

High-Quality Streamable Free-Viewpoint Video: Supplemental Material

Alvaro Collet Ming Chuang Pat Sweeney Don Gillett Dennis Evseev David Calabrese
Hugues Hoppe Adam Kirk Steve Sullivan
Microsoft Corporation

Abstract

In this supplemental material, we describe our default, tower-based camera configuration for the capture stage. We analyze the local configuration of a single tower and visualize the combined camera frustum. We also describe our camera calibration process and show larger prints for the experiments on different camera configurations.

1 Stage Configuration

The current capture stage is a greenscreen volume encircled by 106 synchronized high-speed video cameras, as shown in Fig. 2. These cameras are mounted on 8 wheeled towers (with 12 cameras each), plus 10 cameras mounted in an array overhead. The default configuration is 53 Sentech STC-CMC4MCL RGB cameras and 53 Sentech STC-CMB4MCL IR cameras covering a reconstruction volume of 2.8 m in diameter and 2.5 m in height. We record 2048x2048 images at 30 Hz, though we can go up to 60 Hz for high-speed action. We add unstructured static IR laser light sources on the towers to provide strong IR texture cues.

We reached the current stage configuration after multiple iterations. Our initial implementations contained 16 cameras, then 32, and we developed most of our system using 40 cameras. After establishing targets for quality, acquisition volume, and types of content, we built our current configuration: 106 cameras on movable towers plus an overhead rig, mapped dynamically (at processing time) into pods (logical groups of 2 IR cameras + 2 RGB cameras). We can easily change camera count and reconfigure to different volume sizes/shapes, and the quality improves/degrades gracefully as the configuration changes. Given our unconstrained capture content, we found that a denser sampling of viewpoints was particularly critical to resolve severe occlusions (e.g., multiple actors/props interacting closely) and recover fine features with silhouettes complementing stereo reconstruction.

1.1 Tower configuration

Fig. 1(left) shows an example of one of our wheeled towers. Each tower contains 12 cameras (6 RGB and 6 IR) mapped into 3 pods of 2 RGB and 2 IR cameras each. Stereo depth computations are performed only within a pod. The baseline for each stereo pair is approximately 60 cm, which corresponds to angles of triangulation between 7 and 13 degrees within the capture volume.

Our configuration for RGB cameras is slightly different than for IR cameras. We opted for a horizontal arrangement of RGB cameras within each tower to provide a more even sampling of the view space, as the RGB cameras are used for texturing. The IR pairs are positioned vertically to maximize the coverage of standing subjects for reconstruction (see Fig. 3).

2 Calibration

We perform a checkerboard-based calibration of the camera setup (intrinsic and extrinsic) before every set of takes. To speed up

