KEYWORDS VIRTUAL REALITY // ORTHOSTEREOSCOPIC IMAGING // VIRTUAL REALITY FILM PRODUCTION PIPELINE // CINEMATOGRAPHY

Optimizing 180° Stereoscopic Cinematography for the Apple Vision Pro

By Stanley Lin, Nathan Fairchild, Spencer Lin, Akaash Tumuluri, Jordan Halsey, and Scott Fisher

Abstract

The Apple Vision Pro establishes a new benchmark for image fidelity in head-mounted displays, creating both opportunities and challenges for cinematic VR filmmaking. Producing content that fully leverages its display capabilities requires reconsideration of image acquisition, post-production, and delivery workflows. This paper presents a replicable 180° orthostereoscopic imaging pipeline designed to balance accessibility, reliability, and cinematic quality while meeting the demands of next-generation virtual and mixed reality displays. Using our experimental VR short film *A Larger Living Space* as a case study, we outline an end-to-end workflow that integrates structured capture techniques, AI-based image enhancement, and scalable visual effects methods. Our pipeline emphasizes preservation of fine detail, reduction of noise and artifacts, and seamless integration of practical and virtual lighting within wide fields of view offered by dual-fisheye lenses available on the market. By uniting prosumer-grade tools with scalable post-production methods, our approach demonstrates that high-fidelity immersive video can be achieved outside of large-scale studio contexts. This work contributes a consolidated framework for immersive cinematography and provides a foundation for future exploration into higher bit depth capture, high frame rate imaging, and real-time virtual set integration for VR film productions.

The Apple Vision Pro (AVP) introduces an unprecedented standard of image fidelity in head-mounted displays (HMDs), enabling highly immersive VR cinematic experiences. However, leveraging its full potential requires rethinking both acquisition and post-production workflows for stereoscopic content. In this paper, we present a novel imaging and post-production pipeline that addresses the core challenges of 180° stereoscopic cinematography optimized for AVP delivery while using our original VR short film, *A Larger Living Space*, as a case study **(Fig. 1).** Our approach integrates prosumer-grade camera systems with AI-enhanced post-processing techniques to achieve high-fidelity immersive video. This approach can be broken down into four primary stages: image acquisition, stereoscopic preprocessing and VFX compositing, AI-assisted image restoration and upscaling, and Apple Vision Pro delivery.

We begin by outlining an optimized imaging pipeline built around the Canon R5 C and the RF 5.2mm f/2.8 Dual Fisheye lens, henceforth referred to as the Camera and the Lens, respectively. While this system offers a complete and accessible orthostereoscopic imaging camera package with a 190° angle of view, it also has significant limitations, including suboptimal sensor noise performance, low optical resolving power, and framing/lighting constraints inherent to ultra-wide-angle-of-view (AOV) capture. To address these limitations, we implemented structured image acquisition strategies, including a one-stop overexposure method to improve sensor noise performance, aperture selection to balance resolving power and light transmission, and a hybrid lighting design that leveraged virtual ceiling extensions, embedded LED practicals, and rigged overhead space lights.

1545-0279/26©2026SMPTE

FIGURE 1. Color-graded frame from *A Larger Living Space*, showcasing the fully denoised and upscaled footage with a computer-generated ceiling seamlessly noise-matched to the live-action plate (2025).

Crucially, we developed a scalable post-production pipeline designed to meet the display demands of the AVP. After stereoscopic reprojection, we processed the footage through two discrete AI-assisted restoration stages in Topaz Video AI: an initial denoising pass using Nyx, followed by a separate Proteus enhancement and 2× upscale to a 16384 × 8192 stereo equirectangular master, corresponding to 8192 × 8192 pixels per eye. This increased equatorial sampling density from approximately 22.8 to 45.5 pixels per degree while preserving the native 59.94p temporal cadence. The higher-resolution reconstruction was intended to increase sampling density and improve the reproduction of edges and fine textures during headset presentation rather than to recover optical information absent from the original capture. This departs from legacy post-processing techniques like total variation and wavelet denoising, which overly degrade detail critical to the immersive experience. Final masters were graded for P3-D65 primaries with an ST 2084 (PQ) transfer function and encoded as MV-HEVC for stereoscopic playback on Apple Vision Pro.

We also developed a visual effects (VFX) and compositing pipeline using tools such as Houdini, 3D Equalizer, and Nuke, relying on Pixar USD for render-agnostic scene assembly. Our virtual set extensions (particularly for ceiling replacement) enabled sophisticated lighting design within the extreme AOV of the Lens without compromising the frame. The completed stereo master was encoded in MV-HEVC with embedded spatial metadata for stereoscopic playback on AVP.

This research provides a blueprint for high-fidelity immersive film production using accessible equipment and offers a comprehensive, documented technical workflow tuned for the Apple Vision Pro. Beyond technical enhancements, our imaging pipeline contributes meaningfully to immersive storytelling by supporting more cinematic lighting, refined viewer gaze control through luminance contrast, and more naturalistic motion presentation.

Our work aims to democratize high-quality VR filmmaking by making high-fidelity cinematic results attainable for lower-budget productions. We hope to spark further exploration into 180° stereoscopic content design, enabling independent creators to meet emerging standards in XR media consumption. Future work will explore higher-bit-depth acquisition, 120p capture, and real-time motion-tracked VFX set extensions using camera telemetry data for accurate CG scene integration and single- or multi-axis camera movements.

Inspirations and Related Works

Recently, interest in stereoscopic VR capture systems has increased in both commercial and academic domains. Camera systems such as the KanDao QooCam EGO, Z CAM K1 Pro, and the upcoming Blackmagic URSA Cine Immersive each

target immersive imaging with varying trade-offs in resolution, affordability, and workflow maturity. The Blackmagic URSA Cine Immersive, in particular, is positioned as the most complete high-end solution: bundled with DaVinci Resolve Studio, it offers a highly integrated camera-to-headset workflow. However, this functionality is only accessible through the purchase of the camera itself, which retails for $29,995 USD,[1] effectively limiting its availability to creative teams with substantial budgets.

