

The Potential of Light Fields in Media Productions

Jonas Trottnow
Simon Spielmann
Filmakademie Baden-Württemberg
research@filmakademie.de

Thorsten Herfet
Tobias Lange
Kelvin Chelli
Saarland Informatics Campus
herfet@cs.uni-saarland.de

Marek Solony
Pavel Smrz
Pavel Zemcik
Brno University of Technology
isolony,smrz,zemcik@fit.vutbr.cz

Weston Aenchbacher
Mairéad Grogan
Martin Alain
Aljosa Smolic
Trinity College Dublin
smolica@scss.tcd.ie

Trevor Canham
Olivier Vu-Thanh
Javier Vázquez-Corral
Marcelo Bertalmío
Universitat Pompeu Fabra
trevor.canham@upf.edu

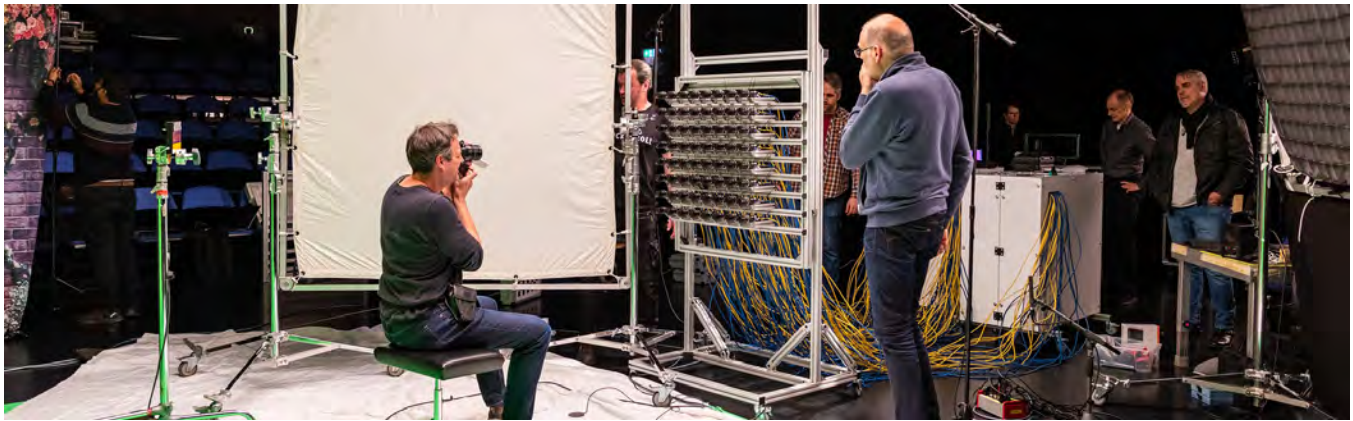


Figure 1: Movie set of light field shoot

ABSTRACT

One aspect of the EU funded project SAUCE is to explore the possibilities and challenges of integrating light field capturing and processing into media productions. A special light field camera was built by Saarland University [Herfet et al. 2018] and is first tested under production conditions in the test production “Unfolding” as part of the SAUCE project. Filmakademie Baden-Württemberg developed the contentual frame, executed the post-production and prepared a complete previsualization. Calibration and post-processing algorithms are developed by the Trinity College Dublin and the Brno University of Technology. This document describes challenges during building and shooting with the light field camera array, as well as its potential and challenges for the post-production.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SA '19 Technical Briefs, November 17–20, 2019, Brisbane, QLD, Australia

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6945-9/19/11...\$15.00

<https://doi.org/10.1145/3355088.3365158>

CCS CONCEPTS

• **Applied computing** → **Media arts**; • **Hardware** → *Sensor devices and platforms*.

KEYWORDS

light field, camera hardware, post-production, compositing

ACM Reference Format:

Jonas Trottnow, Simon Spielmann, Thorsten Herfet, Tobias Lange, Kelvin Chelli, Marek Solony, Pavel Smrz, Pavel Zemcik, Weston Aenchbacher, Mairéad Grogan, Martin Alain, Aljosa Smolic, Trevor Canham, Olivier Vu-Thanh, Javier Vázquez-Corral, and Marcelo Bertalmío. 2019. The Potential of Light Fields in Media Productions. In *SIGGRAPH Asia 2019 Technical Briefs (SA '19 Technical Briefs)*, November 17–20, 2019, Brisbane, QLD, Australia. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3355088.3365158>

1 INTRODUCTION

Besides VFX, nowadays movie productions have to rely on what has been captured on set. It is common to change or extend environments by adding synthetic elements, editing camera parameters (e.g. lens specific blurs, shading, camera position and angles) in post-production. These changes are often time and labor intensive

or even yet impossible. E.g. High quality chroma keying needs significant on set preparation and manual post-processing. Light field data could serve as a powerful new tool [Ziegler et al. 2015], not only as image data, but also as flexible and reusable asset. On top it can serve as input for upcoming holographic and AR/VR devices.

