Reality Capture: Present and Future

DEVON COPLEY

FORMER HEAD OF PRODUCT, NOKIA OZO VR PLATFORM

3GPP/VRIF WORKSHOP ON VR ECOSYSTEM & STANDARDS, DEC 2017

Introduction

“Reality Capture” is the term we used inside Nokia to describe our goal with the OZO Camera project.

There are a variety of reality capture technologies that are in field use and under active development. They each have benefits and constraints, and what works for one use case may be inappropriate or impractical for another.

This presentation is an attempt to survey the state of the art and speculate on the future trajectories of these technologies. I don’t claim to be an expert on all of these topics; please help correct any errors!

Reality capture technologies

360 Video capture

- Standard 360 video
- “3DOF+” video (video+depth)

Light-field capture

Volumetric capture

- Photogrammetry
- Range imaging

360 Video Capture

Standard 360 video capture: overview

The most mature and widely deployed reality capture technology is built upon existing video technology, typically combining two or more standard planar sensors with fisheye lenses. These images are “stitched” together into a video frame using a spherical projection. A variety of different devices and rigs are available across two orders of magnitude of cost.

360 video capture: pros & cons

PROS

- Capture technology is readily available at a wide variety of price points
- Most post-production tasks now supported by standard video production software (e.g. Adobe)
- Standards are emerging (OMAF)
- Video codecs are widely supported throughout the ecosystem, so both encoding and decoding are efficient
- Very low demands on client devices
- Stereoscopic capture provides some parallax and sense of depth for viewer

CONS

- Only enables 3DOF movement for viewer
- Even the best stereoscopic 360 video capture and render systems do not accurately represent binocular+motion parallax
- Existing video codecs were not designed for reality capture and are insufficient in a number of ways
- Representation of entire sphere is wasteful
- Current capture and post-processing tools are only beginning to reach beyond 4K horizontal resolution

“3DOF+” video capture

Various companies are working on enhancing 360 video with **depth information**, in order to enable accurate parallax for a limited range of motion during playback. Facebook’s marketing people call this “6DOF;” the engineers at MPEG call it “3DOF+.”

Currently, no commercially available cameras, production software, or 360 players enable 3DOF+. However **this technology will become available during 2018 from several providers**, **most notably Facebook.**

3DOF+ uses either image disparity information or additional sensor data to generate a depth map that corresponds to the RGB image. This allows a capable player to re-project the video frames to reflect the viewer’s virtual movement in space.

“3DOF+” video capture: pros & cons

PROS

- Leverages much of the existing 360 video infrastructure
- Dramatically improves the sensation of presence due to accurate motion and binocular parallax
- Enables content insertion and depth-buffer keying for mixed reality applications

CONS

- Movement only in a limited range
- Readily visible artifacts (will be reduced if not eliminated via machine learning techniques)
- Theoretically impossible to handle specular effects (reflections) or translucence
- 2x-3x data load versus standard 360
- No current standards for depth layer representation or delivery; planned for future OMAF standard
- Relatively heavy client-side demands

360 video capture: forecast

2018-2019	• Monoscopic 360 video remains most common reality capture mode for all but the highest-end productions. • “3DOF+” is introduced and begins to gain traction in high-end offline productions, supplanting stereoscopic 360 almost entirely • Video production tools enable 3DOF+ support
2020-2021	• 3DOF+ standards emerge; 3DOF+ enters live, semipro and consumer applications
2022+	• 360 and 3DOF+ remain in use for live and resource-restricted capture applications but are supplanted by other technologies for most other use cases

Light-field Capture

Light-field capture: overview

A **light-field** or **plenoptic** camera is a fundamentally different technology for capturing imagery.

It uses an array of planar sensors (or, alternatively, an array of micro-lenses in front of a single sensor) to capture the same scene from a range of locations. Rather than capturing an image as a 2-D grid of light intensity and color values, this technique allows the capture of a “4D light field,” representing intensity and color as *vectors within a volume.*

This capture technology promises extremely realistic immersive imagery, with accurate parallax and correct specular effects. However, it has a number of drawbacks, most notably the massive data loads required.

Light-field capture: pros & cons

PROS

- Provides the most realistic imagery of any currently feasible capture technology
- Significant improvements over 3DOF+ video: fewer artifacts, correct specular effects. Handles reflections, smoke, and translucency well.
- More accurate depth extraction enables mixed reality applications
- Easily rendered down to 3DOF+ or 360 video

CONS

- Enormous data rates. Lytro's first-gen Immerge camera generates **400GB/sec!**
- Size and complexity of capture device
- Complete lack of tools, workflows, or delivery standards
- 6DOF light-field rendering constrained to the volume of the camera itself. Limited benefits for rendering a POV outside the camera volume.
- Rendering requires specialized hardware

Light-field capture: forecast

2018-2019	• Light-field remains largely experimental, useful for one-off projects delivered to controlled consumption environments.
2020-2021	• Light-field capture becomes more common for high-end narrative and documentary content. Delivery becomes feasible through increased bitrates and improved compression; light-field rendering becomes common for both live-action and synthetic content.
2022+	• Light-field displays become feasible. Capture devices are available for semipro and consumer tiers. Delivery formats and editing tools become standardized. Light-field capture supplants 360 video for most use cases, although live capture is feasible only for highest-value events.

Volumetric Capture

Volumetric capture: photogrammetry

Photogrammetry describes a range of techniques by which a 3D representation is derived from a series of planar images taken from different locations. It is an extremely active area of research.

