Title:
Neural VR: Transforming AI-Generated Panoramas into Explorable Interior Space via Neural Scene Reconstruction

Authors:
Chieh-Jen Lin, t60011@mail.tut.edu.tw, Tainan University of Technology

Keywords:
Neural Scene Reconstruction, Virtual Reality, AI-Generated Panorama, Point Cloud Segmentation, Semantic Transfer, Explorable 3D Environment, Interior Design Visualization

DOI: 10.14733/cadconfP.2026.44-49

Introduction:

Virtual Reality (VR) offers unparalleled spatial immersion for interior design communication, enabling designers and clients to experience proposed spaces at the human scale before construction. However, creating VR-ready 3D environments traditionally demands extensive modeling effort—detailed geometry, material assignments, and lighting configurations—that conflicts with the rapid iteration required during early design stages. Recent advances in generative artificial intelligence have partially addressed this bottleneck: text-to-image diffusion models such as Stable Diffusion [8], combined with conditioning mechanisms like ControlNet [10], can now produce photorealistic 360-degree visualizations from simplified parametric models within seconds [4]. Yet these AI-generated panoramas remain fundamentally static, offering only fixed-viewpoint experiences that fall short of the spatial exploration VR promises.

The gap between AI-generated panoramic imagery and explorable VR environments represents a critical limitation in current design workflows. While panoramas can be displayed in VR headsets, users remain anchored to a single viewpoint, unable to walk through spaces, examine objects from multiple angles, or verify spatial relationships through natural movement. Although single-image 3D reconstruction tools such as TripoSR [7], InstantMesh [9], and Hunyuan3D [11] can generate geometric meshes from individual object photographs, they cannot reconstruct entire interior scenes with consistent geometry from a single panoramic image. True VR exploration requires volumetric 3D representation—geometry that persists as viewpoints shift and parallax reveals depth relationships.

Neural scene reconstruction techniques, including Neural Radiance Fields (NeRF) [6] and 3D Gaussian Splatting [1], have demonstrated remarkable capabilities in synthesizing novel views from image collections. These methods learn continuous volumetric representations that support free-viewpoint rendering, theoretically enabling the transformation of 2D panoramas into explorable 3D spaces. However, directly applying neural reconstruction to AI-generated panoramas presents unique challenges: single-viewpoint input provides limited geometric constraints, semantic boundaries may not align with learned density fields, and reconstruction artifacts can undermine the spatial coherence essential for convincing VR experiences.

This paper proposes Neural VR, a workflow that transforms AI-generated panoramas into semantically organized, explorable 3D environments suitable for immersive design evaluation. By combining perspective extraction, neural point cloud reconstruction, semantic segmentation transfer, and constraint-aware scene assembly, the approach bridges generative visualization and interactive VR, enabling designers to leverage AI's creative capacity while recovering the spatial navigability that defines meaningful virtual reality experiences. The proposed workflow has been implemented as an integrated toolset combining Grasshopper in Rhino 8, ComfyUI, and Twinmotion, and tested on multiple interior configurations representing typical Taiwanese apartment layouts. This paper reports both the conceptual framework and the empirical evaluation of the resulting VR-ready environments.

Main Ideas:

The fundamental contribution of Neural VR is a complete pipeline that converts static AI-generated panoramas into explorable 3D environments with preserved semantic organization. The workflow addresses the core tension between generative AI's image-based output and VR's requirement for volumetric geometry through a multi-stage transformation process.

Geometric Foundation from Parametric Models.

The low-LOD parametric model that guides AI panorama generation simultaneously provides critical geometric scaffolding for neural scene reconstruction. The model defines room boundaries (walls, floor, ceiling), furniture bounding volumes, and semantic category labels aligned with the ADE20K taxonomy. When the AI-generated panorama is back-projected onto this geometry, as established in prior workflow stages [9], a coarse depth map emerges that can initialize the spatial distribution of 3D Gaussians or provide depth supervision for NeRF training. These geometric priors address a fundamental challenge in single-panorama reconstruction: disambiguating the inherent scale and depth ambiguity present in spherical projections. The parametric model anchors the reconstruction in metric space, ensuring that furniture dimensions, room proportions, and circulation clearances remain consistent with design intent rather than being arbitrarily inferred by the neural network.

