Title:

AI-generated VR: Leveraging AI and VR for Rapid Ideation and Concept Modeling of Interior Design

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

Keywords:
Virtual Reality (VR), Parametric Modeling, AI-generative Panorama, Interior Design, Conceptual Ideation

DOI: 10.14733/cadconfP.2024.141-146

Introduction:

With the decline in hardware and software costs and improved computational efficiency, virtual reality (VR) is making a resurgence in design education and practices [2]. VR offers superior communication to 2D drawings and 3D renderings, delivering a spatial experience closest to the actual built result. However, VR is always limited by difficulties in creating available 3D models and is mainly used only for final presentations rather than as an ideation tool or discussion medium in the early design stage. Limitations also arise in the manipulation within VR, allowing only a limited placement and modification of existing objects like furniture and materials within predefined environments. Additionally, more proper design approaches need to be adapted to VR applications, hindering its widespread adoption in the early design stage. Therefore, the previous research proposed a generative modeling approach based on the Level of Development (LOD) concept by using Grasshopper to transform bubble diagrams of a single-story house into orthogonal plans using Voronoi diagrams[4]. This approach allows the quick generation of interior partition concept models that can be previewed in VR, making it more suitable for interior design education and practices than physical models. The quickly generated 3D indoor prototype can only provide the background for placing various objects in VR. However, the rapid generation of more details for design ideation and discussions, such as lighting, material textural maps, and other smaller objects for different style ideas and design concepts, is still a laborious process and full of challenges.

With the emergence of generative artificial intelligence (GenAI) technologies, like Midjourney, DALL-E by OpenAI, and Stable Diffusion by Stability AI, capable of generating images from simple text prompts and image inputs, are beginning to be applied in the creative ideation phase of architectural and interior design. Research indicates that AI-generated contents have the potential to nurture designers' inspiration, creativity, and skill development [1, 5]. However, GenAI can now only generate 2D images or rough 3D object models, and generated results are random. Therefore, it is complicated to control the consistency of the different generated stages. An exterior perspective can only reveal one or two building facades for architectural design. Controlling the viewing angle to generate the building's appearances in different directions can slightly avoid inconsistency. However, for interior design, a single indoor perspective will show two or three interior walls in addition to the ceiling and floor. Hence, it is often difficult to solve the problem of inconsistency by only controlling viewing angles. Therefore, A more feasible way is to generate images of the entire indoor space at once, applying AI to generate a panoramic view of the whole indoor space. Based on this, it is possible to develop further 3D models for applying VR as an ideation and discussion tool. This study, therefore, proposes the use of Stable Diffusion in conjunction with panorama LoRA, ControlNet models, and rapidly generative 3D models to generate panorama and 3D models for VR that align with users' proposed design concepts.

Main Ideas:

The advantage of VR is that it can provide correct proportions among objects, but also a 1:1 spatial experience. However, it cannot replace the application of physical mock-up models in education and practice because it takes a lot of work to create a usable 3D model. On the other hand, VR devices, whether glasses or headsets, are not only expensive but also usually quite bulky. In addition to the content provided by the game platform, self-made content often requires a computer to offer high-speed 3D computing capabilities for high-quality content without appropriate compilation. One of the easiest ways to display and share VR content is through panoramic image sharing on Facebook, which can quickly display the same content between smartphones, tablets, and VR devices. Although the panorama is not an accurate 3D model and lacks the spatial navigation function and object operability in VR, it is usually enough for quick sharing and creative ideations. Especially when viewed through VR devices, such as Meta Quest 2 or 3, users can directly open a Facebook panorama to get a close-to-real experience of entering an indoor or outdoor space. Using generative AI technology, it can not only quickly generate panoramas but also is possible to obtain additional graphic feature information through AI analysis, such as canny, depth, normal maps, object segmentation, and other information, which can be used to regenerate a 3D prototype by the generated panorama quickly.

The advantage of AI-generated images for creative ideations is that they can quickly generate images by textual prompts. Still, the disadvantage is that those images often have no consistency among different generating stages and cannot control the proportion among objects and space like 3D renderings. For generating panoramic-style images, the LoRA (Low-Rank Adaptation) model must be used to modify the generated results. LoRA is a novel approach proposed by Microsoft's researchers, who suggested modifying specific layers' weights in deep learning models in a low-rank manner, achieving quick and effective adaptation to new tasks while avoiding extensive retraining of the original model [3]. For generating image tasks, a LoRA model can affect the overall style or theme of the generated image through text prompts and given weighting. By applying a panoramic style LoRA, it is easy to generate a panorama by ©nputting textual prompts. However, it still cannot avoid the problem of wrong proportions among spaces and objects inside.

