Brutus: A Mid-Range Multi-Camera Array for Immersive Light Field Video Capture

Jay Busch
jbusch@google.com

Peter Hedman
hedman@google.com

Matthew DuVall
matthewduvall@google.com

Matt Whalen
mswhalen@google.com

Michael Broxton
broxton@google.com

John Flynn
jflynn@google.com

Ryan Overbeck
rover@google.com

Daniel Erickson
danerickson@google.com

Paul Debevec
debevec@google.com

Figure 1: Our mid-range light field video camera array (left) consisting of 24 Z CAM E2 professional 4K cinema cameras mounted to a sturdy acrylic hemisphere is seen recording a performance of a traditional erhu instrument.

ABSTRACT

We present a mid-range multi-camera system for recording panoramic light field video content. The proposed system captures wide baseline (0.9 meters), high resolution (>20 pixels per degree), large field of view (>220°) light fields at 60 frames per second. The array contains 24 time-synchronized cameras distributed across the surface of a 60cm diameter hemispherical plastic dome. We use Z CAM E2 professional cinema cameras mounted to the exterior of the dome using 3D printed brackets. The dome, mounts, triggering hardware and cameras are only moderately expensive and the array itself is straightforward to fabricate. Using the recently developed DeepView Video pipeline, we can record and render subjects as close as 50cm from the cameras.

CCS CONCEPTS

• **Computing methodologies → Image and video acquisition**; **Rendering**.

KEYWORDS

light field, camera array, image-based rendering, view synthesis

Reference Format:
Jay Busch, Peter Hedman, Matthew DuVall, Matt Whalen, Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, and Paul Debevec. 2020. Brutus: A Mid-Range Multi-Camera Array for Immersive Light Field Video Capture.

1 INTRODUCTION

With positional tracking VR headsets becoming higher resolution and increasingly available, the demand for 6-degrees-of-freedom spatial video has never been greater. Significant steps toward a practical 6DOF video production process have recently been presented in [Pozo et al. 2019] and [Broxton et al. 2020], which take different approaches to the problem including the capture methodology. [Pozo et al. 2019] uses a high-quality set of sixteen 8K digital cinema cameras with bespoke optics to achieve over 30 pixels per degree over a spherical field of view, but the system is heavy and its cameras alone would retail for $400k. [Broxton et al. 2020] uses a hemispherical array of 46 action sports cameras, achieving 15 pixels per degree over 180 degrees field of view, in a lightweight system around $5k.

In this work, we present a novel camera array design for immersive light field video which falls between these two systems. Built from 24 Z-Cam e2 cameras, our new system is portable, moderately expensive at $50k, straightforward to build, and easy to use. Our rig significantly improves on quality of the inexpensive rig of [Broxton et al. 2020], producing over 20 pixels per degree, and better synchronized images both temporally and photometrically.

In this poster, we describe the key contributions of our new light field camera rig. We detail the benefits of choosing the Z-Cam e2 cameras, and how we adapted the acrylic dome technique of [Broxton et al. 2020] to our heavier camera array. We also describe our camera layout design and how it affects the quality of the recorded light field.

2 CAMERA ARRAY DESIGN

Our camera array (Fig. 1), consists of 24 Z CAM E2 professional 4K cinema cameras[1]. A single controller camera controls the others via manufacturer supplied daisy-chained sync cables. The cameras can be powered by wall current or external V-mount or NP-F batteries, drawing approximately 1000W. After recording, the data can be downloaded via the RJ45 cables to a local area network by ftp.

The cameras are set to the relatively square 3696×2772 pixel resolution running at 60fps in the H.265 video codec, allowing approximately 2 hours of recording time onto 256GB CFast cards. We maximize dynamic range using the *Z-Log2* 10-bit color space.

We tested the synchronization of the cameras by placing an unlensed subset of the array in an unlit room and manually discharging a camera flash approximately every ten seconds, with the shutter set to 60fps and 360° (i.e. exposing the entire frame duration). The flashes appeared as bright areas starting consistently at the same pixel row down the rolling shutter frames, indicating line-level sub-millisecond sync, with no drift after five minutes. Furthermore, around one third of the near-instant camera flashes exposed the entire camera sensor, suggesting the total readout time to be two thirds of 1/60th sec or about 11ms. As adjacent cameras point within 30° of each other and cover nearly 180° FOV, any point in the scene should be imaged consistently by the array within 2ms.