2 THE LIGHT FIELD CAMERA ARRAY

In contrast to a classical, single camera production, the Saarland University's light field (LF) camera array can capture 64 full-HD images (1920x1200 pixels) with a framerate of up to 41 frames per second [Herfet et al. 2018]. The cameras can be arranged in a variety of configurations and distances. For the planned scenery we used a square configuration (figure 1) with a narrow spacing (8x8 cameras, 100 mm spacing). The main challenges have been the data handling (9,07 GB per second raw data at 41 fps) and the physical setup of 64 cameras with their associated control units. 8x8 cameras is a unique feature of the setup, as recent camera arrays were limited to a 3x3 to 5x5 setup [Ziegler et al. 2015][Sabater et al. 2017].

In addition to the spatial flexibility, the array allows to adjust the sampling phase (the temporal position of the frames captured by each individual camera) with an exactness of $10 \mu\text{s}$, trading spatial with temporal resolution in case of fast moving objects in the scene, this is what we call a 5D light field. For the production at hand, all cameras have been shuttered synchronously, using the same frame rate and sampling phase.

3 TEST SCENARIO

Filmakademie carried out a production to test the light field camera system in a close-to-real, VFX driven media production. The scenario reflects a typical VFX setup "person in front of a green screen". Such a setup enables content creators to separate foreground and background elements as well as modify and recompose both in a post-production process. The produced footage will be used to demonstrate how light field data can simplify and enrich this creative process and serves as foundation for further research.

In a white room a cellist sits with her instrument, the scenery is blurred. With the first notes she plays, the focus slowly expands, unveiling the scenery. During her performance, the focus adjusts to the scenery through tilt-shift simulations, focus shifts and T-Stop variations. All those effects are created and animated in post by exploiting the possibilities of the captured light field. The production has been supported by a professional director of photography being essential since the intention was to define the needs of creatives and identify "real world" problems and possibilities in handling light field cameras and data. Long-term goals are to utilize these tools to enrich and simplify post-production in general and produce realistic simulations of physically possible and impossible lenses.

4 ALGORITHMS AND TOOLS

64 raw video feeds are captured on set by the light field camera array. This enormous amount of data needs to be run through a carefully designed pipeline to be able to get compelling results. As a first step after capturing, slight inaccuracies of the material need to be eliminated. This involves consistent color equalization across all cameras and a geometric rectification. Using this 'cleaned' data, depth maps are generated as well as e.g. the simulation and



Figure 2: Frame from Previsualization

animation of physical (tilt-shift) lens effects serving as foundation for new post-production approaches.

4.1 Calibration and Rectification

The algorithms working with light field data such as synthetic aperture, super-resolution, depth estimation, and/or scene 3D reconstruction, require accurate camera parameters and rectified data. In an ideal camera array setup, the optical centers of cameras would lie at the exact physical positions in the grid and their optical axis would be parallel. Due to the manual setup of the camera system, the camera extrinsic parameters (position and rotation) may slightly differ from the intended pose in the grid. Additionally, the intrinsic parameters of cameras (focal length, aspect ratio, principal point, distortion coefficients) vary from device to device. All of these parameters need to be obtained through a joint camera calibration and the acquired data need to be rectified - all the images adjusted to camera parameters and transformed onto a common plane.

The calibration process does not require any complex scene. A marker-less calibration [Xu et al. 2015] can be performed based on the concept similar to Structure from Motion [Håming and Peters 2010] on a scene moderately rich in features with clearly identifiable fore-, mid-, and background. In our approach, we used the graph-based non-linear optimizer SLAM++ [Ila et al. 2017] with its very efficient non-linear least squares solvers based on fast sparse block matrix manipulation, whose bundle adjustment optimizer refines the camera parameters.

