

네트워크 관점에서의 3D 콘텐츠 표현 방식 분석

유 상 우*, 홍 원 기[°]

Analysis on 3D Content Representations in Network Perspective

Sangwoo Ryu*, James Won-Ki Hong[°]

요 약

확장현실(XR) 기술의 발전으로 3D 콘텐츠의 중요성이 급격히 증가하고 있다. 이에 따라 3D 콘텐츠를 효율적으로 표현하고 전송하는 기술의 필요성이 대두되고 있다. 3D 콘텐츠는 다양한 방식으로 표현이 가능하며, 각 표현 방식마다 용량, 품질 등 다양한 부분에서 차이가 존재한다. 본 연구에서는 포인트 클라우드(Point Cloud), 메쉬(Mesh), NeRF(Neural Radiance Field), 3DGS(3D Gaussian Splatting), 총 네 가지의 주요 3D 콘텐츠 표현 방식을 대상으로, 동일한 콘텐츠를 각 표현 방식으로 제작하여 데이터 크기, 품질, 데이터 손실의 영향 등 네트워크 성능 특성을 분석하고, 서비스 및 네트워크 품질에 영향을 주는 추가 고려 요소들과 관련 연구 및 기술들을 제시한다. 이를 통해 다양한 형태의 3D 콘텐츠에 대한 이해를 제공하고, 더 나은 서비스 및 네트워크 품질 관리를 돕는 것을 목표로 한다.

Key Words : 3D Content Representations, Volumetric Video Streaming, Quality of Service

ABSTRACT

The rapid advancement of Extended Reality (XR) technology has significantly increased the importance of 3D content. Accordingly, there is a growing need for technologies that can efficiently represent and transmit 3D content. 3D content can be represented in various ways, with each representation method differing in aspects such as data size and quality. This study focuses on four major 3D content representation methods: Point Cloud, Mesh, Neural Radiance Field (NeRF), and 3D Gaussian Splatting (3DGS). In this research, we analyze the network performance characteristics of these four representation methods by creating the same content using each method and evaluating factors such as data size, quality, and the impact of data loss. Additionally, we present additional considerations that affect service and network quality, along with related research and technologies. The aim of this study is to provide a comprehensive understanding of various 3D content representation and to assist in improving service and network quality management.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음 (IITP-2024-RS-2024-00437866)

◆ First Author : Pohang University of Science and Technology (POSTECH), Graduate School of Artificial Intelligence, rswoo@postech.ac.kr

° Corresponding Author : POSTECH, Department of Computer Science and Engineering, jwkhong@postech.ac.kr

논문번호 : KNOM2024-02-05, Received November 18, 2024; Revised December 13, 2024; Accepted December 22, 2024

I. 서 론

최근 가상 현실(Virtual Reality, VR), 증강 현실(Augmented Reality, AR), 혼합 현실(Mixed Reality, MR)과 함께, 이를 모두 포함하는 확장현실(Extended Reality) 기술의 발전으로, 3D 콘텐츠의 중요성이 급격히 증가하고 있다. P&S 인텔리전스는 XR 시장 규모를 2022년 383억 달러에서 2030년 3,948 달러로 38.5% 성장을 전망했으며 [1], IDC는 AR/VR 기기 시장이 2023년부터 2027년까지 연평균 32.6% 성장할 것으로 전망했다 [2]. 이러한 확장 현실 공간에서는 사용자에게 다양한 형태의 3D 콘텐츠를 제공하며, 엔터테인먼트, 교육, 의료, 미디어 등 다양한 산업 분야에서의 활용이 확대되면서, 이를 효율적으로 표현하고 전송하는 기술의 필요성이 중요해졌다.