Prior academic work has largely focused on isolated aspects of immersive imaging rather than integrated workflows. Studies and articles have examined the relationship between frame rate and simulator sickness,[2] evaluated noise-reduction algorithms,[3] and benchmarked individual camera functionality.[4] Similarly, industry documentation from manufacturers provides specifications and operational guidelines but rarely addresses how to deploy these devices systematically to achieve cinematic imagery in VR contexts.

To verify this gap, we conducted a scoping review of immersive imaging literature between 2020–2025, searching multiple electronic databases with terms including stereoscopic video, immersive video, and VR cinematography. While we identified research on 360° video and volumetric capture pipelines, we found no studies that consolidated best practices into an end-to-end 180° orthostereoscopic imaging workflow suitable for modern head-mounted displays.

Our contribution addresses this absence. Unlike existing documentation, this work is specifically optimized for the Apple Vision Pro, whose 34 PPD acuity and high dynamic range (HDR) display impose stricter sharpness and noise thresholds than traditional cinema. The result is a practical framework enabling filmmakers and creatives to achieve high-fidelity immersive imagery from prosumer stereoscopic hardware at a fraction of the cost of high-end systems like the URSA Cine Immersive coupled with its immersive media post-production suite.

Imaging Hardware and Production Challenges

In VR cinematography, our camera system has limitations stemming from the Camera's sensor and processing characteristics and the Lens's optical design. The most significant production challenges included the Camera's suboptimal noise performance and limited dynamic range, the Lens's poor optical resolving power, and the Lens's extreme AOV, which introduced substantial framing and lighting constraints.

Camera System

We selected the Camera and Lens based on the following considerations:

- Native RF mount compatibility with full lens-to-body communication.
- Full-frame image format and 8192 × 4320 (8K) sensor resolution, providing sufficient image detail for further AI resolution upscaling and denoising.
- High frame rate (59.94p) support at full resolution and image format, ensuring smooth motion continuity for HMD playback.
- 190° AOV, producing a high degree of VR immersion when viewing footage in HMDs.

Sensor and Recording Limitations

The Camera's 8K, 36 × 24 mm full-frame sensor provides an acquisition resolution sufficient for subsequent AI-based upscaling and denoising processes. When paired with the dual fisheye Lens, the single 8K sensor is divided into two circular fisheye images corresponding to left- and right-eye views, resulting in an effective image diameter of approximately 3684 pixels per eye.[5] In comparison, the Apple Vision Pro (AVP) renders at 3660 × 3220 pixels per eye[6] with a 100° AOV,[7] substantially narrower than the 190° capture AOV of our imaging system. Because the AVP displays only a 100° subset of the original 190° capture at any given moment, the effective resolution per displayed pixel is reduced relative to our full-frame acquisition. This spatial "cropping" effect necessitates further post-production enhancement, including AI-driven upscaling, to ensure that the displayed imagery fully exploits the native per-eye resolution of the AVP and maintains perceptual fidelity within the HMD.

Following stereoscopic reprojection, each 180° equirectangular eye view measured 4096 × 4096 pixels, corresponding to approximately 22.8 pixels per degree at the equator. The subsequent 2× enhancement and upscale in Topaz Video AI produced an 8192 × 8192 image for each eye, or a combined stereo raster of 16384 × 8192 pixels, increasing equatorial sampling density to approximately 45.5 pixels per degree. This increase in sampling density provides a denser reconstructed raster for subsequent compression and high-resolution headset presentation.

High recording frame rates are essential for preserving motion continuity in VR, thereby reducing user discomfort and maintaining perceptual immersion. The AVP supports playback at frame rate multiples of 24p and 30p up to 100p, enabling judder-free rendering when source material is properly matched. Conventional cinematic frame rates such as 24p or 25p, while aesthetically established in traditional filmmaking, have been demonstrated to be inadequate in VR contexts, where sub-60p acquisition may introduce perceptible judder and simulator sickness (SS).[2] For this reason, principal photography was conducted at the Camera's maximum supported frame rate of 59.94p (60p) at full sensor resolution, recorded in Canon 12-bit Cinema RAW Light.

Using the Camera's full 8K 60p recording capabilities requires Canon's 12-bit Cinema RAW Light format, which omits in-camera noise reduction. Consequently, RAW footage exhibited elevated sensor noise, particularly within shadow regions.[4] These limitations were compounded by the project's low-light, gritty apartment setting. In an effort to mitigate these issues, Canon Log 3 encoding was selected for its lower noise floor, improved shadow handling, and enhanced color retention relative to other log profiles.[8] Nevertheless, even at the Camera's native ISO settings of 800 and 3200, the resulting footage displayed significant noise in dark regions and susceptibility to highlight clipping, as illustrated in the left panel of **Fig. 2.**

1545-0279/26©2026SMPTE

FIGURE 2. Side-by-side comparison of unprocessed RAW footage (left) and AI-denoised, upscaled footage (right) (2025).

Optical Characteristics and Limitations of the Lens

The Lens's 190° AOV and 60 mm interaxial distance produce a stereoscopic pair whose capture baseline approximates common human interpupillary distances. In this paper, we use orthostereoscopic imaging in its established stereoscopic-display sense: capture and presentation geometry intended to preserve naturalistic scale, perspective, and binocular depth relationships.[9–11] The term has a longstanding lineage in stereoscopic cinema, interactive display, telepresence, and virtual-reality research. A human-scale interaxial distance alone does not guarantee perfect orthostereoscopy, however, as perceived stereo geometry also depends on lens projection, reprojection, display field of view, headset optics, and viewer IPD.

Optical deficiencies in the Lens further degraded image quality. Because we needed to maximize light transmission to combat sensor noise, our initial approach prioritized shooting at the Lens's widest aperture (f/2.8). However, testing revealed significant degradation in resolving power at this aperture, leading to unacceptable reductions in image sharpness.