The modular implementation of SLAM++ allows to easily integrate constraints into the optimization system, in our case the specific known constraints about the camera positions on the grid. An optimizer refines the parameters of the cameras to satisfy those constraints and at the same time minimizes the reprojection error of 3D points. This significantly improves the estimate of the camera positions and rotations in the scene.

This approach provides a robust and significantly more accurate estimate of the camera parameters compared to the extended OpenCV stereo calibration approach. The output of the calibration process for each of the cameras contains the corresponding intrinsic and extrinsic parameters. The calibration, if implemented in a GPU based system, such as a GeForce GTX 1080 Ti accelerated workstation with an Intel i7-7700 CPU, takes only few seconds to provide the complete camera parameters.

Once the accurate camera parameters are obtained, rectification is straightforward to complete with conventional methods [Hartley 1999][Loop and Zhang 1999]. Homographies are computed for every camera using a set of point correspondences which is then used to unwrap the corresponding image. Depending on the application, the rectified images are either cropped or the post processing algorithms are modified to consider the unequal viewing areas of the rectified images.

4.2 Color Equalization

For light field camera arrays, a color equalization step is important to account for differences in color balance between the different cameras. Two color equalization schemes are proposed, which use SIFT correspondences between reference and test image pairs to estimate color correction functions which ensure color consistency is maintained across all views.

The first method is an extension of the color stabilization method of [Vazquez-Corral and Bertalmio 2014], originally intended to conform footage of common scene elements from multiple cameras/color encoding schemes. It equalizes the colors between a test and reference image pair by determining an optimal 3x3 matrix and non-linear correction term (inspired by the camera encoding pipeline) which minimizing the difference in colors between them.

When a color chart is present in the scene, a second method has also been investigated. We extend the method in [Grogan and Dahyot 2019], which uses color correspondences between images to compute a thin plate spline transformation which maps the colour distribution of one image to match the other. We extend this approach to account for both SIFT correspondences between test and reference image pairs, as well as correspondences between the captured and ground truth color charts. This ensures that colors across the light field remain consistent while also matching the ground truth colors.

With both methods, the center view is taken as the reference view, and all other view's colors are aligned to it. A color propagation scheme has also been considered—to account for high disparities—in which each concentric circle of views around the center is corrected and later used as references for the following circle. Colors captured by each camera are found to be temporally consistent, so the color correction function that is estimated for a single light field video frame can be applied to the whole video sequence without introducing temporal inconsistencies.

4.3 Post-Processing

Post-production steps enabled by light field capture include the creation of depth maps, 3D point clouds, change of the virtual camera aperture, and digital refocus with a variable depth of field.

4.3.1 Depth Maps and Point Clouds. To produce depth maps for the “Unfolding” data, horizontal pairwise disparity estimation is performed on neighboring views, followed by an edge-aware filter along the spatial and angular light field dimensions. This is a refinement of previous work [Chen et al. 2017] and produces a depth map for each camera. Having a separate depth map for each camera is valuable for camera array light fields because the large camera separations mean they view different objects, both with respect to field of view and parallax. The drawback to using pairwise depth estimation is that there can be depth discrepancies

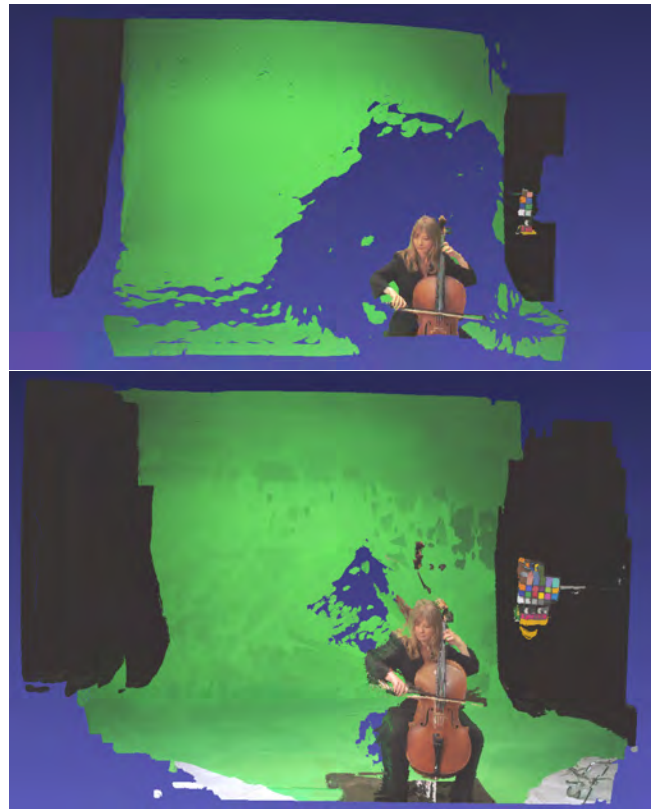


Figure 3: Single-camera (top) and aggregate (bottom) point clouds.

between views. Together with camera calibration inaccuracies, this can lead to surfaces appearing “extended” in depth in point clouds, as seen in Figure 3.

