm by Xu et al[5], implemented using SLAM++ optimizer [6] calculates the optimal geometric transformation for every camera to align their optical centers to a uniform grid with perfect alignment and minimal loss of image information. The final result is uploaded back to the central storage for future use and distribution.

[2]https://ceph.io

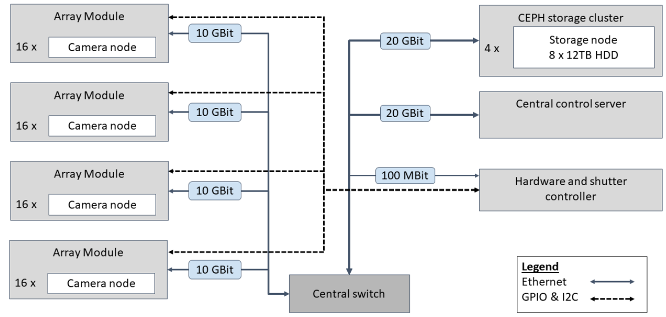

Figure 2. Hardware Architecture

III. Camera Array in Action

One of the first shoots performed using the camera array presented here is called "LF_elements". The captured scenes include walk and movement cycles with multiple people as well as some volumetric smoke and the performance of a fire dancer. The raw version of the frames is available on the SAUCE project website[3]. The follow-up shoot called "Unfolding" features a cellist playing different voices of a self-composed song in front of a green screen. This material will be made available publicly as soon as the project partners have realized their visions for it. To showcase the 5D capabilities of the array especially, the HaToy scene was produced and is shown in figure 3. It revolves around a set of toys on a table, featuring moving elements with varying speeds and complexities. It has not been cleared for public use yet, but can be made available to interested parties upon request.

To fully exploit the 5D capabilities of the camera array, short exposure times must be used, which can result in higher noise level on the captured images. Thus, in order to improve the quality of the captured images, denoising and spatial super-resolution (SR) are applied using the LFBM5D filter [7], [8]. The LFBM5D denoising filter is an extension of the well-known BM3D filter [9] exploiting the non-local self-similarities occurring in natural images, in addition to the

[3]https://www.sauceproject.eu/Downloads

Figure 3. The HaToy Scene

spatio-angular redundancies. The filter relies on 5D patches built from similar 2D patches and filtered in the 5D transform domain, where their spectrum is very sparse and offer a good decorrelation between the true underlying signal and noise coefficients. Noise can thus be filtered by applying hard-thresholding on the 5D transform coefficients in a first step, and Wiener filtering in a second step. The LFBM5D output is then obtained by applying the inverse 5D transform on the filtered 5D spectrum. Note that contrary to the rest of the paper, the 5th dimension of the LFBM5D filter refers to the similarities rather than a temporal dimension. The spatial SR extension uses the sparsisty obtained in the 5D transform domain to regularize the ill-posedness of super-resolution. The algorithm iteratively alternates between a LBM5D filtering step and a back-projection step.

Both denoising and SR were applied on the HaToy scene to demonstrate the effectiveness of the filters. Denoising results are shown in figure 4. Since no ground truth noise free image is available, the quality can not be evaluated with the usual PSNR or SSIM metrics. Thus we use a reference-free method for noise level estimation [10], which shows that the noise level is reduced after applying the LFBM5D filter. SR results are shown in figure 5. The LFBM5D SR filter was applied on the noisy input image and the denoised image, and is compared to bicubic upsampling. Details and high frequency information are clearly better restored by the LFBM5D filter than bicubic upsampling, however when applied to the noisy input, the noise level is also amplified. Applying the LFBM5D denoising prior to applying SR is clearly beneficial.

IV. Conclusion

The camera array is unique due to its modular architecture that makes it easily extendable, the use of off-the-shelf multi-purpose hardware, custom GUI's and the temporal dimension that allows subframing. A modular construction of the camera array enables improvements and adaptation of hardware as well as software components to specific environments. The light field assets that have been captured by the array are valuable for the development of post processing algorithms promising unique visual content.

Figure 4. Left: Noisy image from the 4subframes HaToy scene, σ = 0.57. Right: Corresponding denoised image, σ = 0.29.

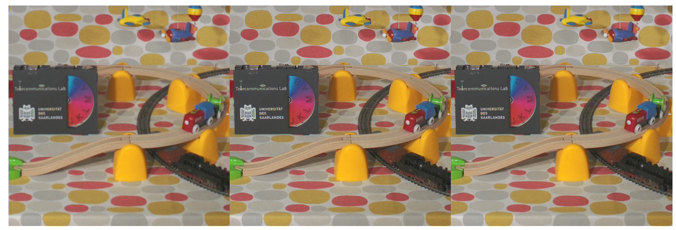

Figure 5. Image from the 4subframes HaToy scene upsampled 2x. Left: bicubic interpolation. Center: LFBM5DSR filter. Right: LFBM5DSR filter applied on denoised images.

The project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 780470.

References

[1] M. Landy and J. A. Movshon, *The Plenoptic Function and the Elements of Early Vision*, pp. 3–20. MITP, 1991.
[2] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGRAPH '96: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, (New York, NY, USA), pp. 31–42, 1996.
[3] T. Herfet, T. Lange, and K. Chelli, "5d light field video capture," in *Proceedings of the 16th ACM SIGGRAPH European Conference on Visual Media Production*, 2019.
[4] M. Grogan and A. Smolic, "L2 based colour correction for light field arrays," in *Proceedings of the 16th ACM SIGGRAPH European Conference on Visual Media Production*, 2019.
[5] Y. Xu, K. Maeno, H. Nagahara, and R.-i. Taniguchi, "Camera array calibration for light field acquisition," *Frontiers of Computer Science*, vol. 9, pp. 691–702, Oct 2015.
[6] S. V. Ila, L. Polok, M. Šolony, and P. Svoboda, "Slam++-a highly efficient and temporally scalable incremental slam framework," *The International Journal of Robotics Research*, vol. 2017, no. 1, pp. 210–230, 2017.
[7] M. Alain and A. Smolic, "Light field denoising by sparse 5D transform domain collaborative filtering," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pp. 1–6, Oct. 2017.
[8] M. Alain and A. Smolic, "Light field super-resolution via LFBM5D sparse coding," in *Proc. IEEE ICIP*, pp. 2501–2505, Oct 2018.
[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
[10] G. Chen, F. Zhu, and P. A. Heng, "An efficient statistical method for image noise level estimation," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 477–485, Dec 2015.