Image Acquisition Optimizations

To address these challenges systematically, we structured our image acquisition pipeline to improve three critical areas of capture: exposure, optics, and lighting.

Exposure and Noise Mitigation Strategies

To compensate for the Camera's noise performance and limited dynamic range, we established a recording configuration optimized for color retention and noise reduction:

- Recording Format: Canon 12-bit Cinema RAW Light
- Resolution: 8192 × 4320
- ISO: 3200 (base 3200)
- Color Profile: Canon Log 3
- Frame Rate: 59.94p

Furthermore, we implemented a one-stop overexposure strategy[8] to retain shadow detail, reduce perceptible noise, and protect highlights. This approach involved systematically exposing 1 stop above our target exposure while monitoring footage through a custom pull-down lookup table (LUT) designed to simulate the correct exposure and intended final look in real time.

We standardized all shots at f/5.6, exposing for f/8 via a 1-stop overexposure method. This approach enabled the sensor to capture increased data in shadow regions while protecting highlights, effectively reducing noise accumulation in underexposed areas.

Optimizing Optical Performance of the Lens

Careful aperture selection was a critical factor in balancing optical resolving power against light transmission. Empirical testing demonstrated the following:

- f/2.8 exhibited significant optical softness, making it unsuitable despite its increased light transmission.
- f/7.1 yielded the highest resolving power, albeit still exhibiting suboptimal sharpness, but at the cost of reduced light transmission.
- f/5.6 provided an optimal compromise, offering an acceptable level of sharpness while securing an additional 2/3 stop of light transmission compared to f/7.1.

Lighting and Framing Adaptations for Ultra-Wide AOV

The ultra-wide AOV of the Lens meant traditional lighting setups had to be moved out of frame or hidden. This increased the distance between fixtures and the illuminated area and/or required smaller, more discrete fixtures. This reduced the total light reaching the sensor from the scene,

FIGURE 3. Frame grab illustrating the set constructed without a ceiling to facilitate overhead lighting (2025).

FIGURE 4. Lighting diagram illustrating a set layout with lighting fixture, diffusion, and bounce placements (2025).

thereby increasing sensor noise. To circumvent these limitations, we employed a hybrid practical-lighting and computer-generated production-design strategy.

Set Design for Optimal Fixture Placement

Our apartment set was designed with doorways and structural corners, allowing conventional off-axis lighting setups to create strong accents and hidden hair lights without compromising framing, as seen in **Fig. 3.** This layered lighting approach allowed us to preserve the scene's visual integrity while using luminance contrast to guide viewer attention to the areas of the 180° frame that matter most to the narrative. Lighting subtly directs the gaze without compromising the viewer's freedom of exploration. We balanced interactive storytelling with cinematic presentation, motivating the viewer to follow the narrative without overt guidance.

Virtual Set Extensions to Motivate Overhead Lighting

To prepare for virtual ceiling set extensions, we built our set on a soundstage without a ceiling, as shown in **Fig. 3.** This allowed us to rig fixtures on the soundstage's overhead lighting grid to provide ambient fill and hard spots into the set. We composited a virtual ceiling into the set in post-production, maintaining the illusion of an enclosed environment. Our

1545-0279/26©2026SMPTE

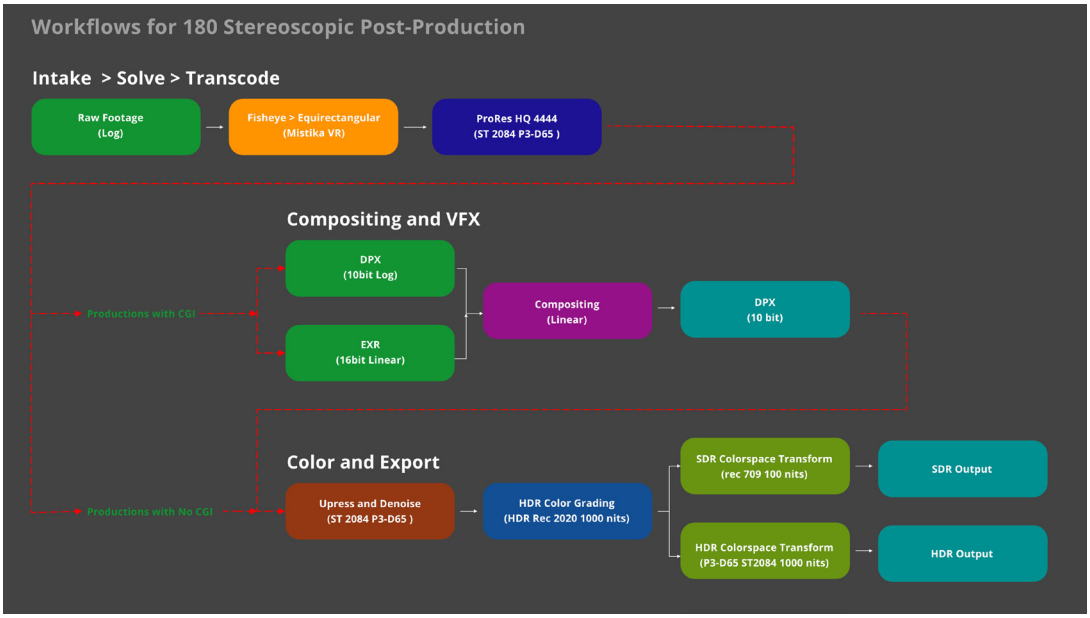


FIGURE 5. Our digital intermediate pipeline flow chart detailing workflows for both CGI and non-CGI projects (2025).

overhead lights were motivated by the house lights on the virtual ceiling, ensuring spatial consistency between practical and CG elements. This ceiling-less set design was our primary method for achieving our target exposure levels while hiding all large lighting fixtures.