이러한 맥락에서 주목받고 있는 기술 중 하나가 볼류메트릭 영상(Volumetric Video)이다. 볼류메트릭 영상은 3차원 데이터로 구성된 영상으로, 확장현실 환경에서 3D 콘텐츠를 보여주기 위해 사용되고, 실제 세상의 물체, 사람, 움직임 등을 가상 세계로 가져와 사실적으로 보여준다. 하지만 볼류메트릭 영상과 같은 3D 콘텐츠는 일반적으로 3차원 공간에서의 위치, 색 등 대용량의 데이터를 가지고 있어, 사용자 및 서비스 제공자 모두에게 전송과 처리에 부담을 준다. 특히 실시간 스트리밍 및 대규모 사용자 환경에서의 서비스를 위해서는 이러한 전송 및 처리 과정을 효율적으로 수행할 수 있도록 하는 연구가 필수적이다.

기존에 3D 콘텐츠를 포인트 클라우드(Point Cloud) 혹은 메쉬(Mesh) 형태로 표현한 것에 더해, 최근에는 NeRF(Neural Radiance Field), 3DGS(3D Gaussian Splatting) 등 새로운 고품질 3D 데이터 표현 방식이 제안되고 있다. 이러한 콘텐츠 표현 방식의 다양화는 표현 방식 간 서로 다른 특성으로 인해 서비스 제공자가 고려해야 하는 요소를 더욱 복잡하게 만들고 있다. 대표적으로 각 표현 방식 별로 제작 과정, 품질, 렌더링 방식 및 속도, 데이터 크기 등이 다르며, 이에 따라 사용 시 나리오, 목표 플랫폼, 주로 사용될 네트워크 환경 등을 함께 고려해야 한다. 특히 네트워크 전송 측면에서의 고려사항들은 원활한 전송을 통한 사용자 경험 최적화, 컴퓨팅 자원 최적화를 통한 운영 비용 절감, 확장성 확보 등 전체적인 서비스 품질 향상과

직접적인 관련이 있기 때문에 서비스 제공자가 중요하게 고려해야 할 요소가 된다.

따라서 서비스 제공자가 더 나은 서비스를 제공하기 위해서는 네트워크의 관점에서 3D 콘텐츠 표현 방식들을 분석 및 이해하고, 적절한 표현 방식을 선택하는 것이 중요하다. 본 연구는 네트워크 관점에서 다양한 3D 콘텐츠 표현 방식을 분석하고, 각 방식의 장단점을 비교하는 것을 목적으로 한다. 특히 같은 콘텐츠를 포인트 클라우드, 메쉬, NeRF, 3DGS, 총 네 가지 표현 방식으로 제작하여 데이터 크기, 품질, 데이터 손실의 영향 등 고려 요소들을 분석하고 추가 고려 요소들 및 관련 연구, 기술들을 제시하여 3D 콘텐츠 표현 방식에 대한 이해를 돕고자 한다. 본 연구의 기여는 다음과 같다.

- 1) 같은 콘텐츠에 대한 서로 다른 3D 표현 방식의 네트워크 성능 특성에 대한 분석 결과를 제시
- 2) 각 3D 표현 방식의 고려 요소들에 대한 관련 연구 및 기술들을 제시

II. 배경 및 관련 연구

본 장에서는 본 논문에서 사용할 3D 데이터 표현 방식 네 가지(포인트 클라우드, 메쉬, NeRF, 3DGS)를 설명하고, 해당 표현 방식들의 대략적인 비교를 담고 있는 연구들을 소개한다.

1) 포인트 클라우드(Point Cloud)

포인트 클라우드는 3D 물체를 점들의 집합으로 구성한 데이터 구조이다. 각 점은 X, Y, Z 좌표와 추가적으로 색상 정보(R, G, B), 법선(Normal) 벡터 정보를 가질 수 있으며, 전체 데이터의 크기는 각 점을 이루는 정보들의 데이터 유형(실수, 8비트 정수 등)과 점의 개수로 결정된다.

