MagicPen: Interactive Sketch-to-3D Generation in Commercial Metaverse Platforms

Ryutaro Kurai*
Nara Institute of Science and Technology

Yuji Hatada†
The University of Tokyo

Yuichi Hiroi‡
Cluster Metaverse Lab

Takefumi Hiraki§
University of Tsukuba
Cluster Metaverse Lab

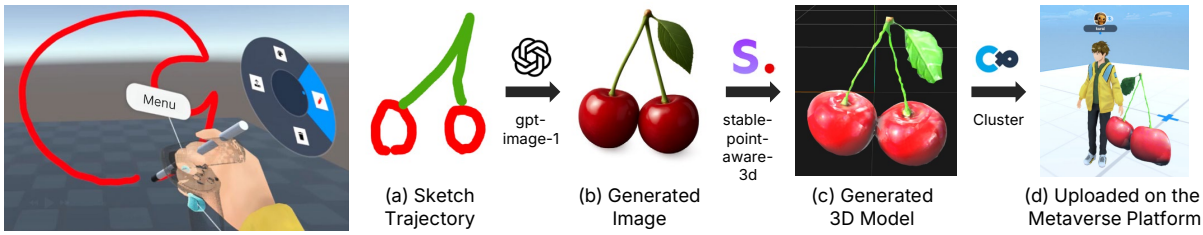

Figure 1: System overview demonstrating the workflow from VR sketch input to integration within Cluster, a commercial metaverse platform. (Left) Users create 3D sketch trajectories with pen-based tools and interface controls for drawing modes, color selection, and upload functionality. The system records pen trajectory coordinates and user viewpoint data, converting the 3D trajectory into a 2D rendered image. (Right) The AI processing pipeline transforms sketch data: (a) sketch trajectory input, (b) image refinement generating detailed 2D representations with realistic texturing, (c) 3D model generation with complete geometric structure, and (d) automatic deployment as interactive Craft Items within Cluster's shared virtual environment.

Abstract

Although User Generated Content (UGC) distinguishes metaverse platforms from conventional VR, 3D content creation remains limited to users with specialized modeling expertise and platform-specific technical knowledge. While generative AI advances have improved accessibility, existing approaches fail to provide both VR-native operation and intuitive input capabilities simultaneously, maintaining barriers for novice users.

We propose MagicPen, a system enabling direct conversion of hand-drawn sketches into 3D objects within VR environments through a three-stage AI pipeline comprising image refinement, 3D model generation, and metaverse integration. Evaluation with 68 diverse sketch inputs revealed 35.5-second average processing times and high-fidelity geometric reproduction, validating seamless Cluster platform integration. The system enables collaborative 3D content creation in social VR spaces, allowing users without modeling expertise to transform individual creative workflows into shared social experiences.

Index Terms: Metaverse, Sketch-based Modeling, 3D Generative AI

1 Introduction

Social VR platforms such as VRChat[1], Cluster[2], and Resonite[3] have evolved into comprehensive metaverse environments supporting both immersive VR experiences using Head-mounted displays (HMDs) and non-immersive participation through PCs and mobile devices. These metaverse platforms differ fundamentally from isolated virtual environments by establishing creative and collaborative spaces premised on multi-user communication [5]. User Generated Content (UGC), the cultural activities and communities emerging from user-driven creations, represents a defining characteristic that distinguishes contemporary metaverse platforms from conventional virtual environments such as online games. Users function not merely as observers but as active participants in metaverse formation, creating customized objects and complete virtual worlds that they share and distribute within their communities. However, 3D content creation still remains limited to users with specialized modeling skills and platform-specific technical knowledge.

One particularly effective approach to overcoming this technical barrier is generating 3D models from hand-drawn sketches, which enables users without specialized expertise to express spatial forms through drawing interactions. Teddy [1], in particular, is known as a pioneering system that provides intuitive 2D-to-3D conversion capabilities. Recent rapid advances in generative AI technology have led to active research in generating 3D models from text and images. Methods such as DreamFusion [4] and Magic3D [3] demonstrated impressive results, however, importing the generated content into social VR environments requires complex workflows.