We selected the Meike 6.5mm MFT fisheye lens for our cameras, measuring it to be sharp (23 pixels per degree), inexpensive, and very wide FOV at 190° horizontal, maximizing the overlap between camera views. To avoid inconsistency in the field, we used adhesive to fix each lens to an aperture of f/8 and a focus distance of 2m, producing sharp images from 0.8m to infinity.

3 GEOMETRIC LAYOUT

We initialized our cameras in evenly spaced rings of 3, 9, and 12 cameras placed at 15°, 45°, and 75° from the apex of the dome, respectively. We then used gradient descent with inverse-square distance forces to space them out more evenly, constraining the vertex locations to lie on the surface of the sphere. During optimization, we mirrored the camera arrangement to find camera locations for a full sphere. While our rig only uses 24 cameras from

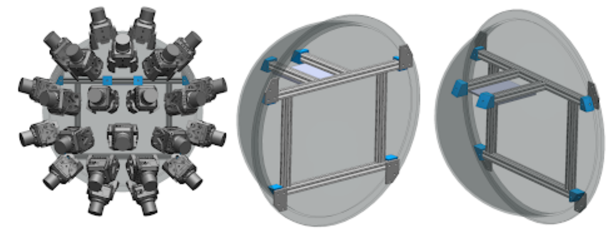

Figure 2: Camera arrangement and reinforced acrylic dome.

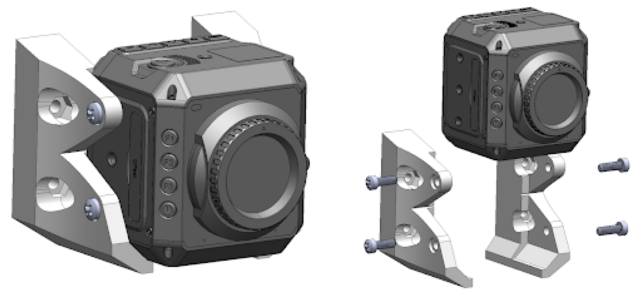

Figure 3: 3D printed brackets hold the cameras securely to the dome at four mounting points.

one hemisphere in the rig, this approach makes it possible to form an evenly spaced spherical arrangement of 48 cameras using two hemispherical domes. From the analytic technique of [Broxton et al. 2020], our array sees each point in the scene in at least three views as long as it is no closer than 50cm from the camera lenses, with about 22cm between camera lenses.

Since each camera-lens pair weighs 1kg, we used thick 12mm acrylic for the dome, placing the cameras on the outside so the dome can be as compact and strong as possible. Each mounted camera protrudes 15cm from the 60cm diameter dome, placing the lenses on a 90cm sphere. The 3D printed mounts were made from shatter-resistant photopolymer resin. We used 3D printed templates to drill holes for the camera's sync, power, and network cables.

4 PRODUCTION USE

We have used our system to record a variety of scenes in both indoor and outdoor locations, focusing on content which involves performers about 1m from the cameras. Since the fully populated dome is somewhat heavy at 40kg, the cameras are transported separately and require 30 minutes to install before a shoot. We plan to release downloadable light field video to view in VR from this new camera array in time for SIGGRAPH Asia 2020.

REFERENCES

Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew DuVall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. 2020. Immersive Light Field Video with a Layered Mesh Representation. *ACM Trans. Graph.* 39, 3 (Aug. 2020).

Albert Parra Pozo, Michael Toksvig, Terry Filiba Schrager, Joyce Hsu, Uday Mathur, Alexander Sorkine-Hornung, Rick Szeliski, and Brian Cabral. 2019. An Integrated 6DoF Video Camera and System Design. *ACM Trans. Graph.* 38, 6, Article 216 (Nov. 2019), 16 pages. https://doi.org/10.1145/3355089.3356555

[1]The rig is called "Brutus" in reference to Shakespeare's *Julius Caesar* ("E2, Brute?")