ImmerseSketch: Transforming Creative Prompts into Vivid 3D Environments in VR

Alfred Lan
National Tsing Hua University
Hsinchu, Taiwan
lan.alfred.y@gmail.com

Tai-Chen Tsai
National Tsing Hua University
Hsinchu, Taiwan
astro5382@gapp.nthu.edu.tw

Chih-Chuan Huang
National Tsing Hua University
Hsinchu, Taiwan
masonhuang0320@gmail.com

Pu Ching
National Tsing Hua University
Hsinchu, Taiwan
geep80108@gapp.nthu.edu.tw

Tse-Yu Pan
National Taiwan University of Science and Technology
Taipei, Taiwan
typan@mail.ntust.edu.tw

Min-Chun Hu
National Tsing Hua University
Hsinchu, Taiwan
anitahu@cs.nthu.edu.tw

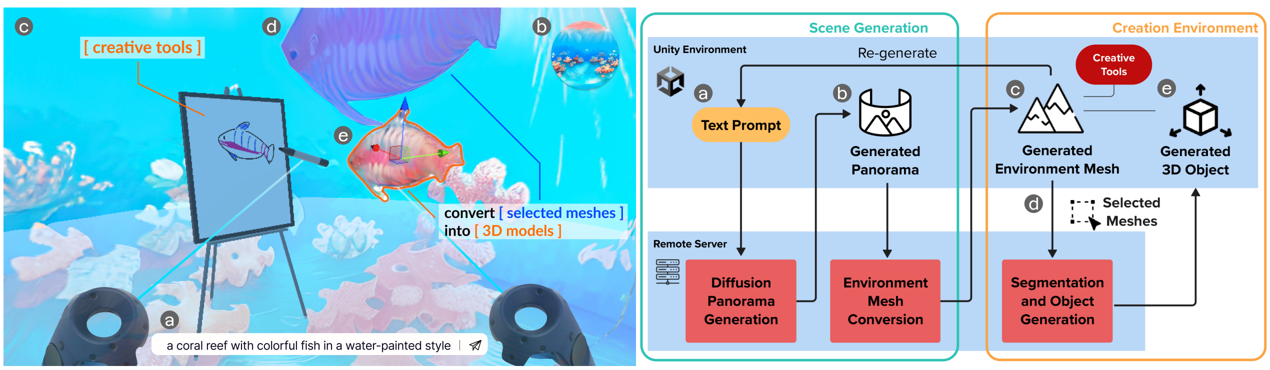

Figure 1: We propose a system that enables creators to deeply immerse themselves in environments of their chosen theme, and offer generated 3D environments and objects as painting references. Left: Application Scenario. Right: Overall Workflow. (a.) Scene description given by the user. (b.) Generated panorama. (c.) Generated mesh of the environment converted from panorama. (d.) Selected section of the environment mesh. (e.) Generated standalone 3D object based on the selected section.

ABSTRACT

We propose ImmerseSketch, a framework designed to transform creative prompts into vivid and detailed 3D content within a Virtual Reality (VR) environment. Our aim is to inspire creative ideation and to provide creators with 3D reference content for painting. However, the computational demands of diffusion models for 3D model generation and the scarcity of high-quality 3D datasets limit the diversity and intricacy of the generated environments. To overcome these obstacles, we focus on generating initial panoramic images through diffusion models and then converting these images into rich 3D environments with the aid of depth estimation. The pilot study shows that the proposed ImmerseSketch can provide an immersive environment and assist the process of creation.

SIGGRAPH Posters '24, July 27–August 01, 2024, Denver, CO, USA

ACM ISBN 979-8-4007-0516-8/24/07
https://doi.org/10.1145/3641234.3671078

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *Interaction design process and methods.*

KEYWORDS

Virtual Reality, VR Creation, Generative Environment

ACM Reference Format:
Alfred Lan, Tai-Chen Tsai, Chih-Chuan Huang, Pu Ching, Tse-Yu Pan, and Min-Chun Hu. 2024. ImmerseSketch: Transforming Creative Prompts into Vivid 3D Environments in VR. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH Posters '24), July 27–August 01, 2024, Denver, CO, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3641234.3671078

1 INTRODUCTION

Traditionally, artists often rely on search engines to discover thematic photographs, which serve as references for details and ignite creative inspiration. However, this method is flawed due to the inherent lack of depth information in photographs, which limits the artists' understanding of an object's spatial sense and three-dimensional structure. Compared to the above-mentioned conventional art creation scenario, Virtual Reality (VR) provides an immersive space that can give users a sense of depth, physicality, and presence. Consequently, research has delved into leveraging VR to enhance the creative process. Past works such as 360proto[Nebeling and Madier 2019] and 360theater[Speicher et al. 2021] utilize VR to preview and prototype spaces made from sketches or panoramic drawings on physical paper during the design process. Recent works such as sPellorama[Chen et al. 2023] harness generative AI to establish a voice-to-panorama system, empowering users to generate 2D panoramas through vocal prompts. Nonetheless, many of these frameworks fall short in constructing three-dimensional environments or furnishing tools for active creation. Their primary function is to facilitate the preview of ideas proposed by users. In this work, we aim to develop a system that generates a variety of 3D environments and objects based on creative text prompts to assist users in their art process and help them ideate.