To solve the problem of wrong proportions among spaces and objects inside, the ControlNet models can be applied to control the results generated by the image features of a given image, including contours, depth and normal maps, object segmentation, and so on [6]. Due to the powerful capabilities of controlling composition, some 3D modeling software, including SketchUp, Rhino, and Revit, has launched plugins or extension apps based on Stable Diffusion and ControlNet as an alternative tool for real-time photorealistic 3D rendering. However, just like the problem with VR applications, the creative ideation stage is unnecessary and usually takes no time to build a complete model for 3D rendering. However, to control the correct proportion among objects and spaces inside the panorama, it is still necessary to provide a panoramic sketch with the proper spatial and object proportions through a conceptual and simplified 3D model, which can control the generation results of panoramic LoRA. Most 3D rendering software, such as Vray, Enscape, Lumion, and Twinmotion, can easily import 3D models to generate panoramas. In addition to requiring additional software operations, most rendering software aims to produce photorealistic rendering effects. 3D rendering images usually provide too many unnecessary image features. Since the contour features of CAD drawings are more apparent than those of 3D rendering, it is more suitable for ControlNet to generate panoramas by inputting CAD-like drawings as 2D images. However, most 3D modeling software has no function to convert 2D drawings into panoramas. This study, therefore, develops Python-based Grasshopper components for generating cubic projection images from any Rhino views and then further generating a CAD-Like panorama based on the display style of the given view.

Based on the above technical requirements and explanations, this study proposes using Rhino and Grasshopper as the platforms to apply previous research results, including the quickly generated 3D prototype of indoor spaces and using AI assistance to generate panorama and 3D models for VR quickly. The approach is to develop Grasshopper components for (1) To generate CAD-like panoramic images from a given 3D prototype, (2) to apply text-to-image by ControlNet and LoRA to generate panoramas that align with the CAD-like panoramic images, (3) to display and share the generated panoramas in VR devices. Each step is detailed below:

Generation of CAD-Like Panoramic Images from Given 3D Prototype

As mentioned above, the 3D rendering results are too realistic, and the 3D modeling software cannot convert 2D-style drawings into panoramic images. A quick development method to generate CAD-like panoramic images is to use Python scripts to generate six cube projection images from the camera position of the Rhino view and then use a Python library such as py360convert (https://github.com/sunset1995/py360convert) to convert the front, rear, left, right, and upper projection images into a 1:2 ratio equirectangular projected panorama. In the past, Rhino 7 could only support Iron Python 2, making it challenging to implement such a function. Thanks to Rhino 8, it began to support Python 3 and can automatically download the specified libraries from the Internet, such as OpenCV and Numpy for py360convert; the panoramic functions without the need for additional rendering software, therefore, can be achieved in Rhino 8. Although the function of generating a panorama from the camera position of any current view in Rhino has been implemented (Fig. 1), the generating style still needs to be set in the Rhino current view in advance, such as Shaded, Technical, or Monochrome mode. Based on the recent test results, the Arctic or Monochrome modes are very suitable for rapidly generating images used by ControlNet to control contours without considering materials, layers, and other texture mapping issues in the creative ideation stages.

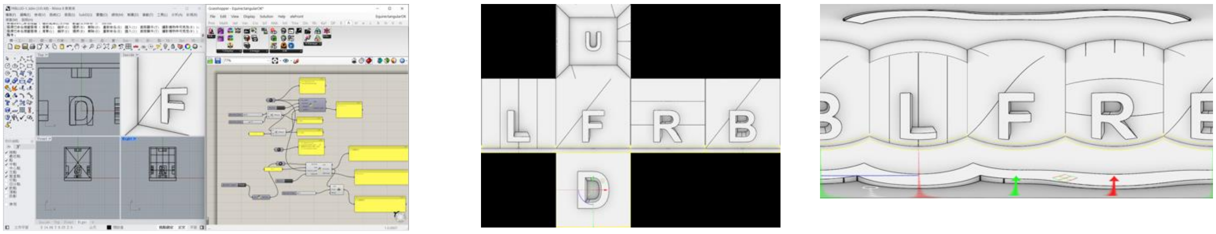

Fig. 1: Generation of CAD-like panoramic image from a view in Rhino: (a) Setting the display mode of the view (Left), (b) the generated cubic projection images (Middle), and (c) the converted equirectangular projection image (Right).