2) 메쉬(Mesh)

메쉬는 3D 물체를 점들과 점들을 이어 모서리, 다각형 면을 구성한 데이터 구조이다. 메쉬는 각 점의 X, Y, Z 좌표와 선 혹은 면을 이루기 위한 점들의 연결 정보, 추가적으로 점 혹은 면의 색상, 법선 벡터 정보를 담을 수 있으며, 전체 데이터의 크기는 각 점, 선 면을 이루는 정보들의 데이터 유형과 점, 선 면의 개수로 결정된다.

3) NeRF(Neural Radiance Field)

NeRF는 2020년 3D 복원을 위해 제안된 데이터

표현 방식으로 [3], MLP(Multi Layer Perceptron, 다중 퍼셉트론)와 같은 신경망을 이용해 시선의 위치 및 방향에 따른 색과 밀도를 표현하는 구조이다. 기존 포인트 클라우드, 메쉬 등의 표현 방식들과 다르게 3D 공간 상에 데이터가 명시적(Explicit)으로 표현되지 않고, 암시적(Implicit)으로 표현되며, 데이터 크기는 구현 방식에 따라 다양한 요소(신경망 크기, 해시맵 크기 등)에 의해 결정된다.

4) 3DGS (3D Gaussian Splatting)

3DGS는 2023년 제안된 표현 방식으로 [4], 3D 물체를 3D 가우시안(Gaussian) 함수의 집합으로 구성한 데이터 구조이다. 각 가우시안은 X, Y, Z 좌표와 크기, 회전, 투명도 정보, 구면 조화 함수(Spherical Harmonics) 형태의 색 정보 등과 같은 속성을 가진다. 전체 데이터의 크기는 가우시안을 구성하는 속성들의 개수 및 데이터 유형, 그리고 가우시안의 총 개수로 결정된다.

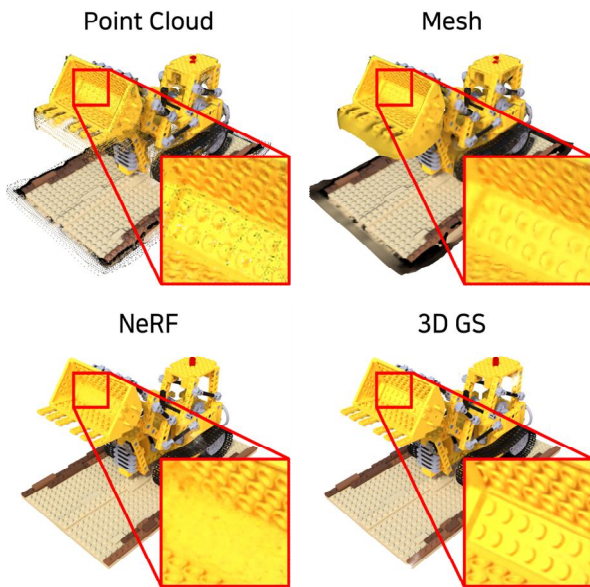


그림 1. 각 표현에 따른 렌더링 예시
Fig. 1. Rendering example of each 3D representation

그림 1은 앞서 설명한 네 가지 표현 방식에 대해 실제로 렌더링을 한 예시를 보여주고, 각 표현 방식의 시각적 특징을 확인할 수 있다. 포인트 클라우드의 경우 일부 빈 공간에 대한 노이즈를 확인할 수 있으며, 메쉬의 경우 적은 수의 면으로 인해 디테일이 뭉개진 모습을 볼 수 있다. NeRF와 3DGS는 학습 기반 방식으로, 전체적으로 높은 품질을 보여주지만, 학습이 완전히 되지 못한 경우에는 일부분이

제대로 표현되지 못한 모습도 확인할 수 있다.