Embedded Practical LED Fixtures

Discrete, high-output, high-color-quality LED fixtures were hidden within practical set decoration pieces such as lamps and magnetically mounted to the walls out of frame to provide subtle fill and hair lighting (refer to "B7C" and "MC Pro" in **Fig. 4**), further enhancing depth and separation between subjects and the background.

Post-Production Workflow

While our acquisition pipeline mitigates many imaging challenges, its most important role is to prepare our footage for post-production. RAW footage directly from the camera still exhibits noticeable noise and optical resolution deficiencies. The final refinement of our footage into a polished VR film is ultimately achieved through our post-production pipeline.

Digital Intermediate Pipeline

We imported Canon Cinema RAW Light footage into Mistika VR and decoded it using Canon Log 3 and Canon Cinema Gamut before stereoscopic reprojection. Following correction and reprojection, the imagery was exported using Rec. 2020 primaries with an ST 2084 (PQ) transfer function. This Rec. 2020/PQ interchange configuration was selected based on Canon workflow guidance to preserve the wide-gamut and HDR image information carried from the RAW source through subsequent post-production stages. P3 therefore describes a downstream mastering and delivery target rather than the native acquisition or intermediate color space.

For footage requiring no CGI integration, the Rec. 2020/PQ material was imported into DaVinci Resolve and transformed into DaVinci Wide Gamut with DaVinci Intermediate for color grading. Footage requiring CGI integration instead entered the ACES-managed compositing workflow described in subsequent sections before returning to the same grading and mastering pipeline. Final imagery was transformed to P3-D65 primaries with an ST 2084 (PQ) transfer function for Apple Vision Pro delivery.

When this production was completed, the required MV-HEVC export workflow was not yet available directly in DaVinci Resolve. We therefore converted the completed P3-D65/PQ stereo master externally using FFmpeg for MV-HEVC delivery. **Figure 5** illustrates both the CGI and non-CGI color-management workflows.

Strengths and Weaknesses of the Camera System for Our Post-Production Pipeline

Canon provides proprietary software for converting dual-fisheye imagery into stereoscopic 180° equirectangular format. However, in our testing, the software did not resolve imagery to an acceptable standard for research or production use. As a result, our workflow employed Mistika VR for this stage of the pipeline. Through camera-lens profiling conducted within Mistika VR, we identified two strengths and five primary weaknesses of our camera system:

- Camera system strengths:
  - ◊ Ability to shoot RAW at 60 frames per second.
  - ◊ A total capture resolution of 8192 × 4320.
- Camera system weaknesses:
  - ◊ Off-centered fisheye images projected from the lens onto the sensor, as seen in **Fig. 6**.
  - ◊ Low optical resolving power of the lens.
  - ◊ Unknown exact IPD.
  - ◊ Suboptimal noise performance, especially in low-light scenarios.

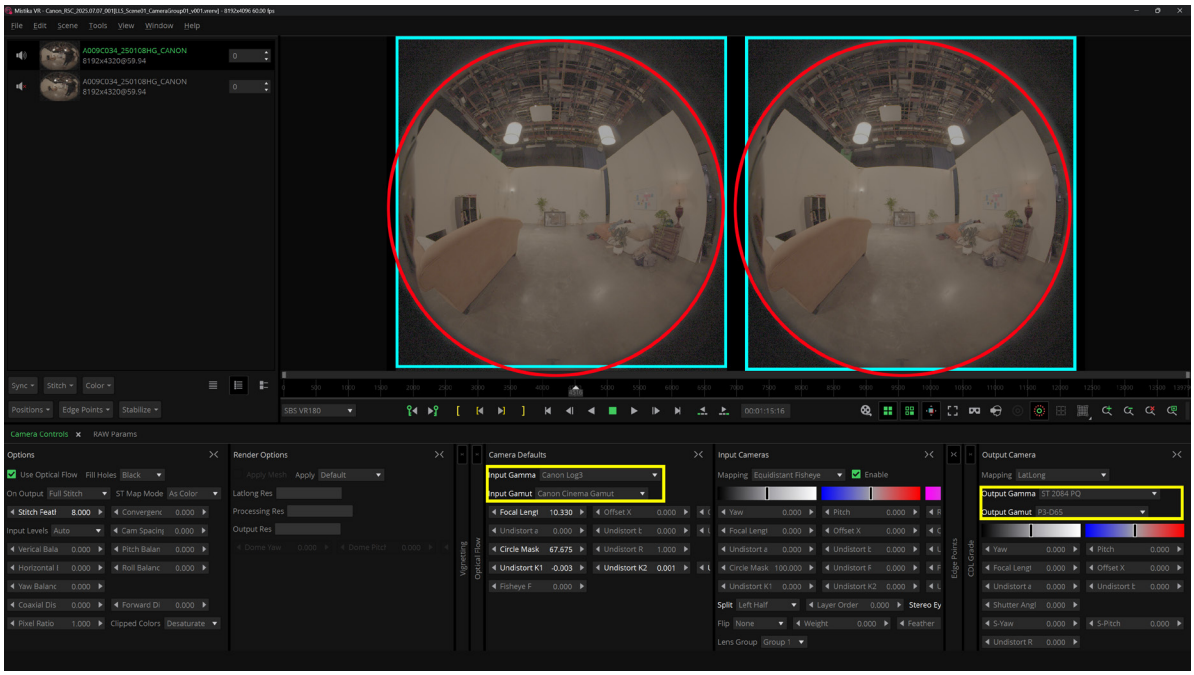


FIGURE 6. A diagram illustrating off-center fisheye images resulting from the camera and lens (2025).