Generation of Panoramas by Textual Prompts with Stable Diffusion, ControlNet, and LoRA

Stable Diffusion, developed by Stability AI (https://huggingface.co/runwayml/stable-diffusion-v1-5), is a very popular open-source text-to-image (t2i) and image-to-image (i2i) generation model. Unlike other similar but paid private models such as Midjourney or DALL-E, Stable Diffusion not only allows local and remote installation but also derives many fine-tuned models (called checkpoints) and personal training of LoRA and ControlNet models. These model files are available from HuggingFace (https://huggingface.co) or CivitAI (https://civitai.com) and can be manuplated through the popular A1111 WebUI interface (https://github.com/AUTOMATIC1111/stable-diffusion-webui). A1111 WebUI can be installed on cloud computing platforms such as Google Colab. However, since Google Colab has banned the free GPU usage times to generate images by Stable Diffusion, installing A1111 WebUI on a local machine becomes a relatively cheap option but is itself a technical challenge. Fortunately, Lykos AI has developed the Stability Matrix (https://github.com/LykosAI/StabilityMatrix), which is a package management software that can install A1111 WebUI and other Stable Diffusion-related packages such as Fooocus (https://github.com/lllyasviel/Fooocus) with one click. It also allows downloading and managing models such as checkpoints, ControlNet, and LoRA files from the CivitAI website, therefore reducing the technical barriers to applying Stable Diffusion.

In addition, since A1111 WebUI provides server and API functions that allow remote operation of t2i and i2i generation functions, it will enable the Grasshopper component to operate remotely. There is already a Grasshopper toolset named Ambrosinus Toolkit that provides the operation of t2i, i2i, and ControlNet v1.x functions by the API of the local installation of A1111 WebUI package (https://www.food4rhino.com/en/app/ambrosinus-toolkit). However, this tool does not support the remote operation of A1111 WebUI server, nor is it compatible with the Stability Matrix package management method. On the other hand, to generate a panorama for displaying on a VR device such as Meta Quest 2 or 3, in addition to ControlNet, the A1111 WebUI needs to install additional extensions. For example, the Asymmetric tiling extension (https://github.com/tjm35/asymmetric-tiling-sd-webui) allows the generated image to avoid apparent seams in the X-axis direction, which is very important for the spatial experience when using a VR device to view a panorama. Therefore, this study developed Python 3 scripts to remotely operate the t2i generation, ControlNet, and Asymmetric tiling extensions and also implemented that the original A1111 WebUI can apply three ControlNet model functions simultaneously through the A1111 WebUI API. In addition, the Human UI component is also used to develop a simplified version of the A1111 WebUI interface to simplify the parameters adjustment for Stable Diffusion and ControlNet operation in the Grasshopper interface (Fig. 2).

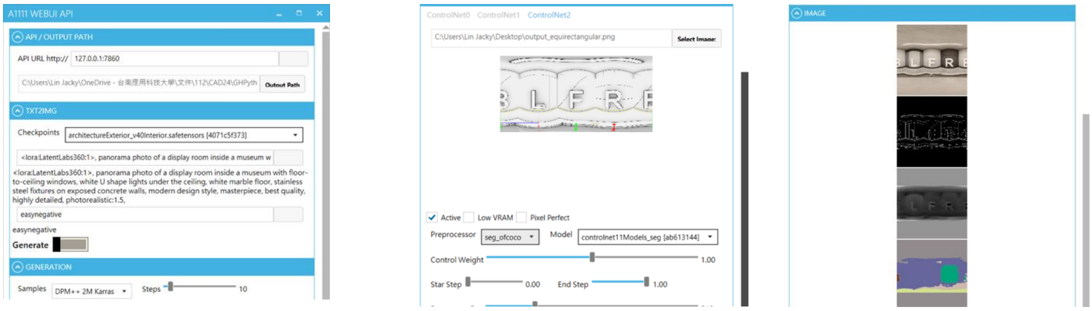

Fig. 2: HumanUI interface for generating panorama by Stable Diffusion, ControlNet, and LoRA: (a) Input textual prompts to activate panorama LoRA (Left), (b) input controlling images for ControlNet (Middle), and (c) the generated panorama by three ControlNet models (Right).

Although the Stability Matrix app can facilitate the management of checkpoint, ControlNet, and LoRA model files, currently, whether the Ambrosinus Toolkit or the results of this research, to install extended functions such as ControlNet and Asymmetric tiling still requires operating the A1111 WebUI interface through a web browser in advance. The difference is that the Ambrosinus Toolkit can only access the local installation. The results of this research can allow students and employees to remotely access the A1111 WebUI and generate images using Stable Diffusion checkpoint to overcome the problem of the local machin''s insufficient computing power. Using AI-generating images requires high-end GPUs, but unfortunately, not every student can afford the high-end equipment.