이러한 다양한 3D 데이터 표현 방식에 대해 대략적인 비교를 한 사례가 일부 존재한다. 먼저 논문 [5]에서는 볼류메트릭 영상의 캡처부터 사용자에게 보여주는 부분까지 볼류메트릭 영상 서비스를 구성하는 모든 부분에 대해 관련 연구를 소개한다. 해당 논문에서는 데이터 표현 방식과 관련해서 포인트 클라우드, 메쉬, 복셀 (Voxel) 등 6종류의 표현 방식에 대해 크기, 품질, 자원 사용량 등을 높음, 중간, 낮음 등으로 제시하였다. 논문 [6]은 네트워크 관점에서 볼류메트릭 영상 서비스에 대한 개요, 구조, 예시 설계 등을 제시하였다. 데이터 표현 방식과 관련해서는 포인트 클라우드, 메쉬, V-PCC [31] 등 5종류의 표현 방식에 대해 크기를 높음, 중간, 낮음 등으로 제시하고 장단점을 간단하게 서술하였다. 논문 [7]은 볼류메트릭 영상 스트리밍에서 3DGS 사용을 위한 방법에 대한 연구이지만, 포인트 클라우드, 메쉬, NeRF, 3DGS, 총 4종류의 표현 방식에 대해 렌더링 품질, 속도, 접근성, 크기를 높음, 중간, 낮음 등으로 간단히 비교하는 부분을 포함하였다. 논문 [8]은 3D 복원을 위한 방식 3가지를 비교하는 연구로, 기존에 포인트 클라우드 혹은 메쉬를 만들기 위하여 사용하는 COLMAP[10, 11, 12] 방식과 NeRF, 3DGS 두 가지 새로운 방식을 비교해 유용성을 평가하였다. 논문 [9]는 볼류메트릭 영상 스트리밍에서 NeRF와 같은 신경망 기반 표현 방식들을 사용하기 위해 고려해야 하는 사항들과 이를 위한 사전 실험 결과들을 제시하였다. 논문 [10]은 빠르고 다양하게 제안되는 NeRF 및 3DGS 기반 3D 복원 방식 기술들을 테스트하기 위한 통합 프레임워크를 제안하며, 여러 데이터셋에 대해 시각 품질, 훈련 시간, 메모리 사용량 등에 대한 정량적인 수치를 제공한다.

이와 같이 기존 연구들이 최종 렌더링 품질을 주로 평가하였다면, 본 연구에서는 네트워크 전송 측면을 함께 고려하여 각 표현 방식에 대한 분석 결과와 관련 연구, 기술을 제시하는 것을 목표로 한다.

III. 방법

이 장에서는 한 콘텐츠에 대해 여러 표현 방식의 모델을 여러 크기/품질로 제작하기 위한 방법과, 이를 분석하기 위한 방법을 서술한다.

1. 데이터 준비

한 콘텐츠에 대해 여러 표현 방식으로 모델을 제작하기 위하여, 같은 멀티 뷰 데이터로부터 3D 복원을 수행하는 방법들을 사용하였다. 이를 위해 NeRF [3] 논문에서 함께 제안한 NeRF Synthetic 혹은 Blender 데이터셋의 Lego 씬의 훈련 데이터셋을 이용하였다. Lego 씬은 100장의 훈련 이미지와 200장의 테스트 이미지를 포함하고 있어, 3D 모델 복원을 위해 훈련 이미지만 사용하였다.

포인트 클라우드와 메쉬를 제작하기 위해서는 COLMAP [10, 11, 12]을 사용하였다. COLMAP은 멀티 뷰 이미지로부터 포인트 클라우드 혹은 메쉬를 복원하는 대표적인 방법 중 하나이다. 100장의 훈련 이미지와 COLMAP을 이용해 고품질의 메쉬를 복원하였으며, 해당 메쉬는 Open3D [14] 라이브러리를 이용해 다운샘플링을 수행하여 저화질의 메쉬를 제작하였다.

포인트 클라우드는 고품질 메쉬에서 꼭짓점(Vertex)만 추출하여 복원했으며, Open3D 라이브러리를 이용해 점의 개수를 줄이는 다운샘플링(Downsampling)을 수행하여 저화질의 포인트 클라우드를 제작하였다. 각 품질에 따른 포인트 클라우드와 메쉬의 점, 면 개수는 유사한 데이터 크기 범위를 위해 실험적으로 조정하였다.