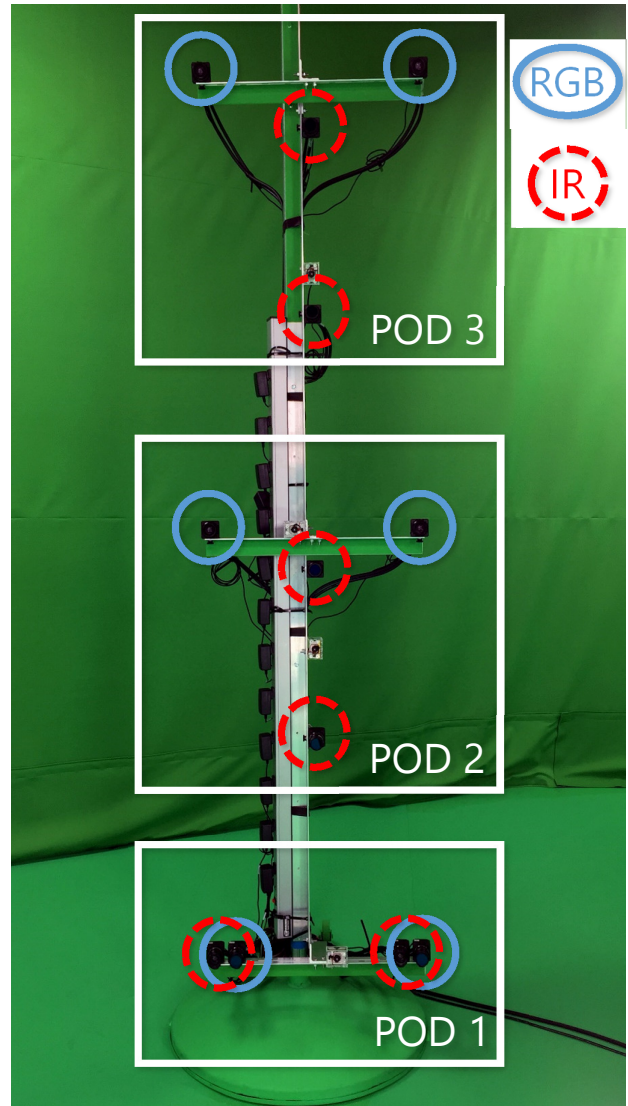


Figure 1: Configuration of a single tower, and its mappings into camera pods and stereo pairs.

this typically tedious process, we built a wheeled octagonal tower which we call the Octolith, as shown in Fig. 4. With the Octolith, we only need to capture a few frames (typically five) at different positions within the capture volume to robustly compute the intrinsic and extrinsic calibration parameters.

3 Experiments on camera configuration

Fig. 5 shows larger prints of the evaluation of geometry and texture in different camera configurations (main paper, Fig. 13).