2 METHOD

The system flow of the proposed ImmerseSketch tool is shown in Fig. 1, which includes three main modules: (1) Diffusion Panorama Generation Module, (2) Environment Mesh Conversion Module, and (3) Segmentation and Object Generation Module.

2.1 Diffusion Panorama Generation

At the start of the user flow, users are given a textbox to enter prompts (by speech or typing) for their desired environment. These prompts are then input into a Stable-Diffusion-based framework specializing in producing panoramas, namely Diffusion360[Feng et al. 2023]. The framework produces multiple panoramas as options for the user, each of which is showcased on an individual preview orb within the world space. Users can point towards a given orb and select it using their VR controller. Subsequently, users have the option to convert a selected panorama into a 3D environment mesh. Alternatively, if none of the options meet the user's needs, they can backtrack to the previous step to input new prompts.

2.2 Environment Mesh Conversion

We utilize DepthAnything[Yang et al. 2024], a zero-shot depth estimation method known for its strong generalization capabilities, to convert a selected 2D panorama image into a 3D environment mesh. Our framework estimates the depth of the selected panorama image and integrates the predicted depth information with RGB data to generate a 3D environment mesh. The conversion process projects each pixel outward from a central point, using the pixel's estimated depth as the vertex distance. Subsequently, the 3D mesh is constructed by connecting these vertices.

2.3 Segmentation and Object Generation

The converted environment mesh in the previous step is a singular mesh, and objects in the selected panorama scene are not standalone. To enable users to rotate and observe an object selected in the panorama independently, we employ One-2-3-45[Liu et al. 2024] for transforming a single image of an object into the corresponding 3D model. Users are prompted to select an object with their VR controller. A screenshot is then captured from the user's head-mounted display view. This screenshot, along with the projected pointing location, is provided to the Segmentation Module, which outputs the predicted segmentation map of the selected object. Once the segmentation result is completed, the segmented image is taken as the input for One-2-3-45 to generate a 3D model. Upon completion, the object is displayed in an inventory menu, enabling users to bring it into the world space for rotation and observation.

3 PILOT STUDY

The pilot study was conducted based on 14 participants who have a passion for painting. The experimental scenes cover three styles: realistic, natural, and futuristic fantasy. Before the task, participants received a brief introduction to the system and a 10-minute VR controls familiarization session. Each participant was then instructed to create three sketches, one for each scene, within a time frame of 5 to 10 minutes per sketch. Results created by participants are shown at https://kalmalfred.github.io/ImmerseSketch. After completing the tasks, each participant completed a System Usability Scale Questionnaire(SUS), a Slater-Usoh-Steed Presence Questionnaire(SUS-PQ), and participated in an interview to share their experience. The SUS employs a 5-point scale to assess system usability. Our system achieved a mean score of 77.1 with a standard deviation of 12 (Level B), indicating successful operation with minimal cognitive load. The SUS-PQ, using a 7-point scale, rated our system at 4.3±1.6, reflecting high user perception of the virtual environment. During the interviews, participants unanimously agreed that the generated environments aided in ideating concepts and provided an immersive, relaxing setting for painting.

4 DISCUSSION AND FUTURE WORKS

Our system is currently in its experimental phase, offering opportunities for refinement. Our primary objective for future work is to enhance immersion and the quality of generated 3D environments. Based on participant feedback, one enhancement involves integrating soundscapes generated from prompts to enrich user experience. Additionally, recent advancements in 3D generation, such as DreamGaussian, demonstrate improvements in quality and speed using 3D Gaussian Splatting. We plan to incorporate this technique to improve the consistency and quality of generated environments while maintaining efficient rendering.

REFERENCES

Timothy Chen, Miguel Ying Jie Then, Jing-Yuan Huang, Yang-Sheng Chen, Ping-Hsuan Han, and Yi-Ping Hung. 2023. sPellorama: An Immersive Prototyping Tool using Generative Panorama and Voice-to-Prompts. In *ACM SIGGRAPH 2023 Posters* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 42, 2 pages. https://doi.org/10.1145/3588028.3603667

Mengyang Feng, Jinlin Liu, Miaomiao Cui, and Xuansong Xie. 2023. Diffusion360: Seamless 360 Degree Panoramic Image Generation based on Diffusion Models. *arXiv preprint arXiv:2311.13141* (2023).

Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Mukund Varma T, Zexiang Xu, and Hao Su. 2024. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization. *Advances in Neural Information Processing Systems* 36 (2024).

Michael Nebeling and Katy Madier. 2019. 360proto: Making interactive virtual reality & augmented reality prototypes from paper. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

Maximilian Speicher, Katy Lewis, and Michael Nebeling. 2021. Designers, the stage is yours! medium-fidelity prototyping of augmented & virtual reality interfaces with 360theater. *Proceedings of the ACM on human-computer interaction* 5, EICS (2021), 1–25.

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth anything: Unleashing the power of large-scale unlabeled data. *arXiv preprint arXiv:2401.10891* (2024).