VolumeMatic and the AI3D.foundation Benchmark Card for 3D Generative AI: Text to spatial computing volume app with generated content

Yosun Chang
yc@AIMagical.com
AIMagical AI3D.foundation
AReality3D Permute.xyz
San Francisco, CA, USA

Figure 1: Screenshots from VolumeMatic app, from left to right, top to bottom: AI3D Gen prompt widgets showing various service endpoints, Preview Model Prompt, Code Gen Prompt, Generated spatial computing game, Preview Image Prompt [Chang 2024b]

ABSTRACT

VolumeMatic is an Apple Vision Pro natural language text to interactive volume app creator app, utilizing the "chat to create" motif. The user can easily create 3D using various AI3D creation methods, such as text to 3D and image to 3D from different models and providers, utilizing an abstractified AI3D multimodal API. We utilize object detected and semantic relations among different HCI elements to enable natural language interactive spatial computing app creation. The app hopes to launch an AI3D Foundation to help accelerate the advancement and impact of AI for 3D and interactive content (AI3D). We also present the AI3D Benchmark Card to quickly summarize the results from different models, with a ground truth mesh.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Mixed / augmented reality**; **Graphical user interfaces**; **Natural language interfaces**; **Command line interfaces**; **User interface design**; • **Computing methodologies** → **Natural language processing**; **Discourse, dialogue and pragmatics**; **Graphics processors**.

KEYWORDS

AI3D, spatial computing, AI app creation, Apple Vision Pro

ACM Reference Format:
Yosun Chang. 2024. VolumeMatic and the AI3D.foundation Benchmark Card for 3D Generative AI: Text to spatial computing volume app with generated content. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Appy Hour (SIGGRAPH Appy Hour '24), July 27 - August 01, 2024.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3664294.3664361

SIGGRAPH Appy Hour '24 , July 27 - August 01, 2024, Denver, CO, USA

ACM ISBN 979-8-4007-0682-0/24/07.
https://doi.org/10.1145/3664294.3664361

1 INTRODUCTION

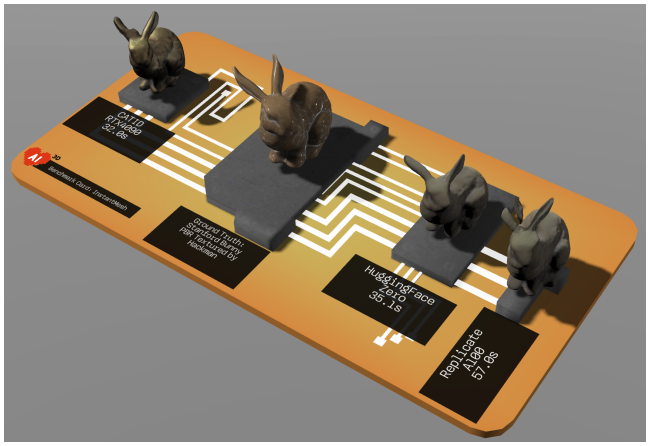
VolumeMatic starts with a freeform chat interface, where the user can chat to create 3D models using AI. The user can also chat to turn the created 3D models into interactive elements in a game or app. Everything happens inside an Apple Vision Pro volume app, and the output is another Vision Pro volume app!

The clientside interfaces with various model endpoint service backend that generates the 3D model and interprets text to commands and interactivity, as described in the AI3D API section below.

Previously, we built Napkinmatic 3D [Chang 2023b] and Napkinmatic [Chang 2023a], both of which required drawing skills. Being text input focused, we hope that this app is more accessible!

1.1 Volume Motifs

Discord [Discord Inc. [n. d.]] chat has become the current prevalent method for both image and AI3D creation, so we extended the interface concept and the contextual prompt widgets into spatial computing. In 3D, the "chat box" can appear on any of the 6 surfaces of an Apple Vision Pro cube volume, subject to user preference, as they view and interact with their creation from different angles. Similar to a screen, but in 3D, the interior of the cube contains the extent of the content and interactivity. The external surfaces thus become a sort of god's eye view in natural language "source creation" mode.



2 AI3D API

The AI3D API, maintained by the AI3D Foundation, provides a unified access point to various AI3D-related technologies across the ecosystem.

2.1 AI3D Endpoint Providers

Different commercial and open source models produce different results and incur different costs.

We include all commercial providers of AI3D creation. Currently, the list includes: CSM [csm 2023], Luma [lum 2023], Meshy [mes 2023], Sudo [sud 2023], and Tripo3D [tri 2023b].

Various model inference-per-second services such as Replicate [Rep 2023], fal [fal 2023a], and HuggingFace [Hug 2023] are also included, running open-source models such as LRM [LRM 2023], LGM [LGM 2023], TripoSR [tri 2023a], and InstantMesh [Git 2024].

2.2 Hosted AI3D Models

The AI3D Foundation also hosts a number of models and common workflows. One example: mixing different models in a pipeline for text to 3D: text to image via SDXL-lightning [Rep 2024a] ($\sim 1s$) or LCM ($\sim 200ms$ on [fal 2023b]) and image to 3D via Tripo3D [tri 2023b] or InstantMesh [Git 2024] ($\sim 15-30s$).

