360RVW: Fusing Real 360° Videos and Interactive Virtual Worlds

Mizuki Takenawa
The University of Tokyo
Tokyo, Japan
takenawa@hal.t.u-tokyo.ac.jp

Naoki Sugimoto
MMMakerSugi
Tokyo, Japan
naoki48916@gmail.com

Leslie Wöhler
The University of Tokyo JSPS
International Research Fellow
Tokyo, Japan
woehler@hal.t.u-tokyo.ac.jp

Satoshi Ikehata
National Institute of Informatics
Tokyo, Japan
sikehata@nii.ac.jp

Kiyoharu Aizawa
The University of Tokyo
Tokyo, Japan
aizawa@hal.t.u-tokyo.ac.jp

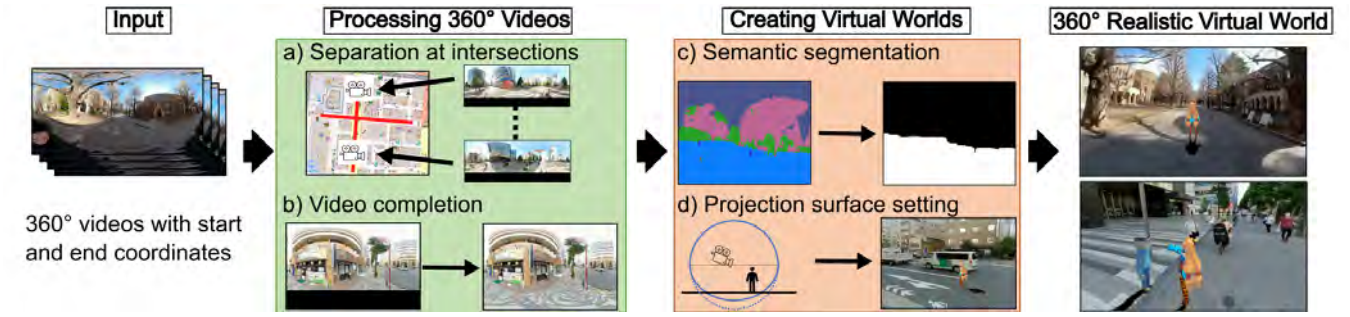

Figure 1: We propose a system to create 360° realistic virtual worlds (360RVW) from omnidirectional street videos that allow scene exploration and user interactions.

ABSTRACT

We propose a system to generate 360° realistic virtual worlds (360RVW) for the interactive spatial exploration of omnidirectional street-view videos. Our 360RVW enables users to explore photorealistic scenes with digital avatars, and interact with others. To create the virtual worlds our system only requires 360° videos with annotations of the start and end camera coordinate as input. We first detect street intersections to divide the input videos and remove the camera operator from the recordings using a video completion technique. Next, we analyze the 3D structure of the scene using semantic segmentation to define walkable areas. Finally, we render the environment using an ellipsoid projection surface to achieve a more realistic integration of the avatar into real-world 360° videos. The whole process is largely automated, enabling users to produce realistic and interactive virtual worlds without specialized skills or time-consuming manual interventions.

CCS CONCEPTS

• **Information systems** → **Multimedia content creation**; • **Computing methodologies** → *Virtual reality*.

KEYWORDS

omni-directional video, virtual reality, interface

ACM Reference Format:
Mizuki Takenawa, Naoki Sugimoto, Leslie Wöhler, Satoshi Ikehata, and Kiyoharu Aizawa. 2023. 360RVW: Fusing Real 360° Videos and Interactive Virtual Worlds. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3581783.3612670

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10.
https://doi.org/10.1145/3581783.3612670

1 INTRODUCTION

In recent years, virtual worlds have allowed users to connect with others and explore locations from all over the world. Even large real-world gatherings have successfully been transferred to online events with a Halloween celebration held in a virtual recreation of Shibuya attracting over 400.000 visitors [9]. Unfortunately, creating such virtual worlds is time-consuming and requires specialized skill to design convincing 3D environments. Moreover, there are limits to how well 3D models can reproduce the real world because living things or moving cars don't exist. To address this challenge, we use 360° videos instead of building complex 3D models. Unlike 3D models, 360° videos can easily be created due to the widespread availability of 360° cameras and convey a realistic impression of the environment, however, they are generally non-interactive and do not allow free spatial exploration. Therefore, in this paper, we seek to generate photorealistic and interactive low-cost virtual worlds by fusing virtual 3D environments and 360° videos.

To avoid the manual creation of 3D worlds, many systems using 360° media have been proposed. In Google Street View (GSV) [5],

360° photos are used to give users an impression of the environment, however, the images are only taken at discrete positions and do not offer interactivity. To improve the experience, Geollery [1] combines images from GSV with social media functions and 3D city models. Considering 360° videos, Movie Map [8] proposed a system to create a connected world from 360° street-view videos enabling users to navigate the world by choosing a walking direction at intersections. Other works focus on how to incorporate user interactions in 360° videos considering video streaming [6] or shared viewing experiences [3]. Furthermore, Tourgether360 [4] proposes a collaborative interface that displays the camera trajectory and the eye of users' avatars to indicate their viewing direction in a 360° video landscape, increasing the spatial awareness and social presence of users.

In this paper, we enrich omnidirectional videos to create 360° realistic virtual worlds (360RVW) by adding free spatial exploration via digital avatars. In contrast to previous works, users can move their avatars in a 3D representation of the real-world video which includes walking people and moving cars and interact directly with others. Furthermore, our system only requires 360° videos and the start and end camera coordinate as input to afterwards automatically generate virtual environments without manual intervention.

2 SYSTEM

Our system enables the automatic generation of virtual environments and only requires 360° videos with defined start and end coordinates as input (Fig. 1).