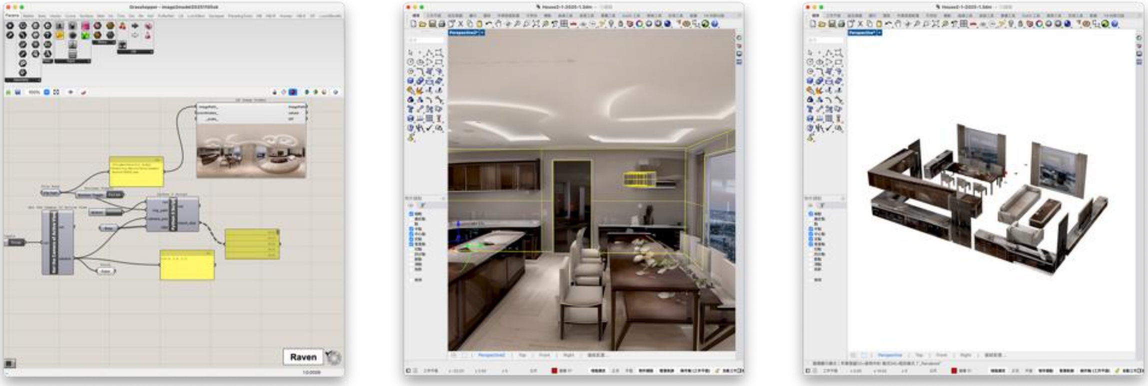

Fig. 1: Panoramic back-projection onto low-LOD parametric models: (a) Grasshopper scripts for panoramic back-projection, (b) Panoramic back-projection on Low-LOD models, (c) Low-LOD objects with reverse-projected textures providing depth priors for neural scene reconstruction.

From Panorama to Perspective: Extracting Reconstructable Views.

Equirectangular panoramas encode 360-degree visual information but introduce severe geometric distortions that confound neural reconstruction algorithms trained on conventional photographs. Neural VR addresses this by extracting multiple rectified perspective views from the source panorama, each oriented toward semantic regions identified in the original parametric model [4]. This perspective rectification approach builds upon the view synthesis strategy established in prior research [2]. For every furniture piece, architectural element, or functional zone, the system computes optimal virtual camera parameters—position, orientation, and field-of-view—that capture the region with minimal projection distortion. This decomposition transforms the single-viewpoint panorama into a collection of perspective images suitable for neural reconstruction while maintaining semantic correspondence with the design model.

Neural Point Cloud Generation for VR-Ready Geometry.
Perspective views serve as input to neural single-image reconstruction networks that infer 3D structure from 2D observations [7, 9]. Unlike traditional mesh reconstruction that requires watertight surfaces, point cloud representation tolerates the geometric ambiguity inherent in single-view inference while preserving the rich color information embedded in AI-generated imagery. Each reconstructed point inherits RGB values from the source panorama, capturing material appearances, lighting effects, and stylistic details that would require extensive manual specification in conventional VR authoring. The resulting colored point clouds form an intermediate representation that bridges image-based generation and spatial navigation—dense enough to convey visual presence, yet flexible enough to accommodate the uncertainty of monocular depth estimation.

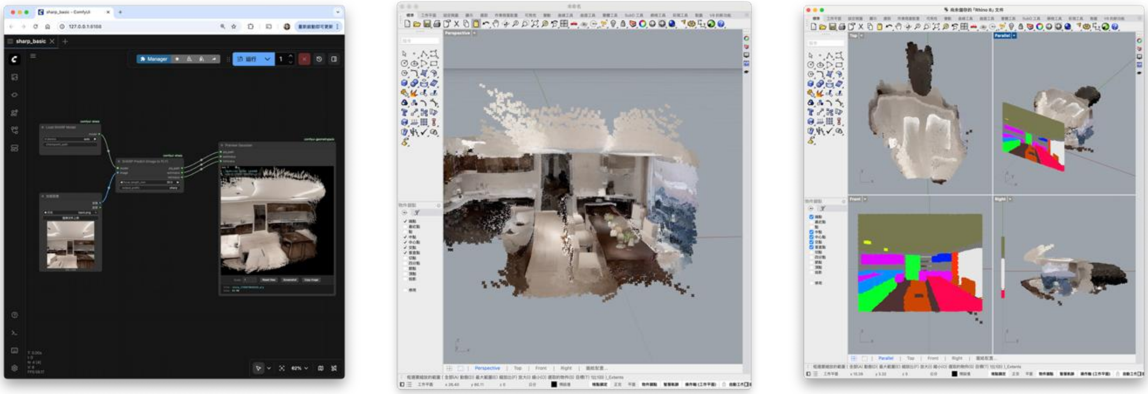

Fig. 2: Generating Point Cloud: (a) Generating point cloud from an extracting image by Apple SHARP model [5] in ComfyUI, (b) Input point cloud of the image in Rhino, (c) Segmentation of point cloud by the semantic segmentation map.