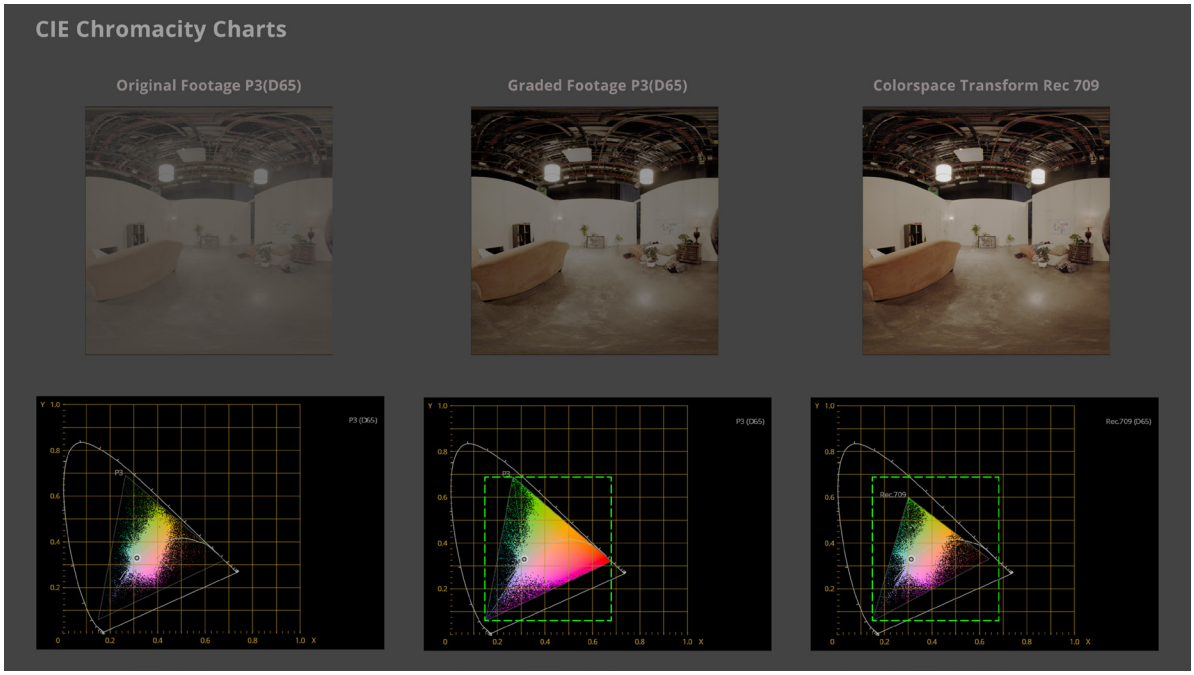


FIGURE 7. Comparison between ungraded footage, graded footage within the P3 color space, and a color space transform to Rec. 709 Gamma 2.4 (2025).

◊ Focus inconsistency between the two fisheye projections, resulting from each lens group needing to be independently focused before each shot.

Camera System Testing Methodology

Because of the Lens's extreme AOV and distortion characteristics, a standard lens grid calibration shoot was insufficient. Using the software 3D Equalizer, we determined that a more rigorous approach involving a detailed series of tests supplemented with high-precision photogrammetry was required.

Photogrammetry

Conventional grid-based calibration proved inadequate for the extreme distortion characteristics of the fisheye optics. To establish a more rigorous spatial reference, we combined 3D Equalizer lens and camera analysis with a high-precision photogrammetric reconstruction of the practical set. We incorporated custom marker sets into the RealityCapture reconstruction, while a Leica Disto site survey provided independently measured physical dimensions to constrain global scale and orientation. Integrating these surveyed measure-

1545-0279/26©2026SMPTE

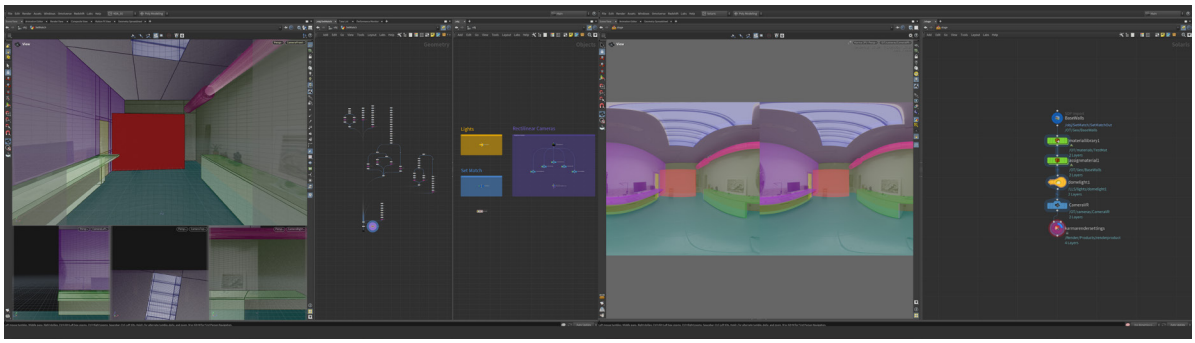


FIGURE 8. 3D viewport used for alignment and the stereo renders over the backplate (2025).

ments prevented the photogrammetric reconstruction from relying on image-derived scale alone.

The resulting photogrammetric geometry served as a spatial reference throughout camera validation and CGI integration. Initial base geometry derived from on-set measurements was aligned to the reconstructed environment and subsequently rendered against the stereo 180° equirectangular backplates. This provided visual feedback for evaluating correspondence between the known physical geometry, the photogrammetric reconstruction, and the stereoscopic plate. Once the scan had been oriented to the base geometry, no subsequent scale adjustment was applied. The calibrated environment was then carried forward into Houdini, where the modular set geometry and custom alignment tools described later in the Section, *A Larger Living Space—Pipeline Integration Example*, provided an additional means of validating geometric correspondence.

Footage Assets

To create a full lens profile for the Lens, we determined that we would shoot a series of tests to incorporate into 3D Equalizer. This would include a typical center grid lens shot with a static camera, a series of nodal camera pans for matchmoving, and a gimbaled handheld tracking shot. All of these would keep the main lens grid in full view.