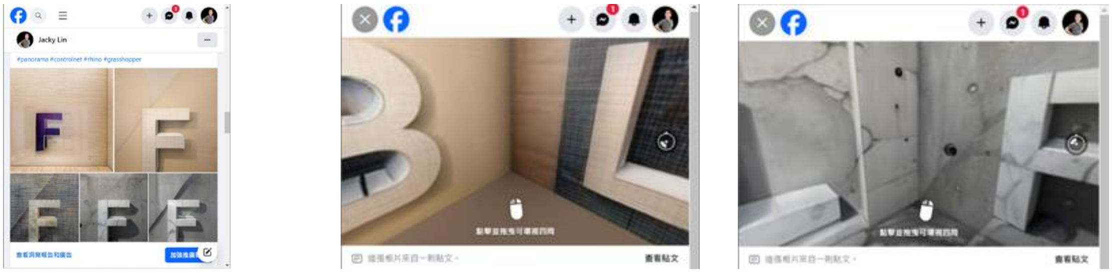

Fig. 3: Sharing panorama on Facebook: (a) Serial panorama album (Left), (b) operating a panorama on PC by mouse (Middle), and (c) quickly switching to another panorama (Right).

Display and Sharing the Generated Panoramas in VR Devices

As mentioned above, the easiest and fastest way to present the visual effects of panorama as virtual reality on smartphones, tablets, and VR devices is through the Facebook panorama sharing function. However, if the AI-generated panorama is uploaded directly to the Facebook website, it will not produce a panoramic visual effect. Since Facebook panorama sharing function is originally suitable for sharing panoramic pictures taken by a panoramic camera, AI-generated panoramas need to forge their camera source, that is, the Make and Model in EXIF metadata such as Ricoh THETA S. Secondly, due to the restrictions of the Facebook website, it cannot upload panoramas to Facebook through a personal computer. Still, it must be uploaded through the Facebook App by a smartphone to recognize a panorama. The EXIF metadata of the generated panorama can be modified when the file is saved by applying the Python 3 scripts. However, uploading via a smartphone still affects the efficiency of sharing results generated form a PC. Currently, a more convenient way on PC is to use a cloud drive that can be synchronized with a smartphone, such as One Drive, Google Drive or iCloud to synchronize the panorama files with the smartphone, and then upload them to Facebook for sharing. Not limited to expensive and cumbersome VR devices, sharing on Facebook may be more suitable for discussions and remote communication in design studios (Fig. 3).

Conclusions:

With the launch of cheaper all-in-one VR devices such as Meta Quest 2 and 3, sharing VR content will become easier. The next difficulty will be how to produce VR content quickly. Compared with traditional VR production, which has cumbersome processes and the manipulation is limited by high-end 3D computing equipment, this research uses quickly AI-generate panoramas, which can sufficiently display spatial effects on relatively cheap VR devices and smartphones to reduce hardware requirements for sharing VR contents. In addition, compared with traditional VR production, which can only operate a group of models at a time to make partial modifications, sharing multiple panoramas at once through Facebook and quickly switching among them is more suitable for creative ideations and discussions further to inspire new ideas and confirmation of design concepts. By combining Stable Diffusion, ControlNet, and the panoramic LoRA with Grasshoppe"s parametric modeling capabilities, this research demonstrates a method to rapidly generate VR scenes that align with existing conceptual interior spaces and design ideas by textual prompts. However, AI-generated panoramas have the potential to generate further 3D models, such as back-projecting to 3D prototypes to improve their quality or using ControlNet to obtain more image features such as depth and normal maps, and then to generate 3D models based on those features. These applications are still under development due to time constraints. It is scheduled to continue to explore more applications of AI-generated images combined with VR and generative models in interior design in future research.

Acknowledgments:

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

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

References:

[1] Dortheimer, J.; Schubert, G.; Dalach, A.; Brenner, L.; Martelaro, N.: Think AI-side the Box! Exploring the Usability of Text-to-Image Generators for Architecture Students, in: eCAADe 2023, 2023, 567-576. https://doi.org/10.52842/conf.ecaade.2023.2.567

[2] Faraj Al-Suwaidi, M.; Agkathidis, A.; Haidar, A.; Lombardi, D.: Immersive Technologies in Architectural Education: A pedagogical framework for integrating virtual reality as the main design tool in a fully virtualised architectural design studio environment, in: eCAADe 2023, 2023. https://doi.org/10.52842/conf.ecaade.2023.1.089

[3] Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models, in, 2021, arXiv:2106.09685. 10.48550/arXiv.2106.09685

[4] Lin, C.-J.: Plan1-2-3: Developing LOD-Based Parametric Floor Plan of Interior Design for Virtual Design Studio, in: CAD'23, 2023, 248-253. https://doi.org/10.14733/cadconfP.2023.248-253

[5] Stigsen, M.B.; Moisi, A.; Rasoulzadeh, S.; Schinegger, K.; Rutzinger, S.: AI Diffusion as Design Vocabulary, in: eCAADe 2023, 2023, 587-596.

[6] Zhang, L.; Rao, A.; Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, 3836-3847. https://doi.org/10.1109/ICCV51070.2023.00355