Scale 3D point clouds are created to visualize the scene by reverse projection, utilizing the depth maps and camera parameters found through calibration (Section 4.1). Because the positions of the cameras are known, in relation to one another, single-camera point clouds can be combined in a common 3D world space to create an aggregate point cloud. In this dense point cloud, each view fills in holes present in the point clouds of other views. Though outliers are largely filtered [Rusu et al. 2008], remaining clusters create a sense of “blockiness” in the background. In the future, depth information might also be used to accelerate the separation of foreground and background elements.

4.3.2 Digital Refocus and Perspective Shift. A common light field operation is to simulate a change in focal length of a single camera with narrow depth of field [Ng et al. 2005]. This is achieved in here by using a generalized shift-and-sum framework [Alain et al. 2019]. Further, this framework does not require the refocus plane to be parallel to the plane of the camera array plane, so it is possible to create physically-impossible “tilt-shift refocus” images. In addition to refocus, this framework includes the ability to change the size and position of the virtual camera aperture. An example result is shown in Figure 4 (bottom). Note that angular aliasing artifacts

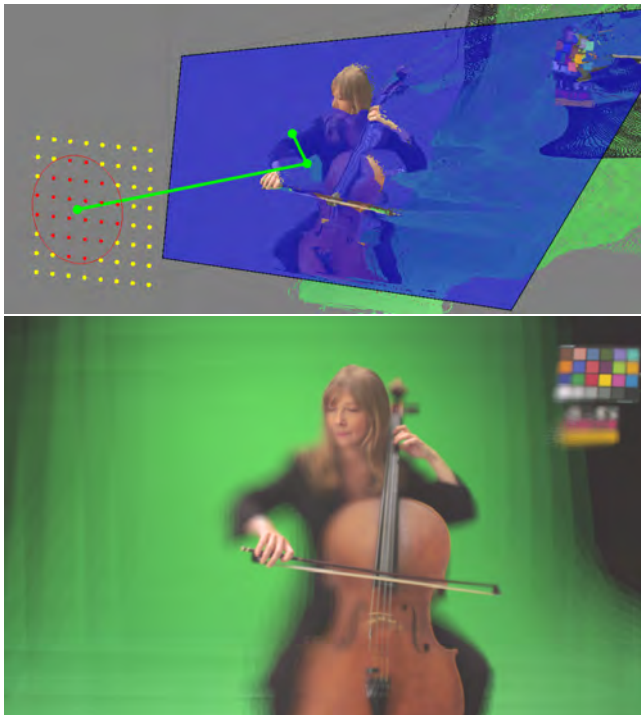


Figure 4: Refocus geometry (top) and corresponding tilt-shift refocus result (bottom).

appear in out-of-focus regions. Removal of these artifacts through view interpolation and filtering is the subject of current work.

When all of the elements above are combined, the user is provided with an interactive tilt-shift refocus tool that allows them to control aperture parameters, quickly set the refocus plane position, and visualize the refocus plane in the point cloud (Figure 4, top). The refocus result can be produced, evaluated, and refined by using fine controls. This allows the artist to easily create previously-impossible focal effects and gives a director more creative options, even after the scene has been shot.

5 POTENTIAL & EVALUATION

To test the production setup and provide ground truth data, a 3D virtual version of the scenery was designed. The set has been created in the DCC application Blender, and made it possible to plan the physical setup, stage the camera and define timings. Furthermore it was possible to produce light field data by rendering synthetic images of 64 virtual cameras without the need of a physical setup. This sped up the whole creative process and gave us the possibility to define the final look in cooperation with a professional director of photography. Additionally, it was easier to avoid physical imponderables when handling the complex, large camera system, being one of the biggest challenges during production.