Noise Testing

To evaluate noise performance, we tested multiple exposure strategies at both base ISO 800 and base ISO 3200. For each setting, exposures ranged from one stop under to two stops over. We ultimately observed the best noise performance at one- and two-stop overexposure.

Focus Testing

We shot a series of tests with incremental focus ranges starting at 9 in. and increasing a foot at a time to determine the best focal point for deep focus.

Footage Ingest, Transcode, Edit, and Output

We ingested the captured RAW footage through Mistika and split it into left- and right-eye streams. Using Mistika's lens guides, the off-center fisheyes were corrected, and an initial geometric solve was calculated. Following stereoscopic reprojection, the corrected imagery was exported using Rec. 2020 primaries with an ST 2084 (PQ) transfer function and subsequently imported into DaVinci Resolve. Within Resolve, the footage was transformed into DaVinci Wide Gamut and DaVinci Intermediate for grading before final transformation to P3-D65 primaries with an ST 2084 (PQ) transfer function for Apple Vision Pro delivery. A comparison between the graded P3 output and a Rec. 709 Gamma 2.4 transform is shown in **Fig. 7.**

CGI Integration Testing

Plates were processed for compositing as 10-bit DPX files. For scene alignment, we mapped the equirectangular plate to an inverted hemisphere oriented around the Z-Axis with five cameras locked at the zero position. We created five 90º rectilinear cameras oriented +X, -X, +Y, -Y, and +Z, with a square filmback (24mm x 24mm).

Camera Validation—Testing Geometry Alignment

We validated scene alignment using baseline measurements of camera height and on-set location data. We then aligned the initial geometry with the photogrammetry model.

To refine alignment, we rendered the base 3D geometry against the stereo backplates, as shown in **Fig. 8**. Comparing stereo 180° equirectangular renders provided feedback for iterative geometry refinement.

Once we refined the 3D base model, we imported and oriented the 3D photogrammetry scan, ensuring no scale adjustments, as shown in **Fig. 9.**

AI-Assisted Denoising and Upscaling

Following compositing, the completed stereo imagery was exported as ProRes 4444 XQ and processed in Topaz Video AI as two discrete restoration passes. A dedicated denoising pass using the Nyx model was performed first, followed by a separate enhancement and 2× upscaling pass using Proteus. The latter increased each 180° equirectangular eye view from 4096 × 4096 to 8192 × 8192 pixels, producing a final side-by-side stereo raster of 16384 × 8192 pixels. At the equator of the equirectangular projection, this increased sampling density from approximately 22.8 to 45.5 pixels per degree while preserving the original 59.94p temporal cadence.

Proteus was operated in manual parameter mode rather than Auto, allowing the Recover Detail, Sharpen, Dehalo, and Anti-Alias controls to be explicitly configured and held con-

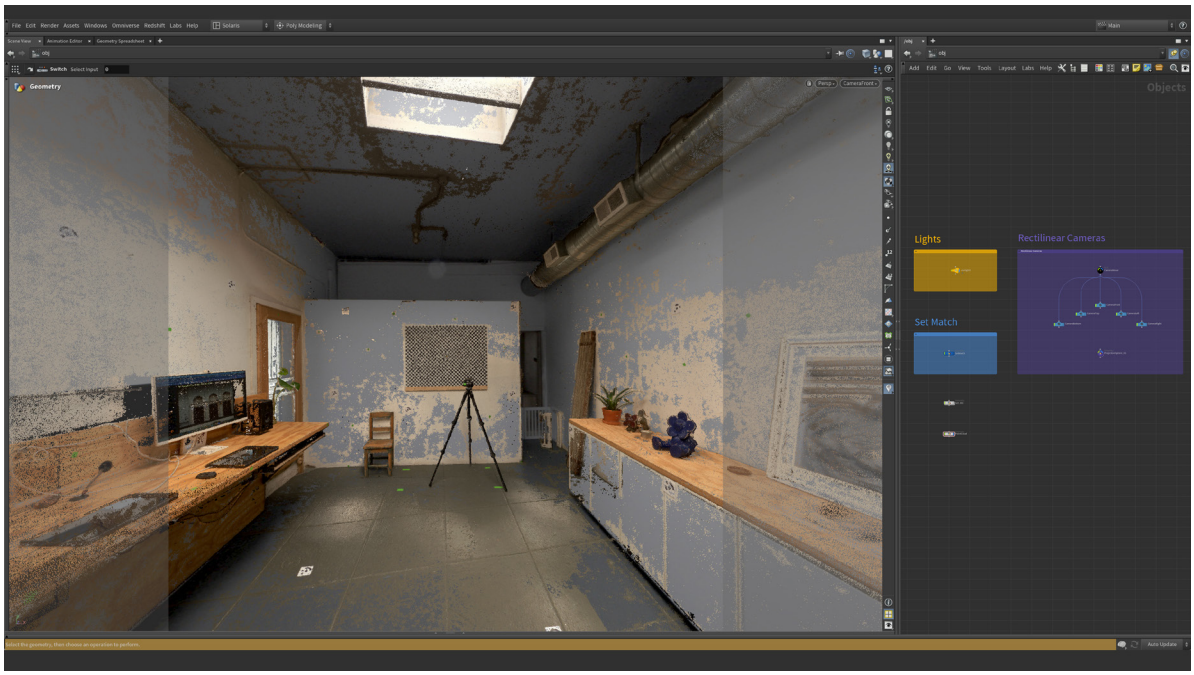


FIGURE 9. 3D viewport showing the aligned photogrammetry with the backplate (2025).

