

Immersive Real World through Deep Billboards

Naruya Kondo
n-kondo@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba, Japan

So Kuroki
kuroki-so273@g.ecc.u-tokyo.ac.jp
The University of Tokyo, Japan

Ryosuke Hyakuta
momosuke@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba, Japan

Yutaka Matsuo
matsuo@weblab.t.u-tokyo.ac.jp
The University of Tokyo, Japan

Shixiang Shane Gu
shanegu@google.com
Google Brain, USA

Yoichi Ochiai
wizard@slis.tsukuba.ac.jp
University of Tsukuba, Japan



Figure 1: Interactive Deep Billboards with a NeRF model (left three) and a World model (right most) in VR.

ABSTRACT

An aspirational goal for virtual reality (VR) is to bring in a rich diversity of real world objects *losslessly*. Existing VR applications often convert objects into explicit 3D models with meshes or point clouds, which allow fast interactive rendering but also severely limit its quality and the types of supported objects, fundamentally upper-bounding the “realism” of VR. Inspired by the classic “billboards” technique in gaming, we develop Deep Billboards that model 3D objects *implicitly* using neural networks, where only 2D image is rendered at a time based on the user’s viewing direction. Our system, connecting a commercial VR headset with a server running neural rendering, allows real-time high-resolution simulation of detailed rigid objects, hairy objects, actuated dynamic objects and more in an interactive VR world, drastically narrowing the existing real-to-simulation (real2sim) gap. Additionally, we augment Deep Billboards with physical interaction capability, adapting classic billboards from screen-based games to immersive VR. At our pavilion, the visitors can use our off-the-shelf setup for quickly capturing their favorite objects, and within minutes, experience them in an immersive and interactive VR world – with minimal loss of reality. Our project page: <https://sites.google.com/view/deepbillboards/>

KEYWORDS

image based rendering, neural networks, billboard

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '22 Immersive Pavilion, August 07-11, 2022, Vancouver, BC, Canada

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9369-0/22/08.

<https://doi.org/10.1145/3532834.3536210>

ACM Reference Format:

Naruya Kondo, So Kuroki, Ryosuke Hyakuta, Yutaka Matsuo, Shixiang Shane Gu, and Yoichi Ochiai. 2022. Immersive Real World through Deep Billboards. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Immersive Pavilion (SIGGRAPH '22 Immersive Pavilion)*, August 07-11, 2022. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3532834.3536210>

1 INTRODUCTION

Data-driven 3D modeling is increasingly in demand from the VR industry, and is essential for reality-grounded VR applications such as shopping, showrooms, tourism, and teleconferences that require transporting real objects into the interactive virtual world. Traditionally, VR uses explicit 3D models such as meshes and point clouds for fast simulation, but they have limited rendering quality and thus fundamentally upper-bound the “realism” of current VR worlds. Inspired from classic “2D” billboards in gaming, we propose Deep Billboards and develop a novel system that drastically enhances real-to-simulation (real2sim) quality of complex objects in interactive VR.

2 DEEP BILLBOARDS

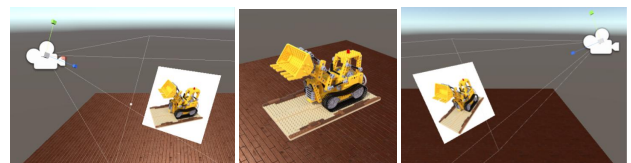


Figure 2: Deep Billboard object on a single canvas.

A “billboard” in computer graphics and 3D video games corresponds to a 2D image that is tilted to always face the observer,

creating an illusion of a 3D object. We propose Deep Billboard, a generalization of the classic billboard, where the 2D texture is re-generated at each frame from an implicit neural rendering model outputting the view-conditioned images (Figure. 2). Compared to standard explicit 3D models such as meshes or point clouds used in VR, Deep Billboards enable object rendering of much higher resolutions, drastically improving the realism of virtual experiences while preserving real-time interactivity. Any implicit neural rendering model can be used for our Deep Billboard (Figure. 3).



Figure 3: Quality of NeRF Billboard (top) and dynamic rendering of World Billboard (bottom).

NeRF Billboard. Neural radiance field (NeRF) has revolutionized the field of data-driven 3D reconstruction [Mildenhall et al. 2020]. While NeRF allows a much higher-resolution 3D construction than classic approaches, and requiring only tens of 2D images of the object from multiple angles. Prior work importing NeRF into VR converting the models back to coarse meshes, due to software incompatibility, high computational demand, and slow rendering of original NeRF models. Our system instead used PlenOctrees [Yu et al. 2021], an extension of NeRF to allow real-time rendering, running on a remote GPU server, to directly use NeRF to update the billboards. To the best of our knowledge, our system is the first to import NeRF models directly into an interactive VR application without loss of accuracy.

World Billboard. While NeRF allows rich rendering of static objects, extension to dynamic scenes is still an active area of research. To show the versatility of our system, we also replace our billboard updater using PlaNet [Hafner et al. 2019], a neural state-space model (SSM) or a *world* model for action-conditioned video prediction in deep reinforcement learning (RL), to achieve data-driven reproduction of time-varying objects from a single video only. From a single 10-minute video labeled with the camera position and orientation, we learned video prediction conditioned on the initial viewpoint image and viewpoint series, and reproduced the time-varying 3D object through the billboard in our interactive VR system.

Physics Interaction. Another disadvantage of a classic billboard is that it does not allow physical interaction. To solve this problem, we propose a system where a billboard is used for visual interaction



Figure 4: Deep Billboard provides basic physical interaction.

while an invisible rough mesh for physics interaction. In Figure 4, we used a mesh extracted from a NeRF model and realized three basic physical interactions. This makes our Deep Billboard to achieve both good visuals and quality physics.

3 SYSTEM

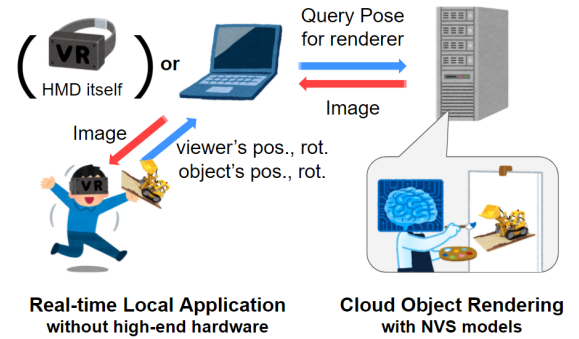


Figure 5: Deep Billboard VR system overview.

Deep generative models require large computational power, so we have built a system that uses cloud rendering to enable even inexpensive VR headsets to handle highly accurate pseudo-3D models as long as they have an internet connection (Figure. 5). Each Deep Billboard object is rendered on a cloud server, and rendered 2D frame is sent to the VR system per frame to achieve real-time interaction with minimal on-board processing. Our system transported a wide range of objects, including actuated toys, hairy plushy to the interactive virtual world with minimal loss of reality.

ACKNOWLEDGMENTS

This work was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan and New Energy and Industrial Technology Development Organization (NEDO) of Ministry of Economy, Trade and Industry of Japan.

REFERENCES

- Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. 2019. Learning Latent Dynamics for Planning from Pixels. In *International Conference on Machine Learning*. 2555–2565.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.
- Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *ICCV*.