Pseudo-3D Scene Modeling for Virtual Reality Using Stylized Novel View Synthesis

Kuan-Wei Tseng*
National Taiwan University
Tokyo Institute of Technology
Taipei, Taiwan
Tokyo, Japan
kuanwei@g.ntu.edu.tw

Jing-Yuan Huang*
Yang-Sheng Chen
National Taiwan University
Taipei, Taiwan
jingyuanhuangg@gmail.com
yeti0193275@gmail.com

Chu-Song Chen
Yi-Ping Hung
National Taiwan University
Taipei, Taiwan
chusong@csie.ntu.edu.tw
hung@csie.ntu.edu.tw

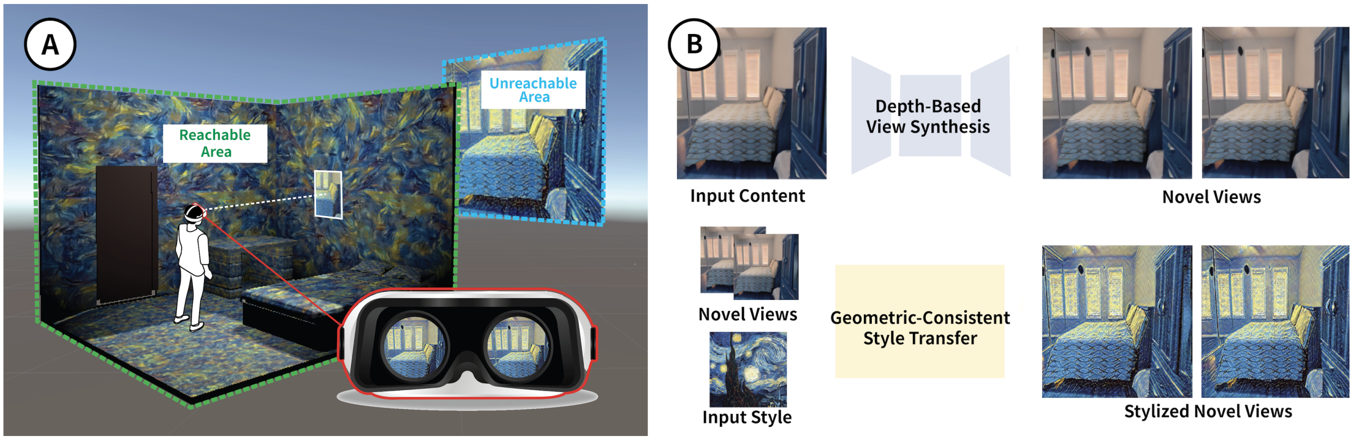

Figure 1: Scenario and method overview. We propose modeling the unreachable area in VR with stylized novel view synthesis as an alternative to actual 3D model or 360 image. A: users can see a stereoscopic image pair synthesized from the monocular content in VR. B: the pipeline of ArtNV[Tseng et al. 2022] applied to our system for novel view generalization and stylization. It integrates depth-based view synthesis and stereoscopic style transfer to generate the left- and right-eye views of VR glasses.

ABSTRACT

Stylized Novel View Synthesis is an emerging technique that combines style transfer and view synthesis. However, none of the existing works explore their applications in Virtual Reality (VR). This work devises a novel application for stylized novel view synthesis. We propose to replace actual 3D scene models or 360 images with stylized stereoscopic images for the areas outside the major play area but are still visible to the user. User study results reveal that users can feel 3D sense and tell them from plane texture. Codes and other materials are available at: kuan-wei-tseng.github.io/ArtNV

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; **Image-based rendering**; **Virtual reality**.

KEYWORDS

View Synthesis, Style Transfer, Virtual Reality (VR)

ACM Reference Format:
Kuan-Wei Tseng, Jing-Yuan Huang, Yang-Sheng Chen, Chu-Song Chen, and Yi-Ping Hung. 2022. Pseudo-3D Scene Modeling for Virtual Reality Using Stylized Novel View Synthesis. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '22 Posters), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532719.3543232

1 INTRODUCTION

Recent advances in novel view synthesis and stylization have shown great potential for applications in 3D displays. Given the content image and a style template, it aims to generate unseen novel views at specified camera poses with designated artistic styles.

To integrate both view synthesis and neural style transfer, a primary challenge is the spatial consistency between views. That is, if the 2D points on different views correspond to the same 3D point, they must have the same style transfer. Previous works handle the consistency by training a shared 3D representation in an implicit [Chiang et al. 2022] or explicit [Huang et al. 2021] manner. Though achieving great consistency, they require multi-view inputs and is time consuming. Mu et al. [Mu et al. 2022] propose a 3D photo stylization model that can generate stylized novel views with

*Both authors contributed equally to this research.

SIGGRAPH '22 Posters, August 07-11, 2022, Vancouver, BC, Canada
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9361-4/22/08.
https://doi.org/10.1145/3532719.3543232

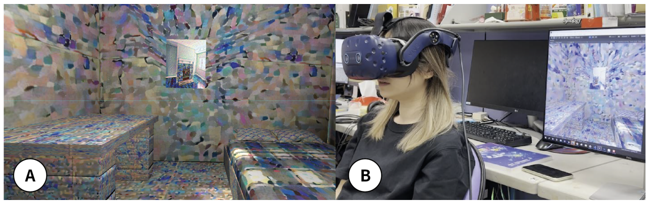

Figure 2: First/third-person view of the proposed system.

only a content image and a style template as input. However, their applicability has not been addressed in VR yet.