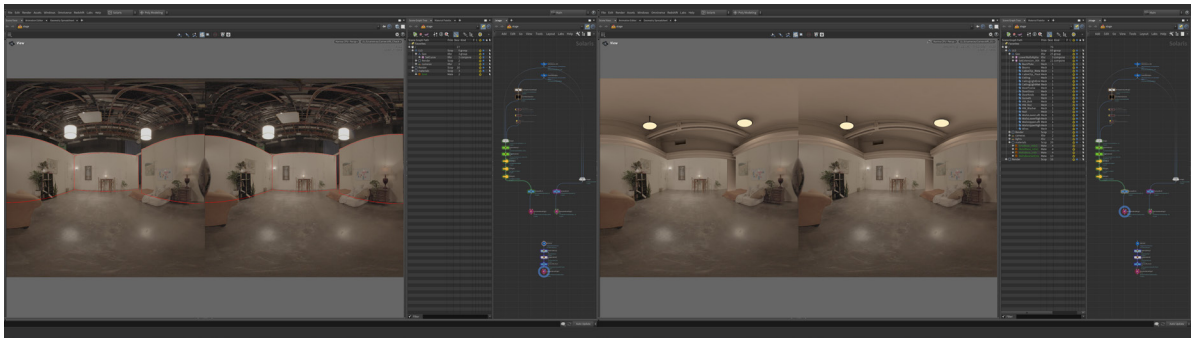


FIGURE 10. Renders showing the alignment tool and the render over the backplate (2025).

stant across the production. Auto mode's per-shot parameter estimation was rejected because shot-to-shot variation could produce a non-uniform rendition of detail and edge structure across cuts and, more critically for stereoscopic imagery, could result in divergent processing between corresponding left- and right-eye views. Maintaining a consistent parameter configuration across both eyes and throughout the production reduced the likelihood that the enhancement itself would introduce binocularly inconsistent high-frequency detail.

The ordering of the restoration stages was likewise deliberate. Denoising was performed before enhancement and upscaling because enlarging imagery before noise removal can promote stochastic sensor noise into more coherent spatial structures. A subsequent restoration or enhancement stage may then interpret and preserve those structures as image content rather than remove them as noise. Performing the Nyx denoising pass first therefore provided Proteus with a cleaner source image from which to reconstruct edges and fine textures.

Traditional mathematical denoising techniques, such as Total variation regularization and wavelet-based noise suppression, can reduce image noise at the expense of edge definition and high-frequency texture.[3] This trade-off becomes particularly visible during close HMD viewing. Neural video-denoising approaches can instead use spatial and temporal image structure when differentiating persistent scene detail from stochastic noise.[12] For this production, the objective of AI-assisted restoration was therefore not maximum noise removal, but sufficient suppression of sensor noise while retaining perceptually important fine structure for stereoscopic headset presentation.

The resulting 16384 × 8192 image should be understood as a higher-density reconstruction rather than as a native 16K optical capture. The resolving power of the original Camera and Lens remains the upper bound on captured scene information. The 2× upscale instead provides a denser sampling grid for subsequent compression and headset presentation without implying a corresponding doubling of captured optical resolution.

Apple Vision Pro Delivery and MV-HEVC Encoding

Following restoration, upscaling, and final color mastering, the completed stereoscopic master was encoded for Apple

1545-0279/26©2026SMPTE

FIGURE 11. Nuke compositing network (2025).

Vision Pro playback using Multiview High Efficiency Video Coding (MV-HEVC). Unlike conventional side-by-side stereo, in which both eye views occupy different portions of a single video frame, MV-HEVC represents the left- and right-eye images as separate layers within a single video track. The layers share a common frame size, allowing each eye to remain discrete during decoding while the codec exploits the substantial similarity between the two views to reduce redundant image information.[13]

When this production was completed, direct MV-HEVC export was not yet available in DaVinci Resolve. The final P3-D65/ST 2084 PQ stereo master was therefore converted externally using FFmpeg. The left- and right-eye imagery was encoded into the corresponding multiview layers, and the required spatial metadata was embedded in the resulting delivery file for correct stereoscopic interpretation on Apple Vision Pro. The completed master was subsequently reviewed natively in the headset to verify correct eye assignment, projection geometry, motion cadence, color reproduction, compression behavior, and overall binocular presentation.

A Larger Living Space—Pipeline Integration Example

To ensure our physical environment could be reproduced digitally, we built our practical apartment set using a modular flat system. All walls were fabricated from standardized 4 × 10 foot panels, which were subsequently replicated in a three-dimensional environment. These virtual modules served as a registration framework, aligned precisely with a base geometric model derived from on-set measurements. A custom Houdini Digital Asset (HDA) was developed to enforce consistent alignment across shots. The HDA also enabled the rendering of polylines over stereo equirectangular backplates, providing a mechanism to validate geometric correspondence and inter-pupillary distance (IPD) accuracy (alignment tool shown in **Fig. 10**).

The base geometry was used as a structural guide for set extensions. These extensions were first rendered as a quality-control pass against the original backplate, then used as matte cut-outs to support compositing operations. An occlusion pass was generated in parallel to improve the perceptual blending of digital extensions with the live-action plate. The set extensions were rendered within a Universal Scene Description (USD) pipeline, using Solaris and the Karma renderer with a MaterialX shading framework. The same USD-based workflow also facilitated the production of supplementary computer-generated imagery (CGI) in Unreal Engine, ensuring interoperability across platforms.

All 3D outputs were generated as multilayer, linear EXR files, incorporating the following render passes: beauty, combined albedo, combined reflection, combined refraction, emission, position, normals, UVs, and a Fresnel component. These passes provided the necessary information to support advanced compositing operations. To preserve the raw fidelity of the renders, no denoising was applied during the rendering process.

Compositing was completed in Nuke with a procedural network, shown in **Fig. 11**, that allows for work to be easily propagated across multiple shots. The 3D renders are grain- and grade-matched against the uncorrected footage, and grain is matched again in the compositing phase to ensure as close a match as possible before the denoise and uprez phase.