There are three broad categories of photogrammetry:

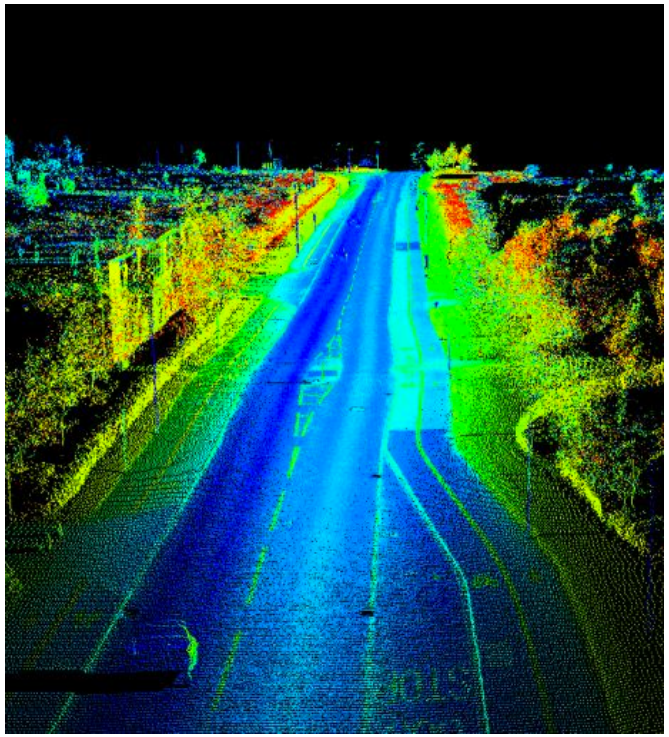
- **Convergent photogrammetry,** also known as “outside-in,” uses an array of cameras in a roughly spherical arrangement capturing a scene or object in the middle of the sphere.
- **Stereophotogrammetry** extracts depth information from two overlapping images.
- **Structure from motion** or SfM, a relatively recent innovation, extracts 3D scene information from a series of images captured in motion through a space, simultaneously deriving camera pose and scene geometry.

Photogrammetry typically creates **colored point cloud data** which can be converted in to a **mesh** and texture-mapped with imagery.

Volumetric capture: range imaging (1)

Several direct range-sensing technologies complement visual capture and enable more accurate 3D modeling.

- **Scanning LiDAR** scans the scene with a laser, providing very accurate range data albeit at a relatively low resolution and refresh rate.
- **Time-of-flight LiDAR** illuminates the scene with a pulse of light, capturing the entire scene in a single frame.
- **Structured Light** projects an infrared pattern on the scene which is imaged using an IR camera; the distortion of the pattern allows depth information to be extracted.



Volumetric capture: range imaging (2)

Range-sensing technology is under active development by a range of vendors and will rapidly develop over the next few years.

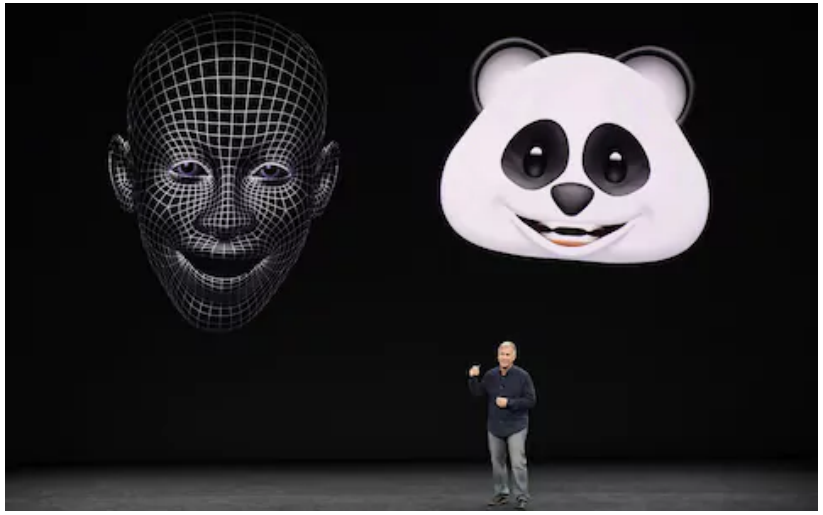
Apple's iPhone X uses IR laser scanning for real-time depth-based facial recognition (and "animojis")

Leica's BLK360 is a new product bringing high-resolution, high-frame rate scanning LiDAR to a new price point (<$20k)

New **solid-state LiDAR** technology from **Velodyne** and others will bring higher frame rates and much lower costs (~$500 in volume)

Intel's RealSense tech is now available and they are pushing it with OEMs

Structure is now shipping a low-cost structured-light depth camera



Volumetric capture: forecast

2018-2019	• LiDAR becomes more and more feasible for many applications • Integration of depth information with RGB data is problematic
2020-2021	• Depth-sensing tech becomes standard on smartphones to service a variety of use cases (AR, inside-out tracking, etc) • Production tools for RGBD content become widely available
2022+	• Pointcloud and animated mesh formats and distribution become standardized

Convergence

Coming soon: capture tech mash-ups

Soon, these different capture and rendering techniques are likely to be combined to create solutions for different use cases.

For example, producers might use photogrammetry techniques to obtain a high-resolution 3D map of a space or landscape, and composite that static 6DOF dataset with real-time capture of foreground content from a 360 or light-field camera.

Or one might virtually superimpose an outside-in-captured hologram into a 3DOF+ capture, with correct parallax and occlusions.

Or one might deliver a high-quality light-field limited 6DOF experience to VR viewers and simultaneously offer an AR experience with the background information removed using depth-buffer keying.

Thank you!

DEVON COPLEY

DEVON@IMEVO.COM