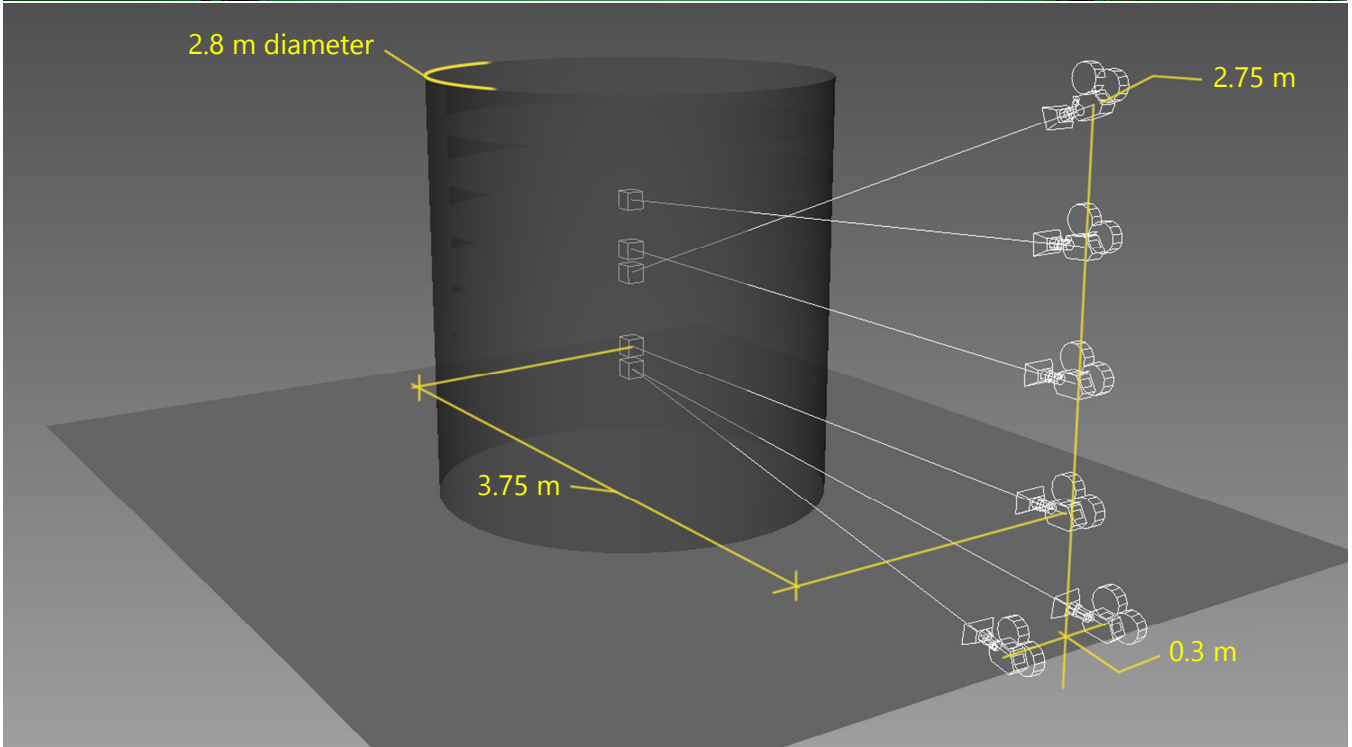


Figure 2: (top) View of our full stage in its default 106-camera configuration. (bottom) Base measurements of the stage in its default configuration. The reconstructed area is a cylinder of 2.8 m diameter with 2.5 m height. We also show a visualization of the IR cameras of a single tower to provide a sense of scale.

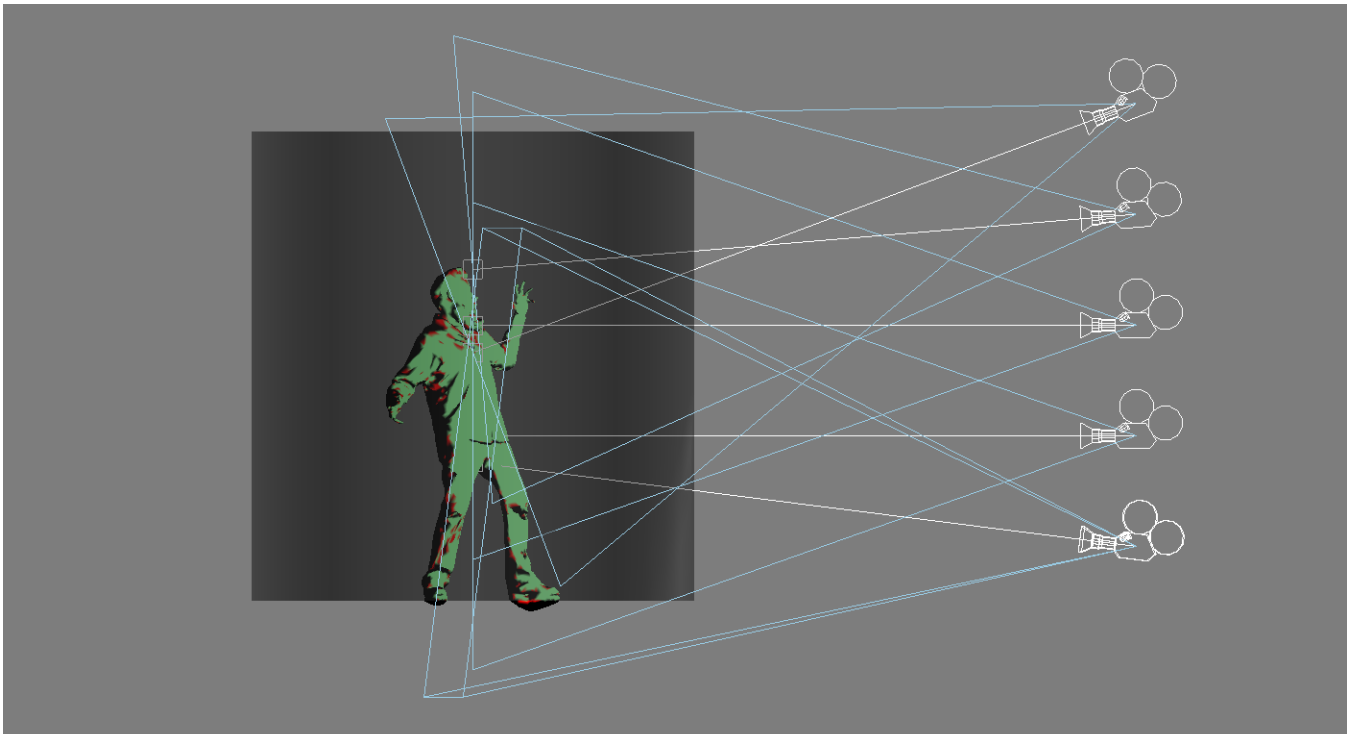


Figure 3: Combined frustum of the IR cameras of a single tower (analogous for the RGB cameras).