Semantic Transfer via Polar Coordinate Matching.
Explorable VR environments require more than raw geometry; meaningful interaction depends on understanding what objects occupy the space. Neural VR transfers semantic labels from the parametric design model to reconstructed point clouds through a polar coordinate matching algorithm. The semantic categories follow the ADE20K taxonomy [12], ensuring consistency with the ControlNet conditioning used during panorama generation. Both the point cloud and the semantic segmentation map are transformed into angular representations centered on the panorama's projection origin. Points are classified by matching their angular position and normalized radial distance to corresponding segmentation pixels, with color similarity providing disambiguation in boundary regions. This approach achieves robust category assignment despite the non-linear coordinate transformations introduced by perspective projection, enabling object-level queries, selective visibility, and category-based interaction in the final VR environment.

Constraint-Aware Scene Assembly for Spatial Coherence.
Individual point clouds reconstructed from separate perspective views must be unified into a coherent scene that respects the spatial logic of interior design. Neural VR evaluates assembled geometry against hierarchical constraints derived from the parametric model: architectural boundaries prevent impossible intersections with walls and floors; circulation constraints ensure navigable pathways remain clear; functional adjacencies maintain logical relationships between related objects. When constraint violations occur—a reconstructed sofa penetrating a wall, or a dining set blocking the kitchen entrance—the system applies category-appropriate corrections including scaling, rotation alignment, and gradient-based repositioning. Points that resist satisfactory resolution are flagged for designer review, preserving human judgment over critical spatial decisions while automating routine geometric corrections.

Enabling True VR Exploration.
The assembled, semantically-organized point cloud environment supports the spatial exploration that distinguishes VR from static visualization. Users wearing head-mounted displays can move freely within defined navigation bounds, experiencing parallax depth cues, examining furniture from multiple angles, and verifying spatial proportions through embodied perception. Because semantic labels persist throughout the pipeline, the VR environment supports object-level interaction: highlighting all seating elements, isolating the kitchen zone, or comparing alternative furniture configurations. This capability transforms AI-generated imagery from a presentation medium into a design tool—a space for discovery rather than merely display.

<u>Implementation and Testing:</u>
Neural VR was implemented as an integrated toolset combining GhPython components for Grasshopper in Rhino 8, ComfyUI workflows for neural reconstruction, and Twinmotion for VR deployment targeting Meta Quest headsets via the Unreal Datasmith Exporter. The Grasshopper components handle perspective extraction, semantic map generation, and point cloud import/classification; ComfyUI orchestrates the neural reconstruction pipeline using pre-trained single-image depth and point cloud generation models; Twinmotion receives the assembled scene through Datasmith and renders it for standalone VR browsing on Meta Quest 3.

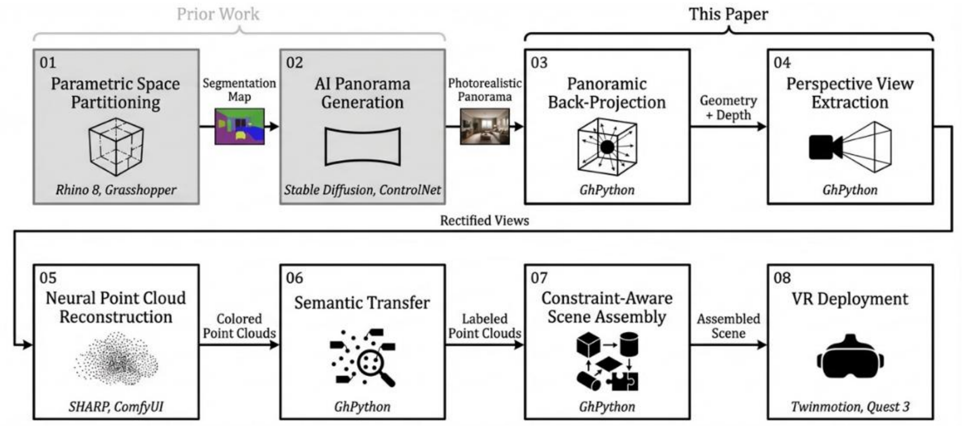

Fig. 3: The Neural VR workflow pipeline. Stages 1–2 build upon prior work [2, 3]; Stages 3–8 constitute the contribution of this paper.