Researchers have explored ways to integrate generated 3D objects into commercial metaverse platforms. MagicCraft [2] offers a tool that uses text input to generate interactive items compatible with these platforms. Roblox Cube [6] generates complete spatial environments from textual descriptions for use within metaverse platforms. However, these approaches lack complete workflows within VR spaces and do not support generating outputs based on user sketches within VR environments.

We propose a system that allows users to convert hand-drawn sketches into 3D objects within commercial metaverse platforms and share them within those platforms. Through an integrated pipeline that combines AI-based generative image transformation and 3D generation techniques, the system automatically produces high-quality 3D models. Furthermore, we integrate the configuration and upload processes necessary for placing generated 3D

*e-mail: ryu.kurai@gmail.com
†e-mail: hatada@cyber.t.u-tokyo.ac.jp
‡e-mail: y.hiroi@cluster.mu
§e-mail: hiraki@slis.tsukuba.ac.jp

[1]https://hello.vrchat.com/
[2]https://cluster.mu/
[3]https://resonite.com/

models in metaverse spaces, realizing a seamless workflow. This enables users without 3D modeling expertise to create and share 3D content collaboratively within their preferred metaverse environment.

2 System Design

Our system architecture consists of three primary components: VR sketch input, AI-powered 3D generation pipeline, and metaverse integration subsystem.

2.1 VR Sketch Input Interface

Users interact with pen-shaped tools attached to their avatar's hands within the VR environment (Fig. 1, left). The interface provides four essential functions: (1) drawing mode toggling, (2) color selection from an integrated palette, (3) sketch deletion capabilities, and (4) upload execution toggling when drawing is finished.

When users complete their sketch and initiate upload, the system records both the 3D trajectory coordinates of the pen tool and the user's viewpoint information (camera coordinates) for subsequent processing.

2.2 Three-Stage AI Generation Pipeline

The right side of Figure 1 shows our AI-based sketch-to-3D pipeline integrated to the metaverse platform. The recorded 3D trajectories undergo processing through three distinct stages designed to transform raw sketch data into metaverse-ready 3D objects. During processing, yellow placeholder cubes are displayed at the original sketch positions, providing users with visual feedback about the generation status.

2.2.1 Stage 1: Image Refinement

The system renders 3D trajectories from the user's recorded viewpoint, generating PNG images that serve as input for refinement using OpenAI's gpt-image-1 API.

We employ carefully designed prompts to transform simple sketch outlines into detailed, 3D-rendered style images while preserving original shape characteristics and spatial relationships:

```
The attached image is an outline showing the
overview of an object.
Based on this outline, imagine and draw the
object.
Render the image in a 3D CG style.
Draw the object so that its complete form is
displayed.
If there are any text strings in the image,
those strings represent the object's name or
description.
Use the text string descriptions for rendering
the image.
Never render any text characters in the output.
```

2.2.2 Stage 2: 3D Model Generation

Refined images feed into Stability AI's stable-point-aware-3d API, which generates high-quality 3D models in GLB format from single image inputs. This API demonstrates robust performance in creating detailed 3D meshes with appropriate texturing from 2D image sources.

2.2.3 Stage 3: Metaverse Integration

Generated GLB files receive platform-specific metadata compatible with the Cluster metaverse platform. This enables automatic creation of "Craft Items" – interactive 3D objects that avatars can manipulate, carry, and place within shared virtual spaces. This generated craft item reappears in the same position as the yellow placeholder cube in the original sketch.

3 Implementation and Evaluation

We implemented the system as an integrated component within the Cluster metaverse platform, supporting VR-HMD users. We implemented the VR interface component of our system using Cluster HUD API.