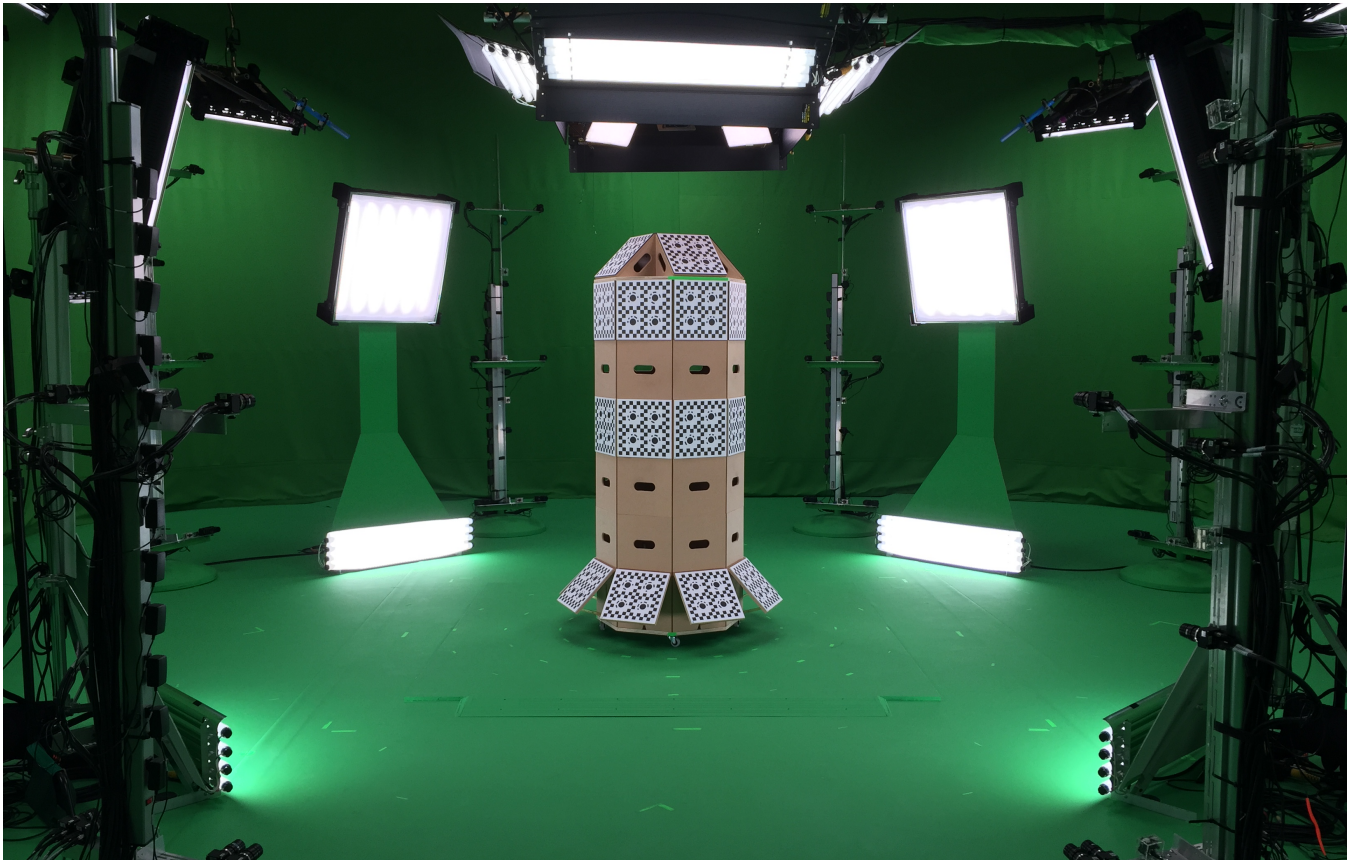


Figure 4: Calibration object (Octolith).



LG/LT



LG/HT



HG/LT



HG/HT

Figure 5: For Kendo dataset, comparison of geometry and texture quality for different camera configurations (Low and High number of cameras for both Geometry and Texture reconstruction).