Testing employed multiple interior configurations representing typical Taiwanese apartment layouts of varying sizes, encompassing living rooms, dining areas, and open kitchens with multiple semantic categories per scene. AI panoramas were generated using Stable Diffusion 1.5 with ControlNet semantic conditioning at 2048×1024 resolution. For each scene, several rectified perspective views were extracted depending on spatial complexity, and each view was processed through the SHARP neural reconstruction pipeline to produce dense colored point clouds.

Preliminary testing indicates that semantic classification via polar coordinate matching performs well for large architectural elements (walls, floor, ceiling) whose angular spans are clearly delineated in the segmentation map, while smaller or geometrically complex furniture categories show lower matching rates due to occlusion and boundary ambiguity. VR navigation testing on Meta Quest 3 hardware confirmed that the Twinmotion renderer maintains interactive frame rates suitable for comfortable exploration. Navigation bounds constrain user movement to spatially plausible regions. A notable limitation observed during VR deployment is that the AI-generated panoramas embed baked lighting effects—shadows, reflections, and ambient occlusion—directly into the texture, which cannot be reinterpreted by Twinmotion's real-time lighting engine. Because the imported scene lacks explicit light sources, the VR viewing experience in Twinmotion appears visually flatter than both the Rhino viewport preview (where the back-projected textures display naturally on the geometry) and the panoramic viewing through platforms such as Facebook established in prior research [2]. This discrepancy highlights the need for either baked-lighting-aware rendering modes in VR engines or post-processing techniques that compensate for the absence of physical light sources in texture-embedded scenes. Initial informal feedback from interior design professionals suggests that the explorable Neural VR environment offers noticeably improved spatial comprehension compared to static panorama viewing, particularly for evaluating circulation paths and furniture scale relationships. Formal user studies with systematic evaluation protocols are planned for the full paper.

Discussion:

Neural VR demonstrates that AI-generated panoramas can be transformed into explorable 3D environments, but several limitations constrain current applicability. First, single-image reconstruction inherently produces incomplete geometry; occluded surfaces and regions outside the original viewpoint lack point coverage, creating visual gaps during VR navigation. Multi-view synthesis from slight viewpoint perturbations—generating auxiliary panoramas from offset positions—could provide additional geometric constraints, though at increased computational cost.

Second, the polar coordinate matching algorithm assumes precise alignment between the point cloud and the segmentation image coordinate systems. Calibration errors or reconstruction distortions degrade classification accuracy, particularly near category boundaries. Learned correspondence networks that adapt to scene-specific characteristics may prove more robust than geometric matching alone.

Third, point cloud rendering, while efficient, lacks the view-dependent lighting effects that mesh-based or Gaussian splatting representations can achieve [6]. For design communication requiring precise material evaluation, the current pipeline serves better for spatial verification than material specification. Integration with 3D Gaussian Splatting reconstruction could potentially combine semantic organization with higher visual fidelity.

Finally, the bounded navigation region—limited to a modest distance from the panorama origin—reflects fundamental limitations of single-viewpoint reconstruction rather than implementation constraints. Extending exploration range requires either accepting degraded visual quality in extrapolated regions or generating multiple panoramas from different positions and fusing their reconstructions.

Conclusions:

This paper presents Neural VR, a workflow that transforms AI-generated panoramas into semantically organized, explorable 3D environments for interior design evaluation. By combining perspective extraction, neural point cloud reconstruction, polar coordinate semantic transfer, and constraint-aware scene assembly, the approach bridges generative visualization and immersive virtual reality. Designers can leverage AI's capacity for rapid, visually rich panorama generation while recovering the spatial navigability essential for meaningful VR experiences.

The contribution reframes AI-generated imagery as intermediate design material rather than terminal visualization output. Where previous workflows treated panoramas as endpoints suitable only for fixed-viewpoint presentation [3, 4], Neural VR positions them as sources for geometric extraction and semantic organization, enabling the transition from passive viewing to active spatial exploration. This capability supports design evaluation tasks—verifying circulation paths, assessing furniture scale, experiencing atmospheric qualities—that static imagery cannot adequately address.

Future work will explore multi-panorama fusion for extended navigation range, integration with 3D Gaussian Splatting [6] for improved visual fidelity, and real-time reconstruction pipelines that enable immediate VR preview during parametric design iteration. As neural scene reconstruction techniques continue to advance, the vision of seamless transition from AI-generated concept imagery to fully explorable virtual prototypes becomes increasingly achievable. Neural VR provides a foundational framework for that integration, demonstrating that the creative potential of generative AI and the spatial immersion of virtual reality can be unified in service of interior design practice.