2.1 Processing the 360° Videos

To separate the input videos for each road segment, we process them following Movie Map [8]. First, the relative camera positions and rotations are estimated with OpenVSLAM [10] in order to assign camera coordinates to each frame based on the start and end coordinates of the videos (Fig. 1 (a)). Next, the coordinates and visual features are used to detect pairs of video frames corresponding to each intersection seen in the videos. Finally, we divide the videos by the intersection frames. This approach enables users to freely explore the virtual world rendered from the videos by changing their direction of movement at road intersections.

In Movie Map [8] the camera operator is excluded from the video by blacking out the ground area. As we need to include a representation of the ground to enable the navigation via avatars, we add an additional processing step by applying the video completion model FGCV [2] before dividing the videos. This module can remove the camera operator from the videos to ensure a natural representation of the floor in the resulting virtual environment (Fig. 1 (b)).

2.2 Creating Virtual Worlds from 360° Videos

Our system uses two main functions to generate a virtual environment from 360° videos. First, we apply semantic segmentation using Orhans' model [7] to obtain labels for each pixel indicating the position of roads and buildings. Based on this, we can construct a binary map that encodes which areas constitute the ground of the scene (Fig. 1 (c)). This enables us to determine whether or not avatars are on the road by projecting the position of the avatar to the coordinate in the binary map. Second, to naturally integrate the avatar into the scene, our system changes the projection from the regular sphere projection to a shallow-bottomed ellipse (Fig. 1(d)). In addition, we set the ground plane to the position corresponding to the road in the scene. With these improvements, as shown in Fig. 2, the avatar appears as if it were in a real 3D scene while a spherical projection results in the avatar appearing to float.

Figure 2: The effect of the projection: (a) In a spherical projection, the avatar seems to float. (b) In the proposed elliptical projection, the avatar appears as a part of the scene.

Figure 3: Applications: (a) Exploring famous tourist attractions. (b) Real-time interaction with other users' avatars.

2.3 System Implementation

Our system handles the processing of 360° videos and generation of the virtual environment as an offline preprocessing step. To display the scenes, we use Unity [11] and display the environments and avatars in real time. Here, the videos are played back according to the forward movement speed of the avatar allowing the user to experience the scene as if the avatar is walking through the video. Moreover, since our system allows users to explore the videos at the same time, they can not only interact with the environment but also with each other. Overall, users are able to experience the collaborative spatial exploration of 360° videos. In our demo, we will showcase locations of Japan including famous sightseeing spots and busy shopping districts.

3 CONCLUSION

In this work, we propose a system to generate new virtual environments from 360° videos that have both a realistic appearance and interactive content. Our system extends 360° videos to realistic virtual worlds by applying segmentation to define explorable areas and uses video completion models as well as a shallow-bottomed ellipse projection to convincingly render avatars in the scene. As shown in Fig. 3, users can freely enjoy viewing and navigating real locations and interact with other users' avatars. We believe that our system will contribute to the development of realistic and immersive multi-user interactive virtual worlds, which can accelerate the development of the metaverse as an emergent online framework.

ACKNOWLEDGMENTS

This research was partly supported by JST-Mirai Program JPMJMI21H1 and JSPS KAKENHI 21H03460.

REFERENCES

[1] Ruofei Du, David Li, and Amitabh Varshney. 2019. Geollery: A Mixed Reality Social Media Platform. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300915

[2] Chen Gao, Ayush Saraf, Jia-Bin Huang, and Johannes Kopf. 2020. Flow-edge Guided Video Completion. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, Cham, 713–729. https://doi.org/10.1007/978-3-030-58610-2_42

[3] Qiao Jin, Yu Liu, Ruixuan Sun, Chen Chen, Puqi Zhou, Bo Han, Feng Qian, and Svetlana Yarosh. 2023. Collaborative Online Learning with VR Video: Roles of Collaborative Tools and Shared Video Control. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 713, 18 pages. https://doi.org/10.1145/3544548.3581395

[4] Kartikaeya Kumar, Lev Poretski, Jiannan Li, and Anthony Tang. 2022. Tourgether360: Collaborative Exploration of 360° Videos Using Pseudo-Spatial Navigation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2, Article 546 (November 2022), 27 pages. https://doi.org/10.1145/3555604

[5] Google LLC. 2023. Google Street View. Retrieved June 21, 2023 from https://www.google.com/streetview/

[6] Alaeddin Nassani, Li Zhang, Huidong Bai, and Mark Billinghurst. 2021. ShowMeAround: Giving Virtual Tours Using Live 360 Video. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 168, 4 pages. https://doi.org/10.1145/3411763.3451555

[7] Semih Orhan and Yalin Bastanlar. 2022. Semantic segmentation of outdoor panoramic images. *Signal, Image and Video Processing* 16, 3 (2022), 643–650. https://doi.org/10.1007/s11760-021-02003-3

[8] Naoki Sugimoto, Yoshihito Ebine, and Kiyoharu Aizawa. 2020. Building Movie Map - A Tool for Exploring Areas in a City - and Its Evaluations. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20)*. ACM, New York, NY, USA, 3330–3338. https://doi.org/10.1145/3394171.3413881

[9] Ayumi Sugiyama. 2021. Virtual Shibuya Halloween event ups its game with personal avatars. Retrieved June 21, 2023 from https://www.asahi.com/ajw/articles/14460054

[10] Shinya Sumikura, Mikiya Shibuya, and Ken Sakurada. 2019. OpenVSLAM: A Versatile Visual SLAM Framework. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. Association for Computing Machinery, New York, NY, USA, 2292–2295. https://doi.org/10.1145/3343031.3350539

[11] Unity Technologies. 2023. Unity. Retrieved June 21, 2023 from https://unity.com/