5D Light Field Video Capture

Thorsten Herfet
https://www.nt.uni-saarland.de/people/thorsten-herfet/
Tobias Lange
https://www.nt.uni-saarland.de/people/tobias-lange/
Kelvin Chelli
https://www.nt.uni-saarland.de/people/kelvin-augustin-chelli/

Saarland Informatics Campus,
Telecommunications Lab,
Saarland University,
Saarbrücken,
Germany

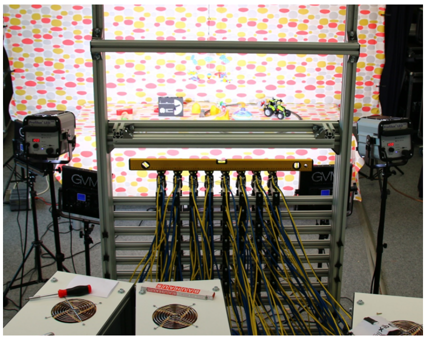

Figure 1: 5D Light Field Capture seen from behind the scene

While plenoptic still imaging, mostly due to the small baseline / frustum and the limited single sensor resolution, has not reached market impact, multi-view video remains to be a very hot topic. With baselines of only a small percentage of the picture width respectively height, multi-view video capture becomes *sparse light field capture*! This short paper introduces 5D sparse light field capture, i. e. a rig based capture system enabling a balanced spatial, angular and temporal resolution.

The Capture Device

Camera rigs with several tens of cameras avoid the limitation of plenoptic dense light field capture:

- The spatial resolution due to several tens of single sensors is dramatically increased.
- The relation between spatial and angular resolution is *healthier* than with most plenoptic cameras.
- The effective aperture and hence light sensitivity is large.
- By careful control of the exposure time the 5th dimension (temporal) can be exploited.

We have built a camera rig consisting of 64 1920x1200@40fps full-HD cameras (Figure 1) and designed a sync plane enabling a controlled adjustment of the temporal behaviour of the rig:

- The *frequency* is controlled in a way that the cameras are gen-locked. They all run with the same frame-rate.
- The *phase* is controlled per camera resp. per sub-array. With this we can subdivide the frame into a variable number of sub-frames and consequently increase the temporal resolution up to a factor as large as the number of cameras with overlapping frustum.

In the paper we will denote light field still imaging as 4D light fields ([3]), synchronized light field video as 4.5D light fields ([2]) and sub-framed light field video as 5D light field video. We choose 5D to denote that in theory each ray can carry four spatial and one temporal dimension. In the current array implementation we're of course physically limited to sub-framed, synchronized light field video.

The HaToy scene

To demonstrate the advantages of sub-framing we've setup a scene with moving toys and tools (Figure 2). A kids mobile, travelling trains, a spin-top and a speed controlled CD drive are assembled to produce all kinds of temporal behaviour: Occlusion and dis-occlusion, translation, rotation and scaling can be seen. At each time instance only parts of the texture on the CD and the spin-top are visible. Due to excellent lighting (we use ~20 thousand Lux of lighting) we can capture with exposure times as small as 350 μs and hence maintain a good spatial resolution even for fast moving objects. The CD drive is controlled to turn with 40 or 80 rps (2.4k or 4.8k rpm), perfectly matching the frame-rate of the array. The full texture can hence only be revealed by sub-framing with at least 2 sub-frames.

Figure 2: The HaToy Scene

Sub-Framing

When sub-framing, we consider *locality*: Neighbouring cameras have the largest scene overlap and hence when producing *N* sub-frames, local sub-arrays with *N* cameras should contain all possible phases. In addition, the phases should be equidistantly spaced.This ensures that the temporal resolution is increased by a factor of *N* while keeping the angular relationship between the motion phases as close as possible. There is no solution fully fulfilling the constraints, hence we implemented a flexible sub-framing pattern by extending the so called *Bit Reversal Permutation* to two dimensions. This is achieved by interleaving the permuted bits to preserve the significance. The general result, bit reversal permuting 8^2 cameras, is given in Figure 3. MATLAB®-code is freely available. For evaluation of subframing patterns we've also setup a Blender ([1]) environment which generates a complete set of sub-frames and by sub-sampling is able to produce arbitrary sub-frame patterns.

$$\begin{bmatrix} 0 & 32 & 8 & 40 & 2 & 34 & 10 & 42 \\ 16 & 48 & 24 & 56 & 18 & 50 & 26 & 58 \\ 4 & 36 & 12 & 44 & 6 & 38 & 14 & 46 \\ 20 & 52 & 28 & 60 & 22 & 54 & 30 & 62 \\ 1 & 33 & 9 & 41 & 3 & 35 & 11 & 43 \\ 17 & 49 & 25 & 57 & 19 & 51 & 27 & 59 \\ 5 & 37 & 13 & 45 & 7 & 39 & 15 & 47 \\ 21 & 53 & 29 & 61 & 23 & 55 & 31 & 63 \end{bmatrix} \quad \begin{bmatrix} 0 & 4 & 1 & 5 & \dots \\ 2 & 6 & 3 & 7 & \dots \\ 0 & 4 & 1 & 5 & \dots \\ 2 & 6 & 3 & 7 & \dots \\ 0 & 4 & 1 & 5 & \dots \\ 2 & 6 & 3 & 7 & \dots \\ 0 & 4 & 1 & 5 & \dots \\ 2 & 6 & 3 & 7 & \dots \end{bmatrix}$$

Figure 3: Bit Reversal Permutation for 64 cameras with 64 (left) and 8 (right) sub-frames

Acknowledgement

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

[1] Blender, made by you. https://www.blender.org/. Accessed: 2019-11-11.

[2] Light field video dataset captured by a r8 raytrix camera. http://clim.inria.fr/Datasets/RaytrixR8Dataset-5x5/index.html. Accessed: 2019-11-11.

[3] Inria light field dataset. https://www.irisa.fr/temics/demos/lightField/LowRank2/datasets/datasets.html. Accessed: 2019-11-11.