Acknowledgments:
The National Science and Technology Council of Taiwan supported this research under grant number NSTC 112-2221-E-165-001-MY3.

Chieh-Jen Lin, https://orcid.org/0000-0001-9981-5864

References:
[1] Kerbl, B.; Kopanas, G.; Leimkuehler, T.; Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering, ACM Trans. Graph., 42(4), 2023, 1-14. https://doi.org/10.1145/3592433
[2] Lin, C.-J.: AI-generated VR: Leveraging AI and VR for Rapid Ideation and Concept Modeling of Interior Design Computer-Aided Design and Applications, 22(4), 2025, 629-639. https://doi.org/https://doi.org/10.14733/cadaps.2025.629-639
[3] Lin, C.-J.: Parametric VR: Leveraging Parametric Modeling and AI-Generated Panoramas for Rapid Ideation for Interior Design, in 2025, 87-92. https://doi.org/https://doi.org/10.14733/cadconfP.2025.87-92
[4] Lin, C.-J.: Parametric VR: A Generative Workflow Integrating Semantic Modeling and AI-Generated Panoramas for Ideation of Early-Stage Interior Design, Computer-Aided Design and Applications, 23(4), 2026, 370-380. https://doi.org/https://doi.org/10.14733/cadaps.2026.370-380
[5] Mescheder, L.; Dong, W.; Li, S.; Bai, X.; Santos, M.; Hu, P.; Lecouat, B.; Zhen, M.; Delaunoy, A.; Fang, T.; Tsin, Y.; Richter, S.R.; Koltun, V.: Sharp Monocular View Synthesis in Less Than a Second, in, 2025, arXiv:2512.10685. https://doi.org/10.48550/arXiv.2512.10685
[6] Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, Computer Vision – ECCV 2020, 2020, 405-421. https://doi.org/10.1007/978-3-030-58452-8_24
[7] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B.: TripoSR: Fast 3D Object Reconstruction from a Single Image, arXiv preprint, 2024. https://doi.org/https://doi.org/10.48550/arXiv.2112.10752
[8] Rombach, R.; Blattmann, A.; Lorenz, D.; Ommer, P.E.B.: High-Resolution Image Synthesis with Latent Diffusion Models, in 2022, 10684-10695. https://doi.org/https://doi.org/10.48550/arXiv.2112.10752
[9] Xu, J.; Cheng, W.; Gao, Y.; Wang, X.; Gao, S.; Shan, Y.: InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models, arXiv preprint, 2024. https://doi.org/https://arxiv.org/abs/2404.07191
[10] Zhang, L.; Rao, A.; Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, 3836-3847. https://doi.org/https://doi.org/10.1109/ICCV51070.2023.00355
[11] Zhao, Z.; Lai, Z.; Lin, Q.; Zhao, Y.; Liu, H.; Yang, S.; Feng, Y.; Yang, M.; Zhang, S.; Yang, X.; Shi, H.; Liu, S.; Wu, J.; Lian, Y.; Yang, F.; Tang, R.; He, Z.; Wang, X.; Liu, J.; Zuo, X.; Chen, Z.; Lei, B.; Weng, H.; Xu, J.; Zhu, Y.; Liu, X.; Xu, L.; Hu, C.; Yang, S.; Zhang, S.; Liu, Y.; Huang, T.; Wang, L.; Zhang, J.; Chen, M.; Dong, L.; Jia, Y.; Cai, Y.; Yu, J.; Tang, Y.; Zhang, H.; Ye, Z.; He, P.; Wu, R.; Zhang, C.; Tan, Y.; Xiao, J.; Tao, Y.; Zhu, J.; Xue, J.; Liu, K.; Zhao, C.; Wu, X.; Hu, Z.; Qin, L.; Peng, J.; Li, Z.; Chen, M.; Zhang, X.; Niu, L.; Wang, P.; Wang, Y.; Kuang, H.; Fan, Z.; Zheng, X.; Zhuang, W.; He, Y.; Liu, T.; Yang, Y.; Wang, D.; Liu, Y.; Jiang, J.; Huang, J.; Guo, C.: Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation, in, 2025, arXiv:2501.12202. https://doi.org/https://doi.org/10.48550/arXiv.2501.12202
[12] Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A.: Scene Parsing Through ADE20K Dataset, in 2017, 633-641. https://doi.org/10.1109/CVPR.2017.544