Although many components of this workflow can be generalized to other stereoscopic HMDs, several parameter choices were driven specifically by the Apple Vision Pro presentation. These include high-resolution reconstruction to compensate for the disparity between capture and display sampling den-

sity, 59.94 fps acquisition to maximize temporal resolution within the Camera's full-resolution recording limitations, HDR grading with a P3-D65/ST 2084 PQ delivery target, MV-HEVC stereoscopic encoding, and final quality-control evaluation within the headset itself. The resulting workflow should therefore be understood not simply as a general stereoscopic post-production pipeline, but as an accessible capture-to-display methodology whose image-processing and delivery decisions were evaluated against the perceptual demands and media architecture of Apple Vision Pro.

Future Work

Significant work remains in advancing techniques and technologies for immersive cinematography. Future research can investigate imaging and processing approaches that extend dynamic range through higher bit-depth acquisition (e.g., 16-bit RAW workflows). Such methods would improve the ability to capture low-light scenes without relying on deliberate overexposure followed by post-production enhancements. As stereoscopic camera systems continue to evolve, further research into higher frame rate capture (≥120p) will be important for minimizing motion discontinuity and reducing simulator sickness. In addition, integrating motion-tracked VFX set extensions with externally recorded camera telemetry offers a pathway to more accurate alignment of computer-generated elements with physical lighting fixtures and camera movement.

Conclusion

The insights gained from *A Larger Living Space* informed the development of a capture-to-delivery workflow optimized around the image-quality and playback requirements of Apple Vision Pro. Although many of its underlying methods can be generalized to other stereoscopic cameras and HMDs, specific acquisition, reconstruction, color-management, and encoding decisions were selected and evaluated against the characteristics of the Apple Vision Pro display and visionOS media pipeline. By combining established filmmaking practices with emerging VR and AI tools, we demonstrate a reproducible approach to high-fidelity 180° stereoscopic cinematography using comparatively accessible production hardware. Our acquisition approach (integrating deliberate overexposure, lensing trade-offs, and adapted lighting strategies) produced footage optimized for subsequent AI-driven denoising and 16K upscaling, allowing us to retain fine detail and minimize noise artifacts. Recording at higher frame rates further enhanced motion realism and viewer comfort. Overall, our findings suggest that immersive VR films can maintain the artistic intent and visual language of traditional cinema while meeting the technical requirements of next-generation HMDs, providing a practical foundation for both filmmakers and researchers advancing cinematic storytelling in VR.

References

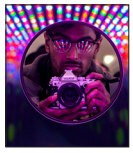
1. B. Allan, "Blackmagic URSA Cine Immersive Camera—Measuring Resolution," *Newsshooter*, 2025. [Online]. Available: www.newsshooter.com.
2. J. Wang et al., "Effect of Frame Rate on User Experience, Performance, and Simulator Sickness in Virtual Reality," *IEEE Trans. Vis. Comput. Graph.*, 29(5):2478–2488, 2023. doi:10.1109/TVCG.2023.3247057.
3. Q. Yang et al., "Low Light Image Denoising Based on Poisson Noise Model and Weighted TV Regularization," *Proc. IEEE Int. Conf. Image Process.*, 3199–3203, 2018. doi:10.1109/ICIP.2018.8451840.
4. J. Behiri, "Canon EOS R5 C Lab Test—Rolling Shutter, Dynamic Range and Latitude," *CineD,* 2023. [Online]. Available: www.cined.com.
5. Canon Inc., "Canon RF 5.2mm F2.8 L Dual Fisheye," Canon Canada, 2025. [Online]. Available: www.canon.ca.
6. D. Heaney, "Apple Vision Pro Extended Teardown Reveals Its True Active Resolution," *UploadVR*, 2024. [Online]. Available: www.uploadvr.com.
7. Nivsus, "Apple Vision Pro FOV: Limitations and the Quest for Enhancement," *vskel*, 2024. [Online]. Available: vskel.com.
8. Canon Europe, "Enhance Filmmaking with Canon Log," Canon Europe, 2023. [Online]. Available: www.canon-europe.com.
9. M. T. Bolas and S. S. Fisher, "Head-Coupled Remote Stereoscopic Camera System for Telepresence Applications," *Proc. SPIE*, 1256:113–123, 1990. doi:10.1117/12.19896.
10. L. Lipton, *Foundations of the Stereoscopic Cinema: A Study in Depth*, Van Nostrand Reinhold: New York, 1982.
11. W. Robinett and J. P. Rolland, "A Computational Model for the Stereoscopic Optics of a Head-Mounted Display," *Presence*, 1(1):45–62, 1992. doi:10.1162/pres.1992.1.1.45.
12. M. Tassano, J. Delon, and T. Veit, "DVDNET: A Fast Network for Deep Video Denoising," *Proc. IEEE Int. Conf. Image Process.*, 1805–1809, 2019. doi:10.1109/ICIP.2019.8803136.
13. Apple Inc., "Converting Side-by-Side 3D Video to Multiview HEVC and Spatial Video," *Apple Developer Documentation.* [Online]. Available: developer.apple.com.

About the Authors

Stanley Lin is a creative technology researcher at USC Ganek Immersive Studio and Visiting Scholar at USC MEML.

Nathan Fairchild is an immersive filmmaker, currently working as assistant director of USC School of Cinematic Arts' Ganek Immersive Studio.

Spencer Lin is an AI & XR researcher focused on augmenting storytelling through AI characters in immersive mediums.



Akaash Tumuluri is a writer, director, and producer, currently serving as lead producer for the Ganek Immersive Studio at USC.

Jordan Reece Halsey is the director of USC's Ganek Immersive Studio and a media artist specializing in immersive technologies and simulation.

Scott S. Fisher is a professor at USC School of Cinematic Arts and Director of USC's Mobile & Environmental Media Lab.

Presented at the SMPTE 2025 Media Technology Summit, Pasadena, CA, 16-19 Oct. 2025.

Copyright © 2026 by SMPTE.

DOI: 10.5594/JMI.2026/YCNL3434
Date of publication: 1 October 2026