Constructing virtual scenes is essential to VR, and it usually requires 3D models or 360 images. Even for unreachable places, we still need to create the scene as the user's sight can still reach it. In this work, we propose to use stylized novel views as an alternative for scene modeling in VR. For the unreachable areas, we can display synthesized stereoscopic images in the VR glasses to make the users perceive 3D. Moreover, the neural style transfer can ensure the reachable and unreachable areas to share the same artistic styles. In detail, we adopt our previous work, ArtNV [Tseng et al. 2022], as the stylized novel view synthesis model. It is a two-stage pipeline consisting of view synthesis and geometric-consistent style transfer stages. Our approach integrates Synsin [Wiles et al. 2020], a view synthesis model , with online neural style transfer [Gatys et al. 2016] using optical flow-based dense matching to achieve stereoscopic consistency. It is noteworthy that our method only needs a single image as content input. Creators can simply take an arbitrary content image to construct a pseudo-3D scene that has same style with the texture of interactable objects. User study experiments reveal that the proposed method allows users to feel 3D senses in a virtual world effectively.

2 DESIGN AND IMPLEMENTATION

2.1 Virtual Scene

We use Unity to construct virtual scenes. The scene is divided into two parts: reachable area and unreachable area. Without loss of generality, as a example shown in Fig. 1, the play area is a room of size 4m*4.5m*5.2m with several pieces of furniture and a see-through window in VR. We use different textures for each piece of the furniture, such as cotton quilts and wood-grain tabletops. Style transfer is performed on these texture maps before attaching them to the furniture. A window is placed on the wall of the room, and its size fits for most of the users to see the unreachable area. Outside the window, the user can see stylized stereoscopic images. Specifically, we customize the shading method in Unity to render the left and right eye's images in a quad close to the window, respectively.

2.2 Stylized Novel View Synthesis

We utilize the novel view synthesis and stylization pipeline in ArtNV. Given the baseline between two views, we first generate a stereo image pair using SynSin, a depth-based novel view synthesis approach. Next, we estimate the dense optical flow between the image pair using RAFT [Teed and Deng 2020] with the occlusion maps obtained by the forward-backward consistency check. We leverage the dense correspondence to construct a consistency loss $\mathcal{L}_{corr}$ for constraining stylized outputs to have similar colors in the overlapping regions. Finally, we perform image optimization-based neural style transfer with the following objectives:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{content} + \lambda_2 \mathcal{L}_{style} + \lambda_3 \mathcal{L}_{corr}, \tag{1}$$

where $\mathcal{L}_{content}$ and $\mathcal{L}_{style}$ are the content and style losses computed with pretrained VGG-19 models, respectively. The combination coefficients are set as $\lambda_1 = 1e6$, $\lambda_2 = 1$, and $\lambda_3 = 3000$.

3 USER STUDY

To evaluate our method, we recruit 10 users (5 males and 5 females) to conduct a user study. We compare our method with that of placing a planar image outside the window with a distance of 2.5m. The users will see a scene with five different styles. For each style, we show the result of our method and the planar-image method in a random order and ask the users which one is more stereoscopic. The results show that 9 out of 10 people prefer our method. We also asked users to explain their criteria. Users commented that "I can tell whether the image is planar if it does not follow my movement naturally." (P1, P2, P3) and "I feel dizzy if the image is stereoscopic" (P5, P9, P10). We further asked users whether they feel discomfort or weird during the experiments. Besides motion sickness, users also responded that they feel some results are defocused. This is probably because the baseline of stereoscopic images is not customized for each testee, which may be improved by pupillary distance calibration before using our system.

4 DISCUSSION AND FUTURE WORKS

We introduce a novel application of stylized novel view synthesis to VR scene modeling. It is an easy-to-implement and low cost alternative to the actual 3D models or 360 images for the unreachable area in the virtual scene. Creators can take a content image and a style template to achieve congruent user experience on both stereoscopy and style. As for the future work, we aim to extend our method to model widely the visible area to increase its utility.

ACKNOWLEDGMENTS

This work was partially supported by Ministry of Science and Technology in Taiwan (MOST 110-2221-E-002-185-MY2 and MOST 111-2218-E-002-020-MBK).

REFERENCES

Pei-Ze Chiang, Meng-Shiun Tsai, Hung-Yu Tseng, Wei-Sheng Lai, and Wei-Chen Chiu. 2022. Stylizing 3D Scene via Implicit Representation and HyperNetwork. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1475–1484.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hsin-Ping Huang, Hung-Yu Tseng, Saurabh Saini, Maneesh Singh, and Ming-Hsuan Yang. 2021. Learning to Stylize Novel Views. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Fangzhou Mu, Jian Wang, Yicheng Wu, and Yin Li. 2022. 3D Photo Stylization - Learning to Generate Stylized Novel Views from a Single Image. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zachary Teed and Jia Deng. 2020. Raft: Recurrent all-pairs field transforms for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 402–419.

Kuan-Wei Tseng, Yao-Chih Lee, and Chu-Song Chen. 2022. Artistic Style Novel View Synthesis Based on A Single Image. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. 2020. SynSin: End-to-End View Synthesis From a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.