2.2.1 AI3D Benchmark Card. The AI3D Foundation produces a "Benchmark Card" that allows for succinct visual comparison of the mesh produced by each model, including the hardware used and inference time expended for each service endpoint.

Shown below is the AI3D Benchmark Card: Instant Mesh [Git 2024] benchmarking the Stanford Bunny [Laboratory 1993] PBR version [Hackmans 2022] using the following endpoints [Taylor 2024] [Hug 2024] [Rep 2024b] - permalink: [Chang 2024a].

3 NATURAL LANGUAGE INTERACTIVITY

VolumeMatic lets you turn natural language chat into generated code. We can process colloquial references and even slang for object references, which works well for small scenes. We show a context-sensitive widget when confusion arises.

3.1 Object Identification and Semantic Relations

We re-calibrate scale based on mesh bounds and re-assign pivots based on detected categories (by default, base center - but sometimes, it may make sense for the pivot to be mesh center).

We can use object detection to identify additional colliders to create for sub-object interactivity.

Strong semantic relations help strongly guide the code generation with the scene objects, building on the author's previous pre gen-AI work DrawmaticAR [Chang 2020], which parses text into known grammar nouns, adjectives, verbs, and applies material colors, animations accordingly to the modified noun object. With the advent of 1M context-window models such as Google Gemini 1.5 Pro [Team 2024], the AI can further query the user when ambiguities in HCI concepts arise, to help create what the user actually wants.

4 CONCLUSION

By combining the needed aspects of 3D app creation with the proper code gen and AI3D endpoints and intuitive context-sensitive widgets in a discord-inspired natural language interface, VolumeMatic hopes to enable anyone who wants to create a spatial computing app to easily create what they envision - interactively, in discourse, with virtually no learning curve required.

REFERENCES

2023. *csm.* Retrieved May 17, 2024 from https://csm.ai
2023a. *fal.* Retrieved May 17, 2024 from https://fal.ai
2023b. *fal LCM.* Retrieved May 17, 2024 from https://fal.ai/models/fal-ai/lcm
2023. *Github: LGM.* Retrieved May 17, 2024 from https://github.com/3DTopia/LGM
2023. *Github: LRM.* Retrieved May 17, 2024 from https://github.com/3DTopia/OpenLRM
2023a. *Github: TripoSR.* Retrieved May 17, 2024 from https://github.com/VAST-AI-Research/TripoSR
2023. *HuggingFace Spaces.* Retrieved May 17, 2024 from https://huggingface.com
2023. *Luma Labs.* Retrieved May 17, 2024 from https://lumalabs.ai
2023. *Meshy.* Retrieved May 17, 2024 from https://meshy.ai
2023. *Replicate.* Retrieved May 17, 2024 from https://replicate.com

2023. *SudoAI.* Retrieved May 17, 2024 from https://sudo.ai
2023b. *Tripo3D.* Retrieved May 17, 2024 from https://tripo3d.ai
2024. *Github: TencentArc/InstantMesh.* Retrieved May 17, 2024 from https://github.com/TencentARC/InstantMesh
2024. *huggingface space: TencentArc/InstantMesh on Zero.* Retrieved May 17, 2024 from https://huggingface.co/spaces/TencentARC/InstantMesh
2024a. *Replicate/bytedance/sdxl-lightning-4step on A100.* Retrieved May 17, 2024 from https://replicate.com/bytedance/sdxl-lightning-4step
2024b. *Replicate/camenduru/InstantMesh on A100.* Retrieved May 17, 2024 from https://replicate.com/camenduru/instantmesh
Yosun Chang. 2020. DrawmaticAR. https://DrawmaticAR.com.
Yosun Chang. 2023a. Napkinmatic. In *ACM SIGGRAPH 2023 Appy Hour* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3588427.3595357
Yosun Chang. 2023b. Napkinmatic 3D. In *ACM SIGGRAPH 2023 Real-Time Live!* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 6, 2 pages. https://doi.org/10.1145/3588430.3597253
Yosun Chang. 2024a. *AI3D Benchmark Card: Stanford Bunny PBR (Permalink).* Retrieved May 17, 2024 from https://ai3d.dev/benchmark/stanford_bunny_pbr/
Yosun Chang. 2024b. *VolumeMatic Demo.* Retrieved Feb 2, 2024 from https://x.com/Yosun/status/1759980927361360301
Discord Inc. [n. d.]. *Discord.* https://discord.com
Hackmans. 2022. *Sketchfab: Stanford Bunny PBR.* Retrieved May 17, 2024 from https://sketchfab.com/3d-models/stanford-bunny-pbr-42c9bdc4d27a418daa19b2d5ff690095
Stanford University Computer Graphics Laboratory. 1993. *Stanford Bunny.* Retrieved May 17, 2024 from https://graphics.stanford.edu/data/3Dscanrep/
Christopher A. aka CATID Taylor. 2024. *Github: catid/InstantMesh on RTX4090.* Retrieved May 17, 2024 from https://github.com/catid/InstantMesh
Gemini Team. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. arXiv:2403.05530 [cs.CL]