표 1, 표 2는 각각 제작된 포인트 클라우드와 메쉬의 점, 면 개수와 데이터 크기를 보여준다. 해당 표의 데이터 크기는 압축이 적용되지 않은 미가공(Raw) 상태의 PLY 파일 데이터 크기이다. Gbps@30Hz는 영상 스트리밍에 사용하는 것을 가정하여, 초당 30프레임의 속도로 해당 파일을 전송하는 경우 사용되는 대역폭이다.

표 1. 여러 품질로 제작된 포인트 클라우드 정보
Table 1. Point cloud information with different quality

Level	# of Points	Size (MB)	Gbps @30Hz
0	102,224	1.462	0.343
1	302,003	4.320	1.013
2	599,129	8.571	2.009
3	885,307	12.664	2.968

표 2. 여러 품질로 제작된 메쉬 정보
Table 2. Mesh information with different quality

Level	# of Points	# of Faces	Size (MB)	Gbps @30Hz
0	37,395	75,720	1.474	0.345
1	110,534	221,918	4.332	1.015
2	222,568	445,842	8.711	2.042
3	322,231	644,697	12.602	2.954

NeRF와 3DGS 모델을 제작하기 위해서는 Nerfstudio [15]를 사용하였다. Nerfstudio는 NeRF 및 3DGS 기반 방법들을 수행하고 테스트할 수 있는 API를 제공하는 프레임워크이다. NeRF 모델을 만들기 위해서는 Nerfstudio에서 제공하는 nerfacto를 사용하였다. nerfacto는 좌표를 해시맵을 이용해 효율적으로 표현하는 해시 인코딩 등 Instant-NGP [16]와 같은 이전 연구들에서 효율적인 학습을 위해 제안되었던 기술들을 모아서 제작한 방식이다. nerfacto는 충분히 작은 신경망(MLP)을 사용하기 때문에 신경망의 크기는 용량에 큰 영향을 주지 않았고, 학습된 모델의 전체 용량은 해시 맵 크기에 영향을 받는 것을 실험적으로 확인하였다. 따라서 NeRF 모델의 크기를 조절하기 위해 해시맵 크기를 결정하는 파라미터 (log2-hashmap-size)를 조정하였다. 해당 파라미터는 기본 모델에서 19로 설정되어 있었고, 일정 값 이하에서는 용량의 변화가 크게 없어 9부터 19까지 단계마다 2 만큼 늘려가며 총 6개의 값을 사용하였다.

3DGS 모델을 만들기 위해서는 Nerfstudio에서 제공하는 splatfacto를 사용하였다. splatfacto는 nerfacto와 비슷하게 여러 연구들에서 제시된 기술들을 모아서 제작한 방식이다. splatfacto 역시 용량에 영향을 주는 파라미터를 관련 문서와 함께 실험적으로 결정했으며, 3DGS에서 가우시안의 색을 표현하기 위한 구면 조화 함수의 차수를 결정하는 파라미터 (sh-degree)와 훈련 과정에서 가우시안의 투명도 임계값 파라미터 (cull-alpha-thresh)를 조정하였다. 이 중 sh-degree 파라미터는 용량에 가장 큰 영향을 주었고, cull-alpha-thresh 파라미터는 일부 영향을 주었다. 두 파라미터를 조정하여 총 6개의 모델을 제작하였다.

표 3, 표 4는 각각 제작된 NeRF와 3DGS 모델의 파라미터 정보와 데이터 크기를 보여준다. NeRF 모델은 신경망 기반 모델로 표준으로 정해진 파일 형식이 없기 때문에, 데이터 크기로 학습된 모델 파

일의 크기를 사용하였다. 3DGS는 NeRF와 다르게 명시적인 표현 방식이기 때문에 포인트 클라우드, 메쉬와 비슷한 형태의 PLY 파일로 저장할 수 있어 해당 파일의 크기를 사용하였다. 해당 표의 데이터 크기는 압축이 적용되지 않은 미가공(Raw) 상태의 모델 혹은 PLY 파일 데이터 크기이다.