3.1 Performance Analysis

We evaluated system performance using 68 diverse sketch inputs representing various object types and complexity levels. Processing time from trajectory data reception to 3D model completion averaged 35.5 ± 10.2 seconds, with one outlier exceeding 600 seconds excluded from analysis. Although this latency prevents real-time interaction, it remains acceptable for creative activities.

Quality assessment revealed strong reproduction fidelity for simple geometric shapes, with limitations observed in complex forms requiring fine detail representation. These findings suggest opportunities for improvement through advanced 3D generation models and multi-viewpoint sketch input support.

4 Applications and Future Work

Our prototype system allows for the creation of 3D content within commercial metaverse environments using sketch-based input methods. The synchronous information sharing of these metaverse platform allows creators to visualize their ideas in real-time and receive immediate feedback, facilitating collaborative brainstorming. This approach transforms traditional, isolated modeling workflows into social, creative experiences.

The system is particularly effective in educational contexts, where students can develop 3D spatial reasoning and creative skills without facing technical barriers. Its collaborative nature enables group projects and peer learning experiences within immersive virtual environments.

Future development will address several technical objectives: (1) reducing processing latency through optimized generative AI model selection, (2) implementing multi-viewpoint sketch support for complex object creation, (3) incorporating style control responsive to scene atmosphere and user preferences [7], (4) developing gimmick addition mechanisms that derive interactive functionality from object context and geometry [2], and (5) expanding compatibility across additional metaverse platforms. We also plan user studies to evaluate interface usability, explore applications in collaborative tasks, and measure educational effectiveness.

5 Conclusion

We presented a VR system that transforms hand-drawn sketches into 3D objects via a three-stage AI pipeline. This pipeline includes image refinement, 3D model generation, and metaverse integration. It enables direct content creation within metaverse environments. Evaluating the system with 68 sketch inputs demonstrated its viability, achieving an average processing time of 35.5 seconds and successfully reproducing the quality of geometric forms. Seamless integration with the Cluster platform as interactive craft items validates the system's practical deployment capabilities for collaborative workflows.

By eliminating the need for specialized modeling expertise, we hope the system will provide non-experts with seamless access to collaborative 3D content creation in social VR spaces, transforming individual creative work into social experiences.

Acknowledgments

This work was partially supported by JST Moonshot Research & Development Program Grant Number JPMJMS2013.

REFERENCES

[1] T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: a sketching interface for 3D freeform design. In *ACM SIGGRAPH 2006 Courses on - SIGGRAPH '06*. ACM Press, New York, New York, USA, 2006. 1

[2] R. Kurai, T. Hiraki, Y. Hiroi, Y. Hirao, M. Perusquía-Hernández, H. Uchiyama, and K. Kiyokawa. MagicCraft: Natural language-driven generation of dynamic and interactive 3D objects for commercial metaverse platforms. *IEEE Access*, 13:1–16, 2025. 1, 2

[3] C.-H. Lin, J. Gao, L. Tang, T. Takikawa, X. Zeng, X. Huang, K. Kreis, S. Fidler, M.-Y. Liu, and T.-Y. Lin. Magic3D: High-resolution text-to-3D content creation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 300–309, 2023. 1

[4] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall. DreamFusion: Text-to-3D using 2D diffusion. *arXiv [cs.CV]*, Sept. 2022. 1

[5] G. D. Ritterbusch and M. R. Teichmann. Defining the metaverse: A systematic literature review. *IEEE Access*, 11:12368–12377, 2023. doi: 10.1109/ACCESS.2023.3241809 1

[6] F. A. T. Roblox. Cube: A roblox view of 3d intelligence. *arXiv preprint arXiv:2503.15475*, 2025. 1

[7] L. Stacchio, E. Balloni, E. Frontoni, M. Paolanti, P. Zingaretti, and R. Pierdicca. Minevra: Exploring the role of generative ai-driven content development in xr environments through a context-aware approach. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):3602–3612, 2025. doi: 10.1109/TVCG.2025.3549160 2