In order to enable creatives to benefit from LFs, several challenges have to be overcome. First of all, efficient compression methods reducing the amount of captured data are required. This would be a starting point to make such data usable in a post-production pipeline. Looking at potential benefits LF-data can provide, the most useful would be to generate high quality, dynamic (moving) depth

maps or even geometric reconstructions. This would not only assist to simplify color keying but also unlocks possibilities that could lead to a workflow to freely edit the captured scenery including its geometric information, lighting and surface characteristics after the actual recording, making the data a smart, reusable asset.

The next steps can be directly derived from the test production. Next to the separation of foreground and background elements, the production required solutions to intuitively change and simulate camera lens effects. Full control over focus, including the focus plane (tilt-shift) and the lens blur appearance (f-stops) in an artist driven post-production is the target. Solutions for changing the camera's perspective would be very useful tools and could help to overcome physical limitations of the LF-camera rig.

Another aspect of future LF usage is the direct presentation of the captured LF footage by a capable display system.

In summary, LFs prove to be very valuable for future media post-productions and make captured footage reusable and future-proof. Technical challenges still exist but are already being solved. Artist friendly tools will be the key to a wider adoption.

Sample light field data captured with the described camera rig has been publicly released [SAUCE 2019].

ACKNOWLEDGMENTS

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

REFERENCES

- M. Alain, W. Aenchbacher, and A. Smolic. 2019. Interactive light field tilt-shift refocus with generalized shift-and-sum. In *Proc. European Light Field Imaging Workshop*.
- Y. Chen, M. Alain, and A. Smolic. 2017. Fast and accurate optical flow based depth map estimation from light fields. In *Proc. of the Irish Machine Vision and Image Processing Conf. (2017-08-30)*.
- M. Grogan and R. Dahyot. 2019. L2 Divergence for robust colour transfer. *Computer Vision and Image Understanding* 181 (2019), 39–49.
- K. Häming and G. Peters. 2010. The structure-from-motion reconstruction pipeline - a survey with focus on short image sequences. *Kybernetika* 46 (2010).
- R. I. Hartley. 1999. Theory and Practice of Projective Rectification. *International Journal of Computer Vision* 35, 2 (01 Nov 1999), 115–127.
- T. Herfet, T. Lange, and H. Priyadarshini Hariharan. 2018. Enabling Multiview- and Light Field-Video for Veridical Visual Experiences. In *Computer and Communications (ICCC), 2018 4th IEEE International Conference on (ICCC '18)*. IEEE, Chengdu, China.
- V. Ila, L. Polok, M. Solony, and P. Svoboda. 2017. SLAM++-A highly efficient and temporally scalable incremental SLAM framework. *The International Journal of Robotics Research* 36, 2 (2017), 210–230.
- C. Loop and Z. Zhang. 1999. Computing rectifying homographies for stereo vision. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, Vol. 1. 125–131 Vol. 1.
- R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. 2005. Light field photography with a hand-held plenoptic camera. *Stanford Technical Report CSTR 2 (2005)*, 1–11.
- R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz. 2008. Towards 3D Point cloud based object maps for household environments. *Robotics and Autonomous Systems* 56, 11 (2008), 927–941. Semantic Knowledge in Robotics.
- N. Sabater, G. Boisson, B. Vandame, P. Kerbirou, F. Babon, M. Hog, T. Langlois, R. Gendrot, O. Bureller, A. Schubert, and V. Allie. 2017. Dataset and Pipeline for Multi-View Light-Field Video. In *CVPR Workshops*.
- SAUCE. 2019. Lightfield Data. (2019). <https://animationsinstitut.de/en/lf>
- J. Vazquez-Corral and M. Bertalmio. 2014. Color Stabilization Along Time and Across Shots of the Same Scene, for One or Several Cameras of Unknown Specifications. *IEEE Transactions on Image Processing (TIP)* 23, 10 (Oct 2014), 4564–4575.
- Y. Xu, K. Maeno, H. Nagahara, and R. Taniguchi. 2015. Camera array calibration for light field acquisition. *Frontiers of Computer Science* 9, 5 (01 Oct 2015), 691–702.
- M. Ziegler, A. Engelhardt, S. Müller, J. Keinert, F. Zilly, S. Foessel, and K. Schmid. 2015. Multi-camera System for Depth Based Visual Effects and Compositing. In *Proceedings of the 12th European Conference on Visual Media Production (CVMP '15)*. ACM, New York, NY, USA, Article 3, 10 pages.