香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. XING Wenpeng
Date	14 May 2024 (Tuesday)
Time:	2:30 pm – 4:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	1) RRS 732, 7/F, Sir Run Run Shaw Building, Ho Sin Hang Campus 2) ZOOM (Meeting ID: 997 8589 7113) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 12:00 nn, 13 May 2024)

Towards Efficient and High-quality 3D Scene Reconstruction and Novel View Rendering

Abstract

To efficiently model 3D scene content and faithfully render novel-view images from learned representations is an essential task for various applications in extended reality (XR), robotics, and virtual production. This thesis focuses on volumetric scene representations, which are capable of accurately modelling the geometry and view-dependent reflectance from multi-view observations due to their volumetric capacity and differentiable nature. We will comprehensively investigate the challenges in volume reconstruction (diverse baselines and content dynamics), optimization, rendering, computational complexity, and memory efficiency, and introduce our solutions centered around this representation.

First, to address the challenges of volume reconstruction from large-baseline multi-views, we introduce the Scale-Consistent Volume Rescaling (SCVR) algorithm to align the disparity probability volumes (DPV) for consistent geometry fusion across sparse Light Fields, enabling high-quality novel view rendering and accurate disparity inference over the Stanford Lytro Multi-View LF dataset.

Second, recognizing the extensive scene-specific training required by traditional neural volumetric representations, the proposed MVSPlenOctree using variance-based features for Multi-View Stereo inference, achieves generic radiance field estimation and introduces an occlusion-robust sampling strategy for more effective volumetric reconstructions. Experiments on the DTU, LLFF, and Synthetic-NeRF datasets prove its effectiveness.

Third, to address the data-intensive needs of volumetric scene representation, we present two novel volume compression methods. The first, Temporal-MPI, encodes 3D dynamics across video sequences into compact temporal bases and coefficients, allowing for quick generation of Time-instance MPI within milliseconds. This method, evaluated on Nvidia's Dynamic Scene Dataset, proves faster and more efficient than existing dynamic scene models. The second technique employs a multi-scale tensor decomposition approach, consisting of Multi-Scale Vector-Matrix Decomposition (MSVMD) and Multi-Scale CANDECOMP/PARAFAC Decomposition (MSCPD), to effectively compress static volumetric fields and reduce storage complexity. Tested on the Synthetic-NeRF Dataset, it consistently surpasses single-scale tensor decomposition methods, especially when using a minimal number of components.

Finally, the thesis investigates the integration of pre-trained diffusion models for novel view synthesis by incorporating conditioning camera parameters, discussing future research directions for 3D content generation and view synthesis in the AIGC era.

***** ALL INTERESTED ARE WELCOME *****