Research Statement

Zhiqi Huang

M.Phil. Student, Waseda University | huangzhq1747@gmail.com | huang-zhiqi.github.io

Research Vision

My research is centered on a simple idea: a strong understanding of 3D structure can make generative models more controllable, consistent, and useful. I view 3D not only as an output modality, but also as a representation for reasoning about humans, objects, materials, anatomy, and future states. Meshes, PBR materials, 3D Gaussian Splatting, volumetric data, depth, normals, and multi-view cues provide constraints that pure 2D appearance often lacks. My goal is to use these 3D priors to solve concrete research problems, and then extend the same view toward video generation, world models, and embodied intelligence, where spatial consistency and interaction require more than plausible frame synthesis.

This direction is strongly shaped by my industry experience. Before graduate study, I spent three years at 4399 Games building cross-platform real-time rendering systems for shipped mobile and PC titles. That work taught me how 3D assets, materials, scenes, and rendering pipelines are represented in game and simulation-like environments, and how they must be optimized for constrained mobile hardware. It also shaped my research taste: a generated 3D result should not only look good, but also be controllable, render consistently, fit into a pipeline, and support downstream use.

Research Experience

My current work follows one story across four research problems: identify the relevant 3D prior, connect it with the available modalities, and use it to make generation better aligned with the target task. In SIE3D, accepted to IEEE ICASSP 2026, I study single-image expressive 3D avatar generation with language control. The key challenge is to preserve a person's identity while allowing semantic control over expression. I approach this as a 3D human understanding problem: the model should construct an editable 3D Gaussian avatar, not only reproduce a 2D portrait. By combining 3D human priors with image identity and text semantics, this work connects identity-preserving reconstruction with controllable digital humans for games, VR, and interactive agents.

My second project, a first-author manuscript under review at ACM MM 2026, studies geometry-aware PBR material generation from long text. In practical graphics pipelines, a material is not just an image; it must be coherent on a mesh, relightable, and compatible with physically based rendering. This work uses geometry and mesh-normal priors to align long-form material descriptions with 3D assets, targeting PBR maps that are semantically aligned, multi-view consistent, and usable in simulation-ready asset pipelines. It extends my 3D story from humans to objects and materials.

My third project, a first-author manuscript targeted at AAAI 2027, applies the same perspective to longitudinal lung-nodule CT forecasting. CT is often treated as a sequence of 2D slices, but the underlying problem is volumetric: anatomy and disease progression happen in 3D. I therefore frame future CT prediction as a 3D understanding problem, using deform-then-edit forecasting to model local change while preserving surrounding anatomy. This project shows that my interest in 3D priors is not limited to graphics; it is a general way to constrain generation when spatial structure matters.

My fourth project, also targeted at AAAI 2027, studies text-conditioned PBR texture generation with prompt-derived reference conditions. Instead of first generating a reference image from text, the method maps material prompts directly into internal reference-condition tokens for a geometry-aware PBR generator. This connects language-level material descriptions with mesh geometry priors to generate coherent albedo, metallic, and roughness maps. Together with my ACM MM submission, this work studies how text, image-like reference information, geometry, and PBR representation can be fused inside a 3D generation pipeline.

Future Direction

These projects point toward my future research agenda: extending 3D-understanding-based generation from static assets to dynamic worlds. For video generation, I want to study models that use intermediate 3D structure such as depth, normals, object geometry, Gaussian fields, or scene layouts to maintain persistent geometry and object identity across time. For world models and embodied intelligence, I am interested in predicting not only future frames, but also how the 3D state of an environment changes under action. Such models could support robot manipulation, AR/VR, interactive simulation, and game-scale virtual worlds, where agents and users need stable spatial structure rather than short-lived visual plausibility.

Across my work, I aim to bridge generative AI with usable 3D systems. My background in real-time rendering gives me a practical understanding of 3D representation and deployment, while my research explores how those representations can guide multimodal generation. I hope to build models that expose structure: where objects are, how surfaces and materials behave, how identities and expressions are controlled, how anatomy changes, and how future states evolve. This is the core thread of my research: using 3D understanding to solve current scientific problems and to build the foundation for future video, world-model, and embodied AI systems.