표 3. 여러 품질로 제작된 NeRF 모델 정보
Table 3. NeRF information with different quality

Level	log2-hash map-size	Size (MB)	Gbps @30Hz
0	9	28.383	6.652
1	11	28.945	6.784
2	13	31.101	7.289
3	15	38.999	9.140
4	17	67.069	15.719
5	19	167.770	39.321

표 4. 여러 품질로 제작된 3DGS 정보
Table 4. 3DGS information with different quality

Level	sh- degree	cull-alpha -thresh	Size (MB)	Gbps @30Hz
0	1	0.4	4.441	1.041
1	1	0.1	5.265	1.234
2	2	0.4	7.106	1.665
3	2	0.1	8.071	1.892
4	3	0.4	10.683	2.504
5	3	0.1	12.333	2.891

2. 시각 품질 분석

제작된 모델의 시각 품질을 평가하기 위해 PSNR, SSIM [17], LPIPS [18]의 총 세 가지 측정 지표를 계산하였다. 세 지표 모두 이미지의 유사도를 평가하기 위한 지표들로, PSNR, SSIM은 화질의 손실량을 평가하기 위해 주로 사용하며, LPIPS는 앞의 두 지표보다 인간 지각적인 면을 더욱 고려한 지표이다. PSNR, SSIM은 높을수록 좋은 수치이며, LPIPS는 낮을수록 좋은 수치이다. 포인트 클라우드와 메쉬는 테스트 이미지와 같은 시점에서 렌더링을 하기 위하여 데이터셋에서 제공하는 정보 및 변환 행렬을 이용해 카메라의 내부 변수(Intrinsic)와 위치를 계산하였으며, Open3D를 이용해 3D 모델을 불러오고 렌더링하여 이미지로 저장하였다. 저장된 이미지들은 테스트 이미지들과의 정해진 품질 지표를 계산하였고, 각 표현 방식 및 품질마다 평균 및 표준 편차를 계산하였다. NeRF와

3DGS 모델은 Nerfstudio에서 제공하는 평가 툴을 활용하였다. 네 가지 표현 방식 품질 지표 계산을 위해 같은 라이브러리 [19] 및 함수가 사용되었다.

3. 패킷 손실 분석

제작된 모델이 패킷 손실이 발생하는 경우 품질에 주는 영향을 평가하기 위해 두 서버로 이루어진 간단한 콘텐츠 전송 시스템을 구현하였다. 전송 서버와 수신 서버 모두 Python 소켓을 이용해 구현하였으며, 패킷 손실의 영향을 그대로 두기 위해 재전송을 보장하지 않는 UDP 프로토콜을 이용하였다. 3D 모델 파일은 정해진 패킷 크기 (2000 byte)로 분리하여 전송하였다. 패킷 손실은 전송 서버에서 리눅스 tc 명령어를 이용하여 0.1%, 1%, 10%의 환경을 만들고, 수신한 모델은 III-2장 시각 품질 분석에서와 같은 방식으로 시각 품질을 평가하였다.

IV. 분석 결과 및 고려 요소

이 장에서는 본 논문에서 제안하는 분석의 결과와 3D 표현 방식 사용을 위해 고려해야 할 요소들을 서술한다.

1. 데이터 크기(대역폭)-품질 분석

먼저 포인트 클라우드는, 렌더링 환경에서 포인트의 크기 설정에 따라 품질이 다르게 나타났다. 그림 2는 포인트 크기에 따른 용량-품질 (PSNR) 그래프이다. 포인트 클라우드의 용량은 일반적으로 총 포인트의 수 혹은 포인트의 밀도와 비례하는데, 포인트의 크기가 작을 때는 포인트의 밀도가 낮기 때문에 빈 공간이 많이 생겨 품질이 낮게 측정되었다.

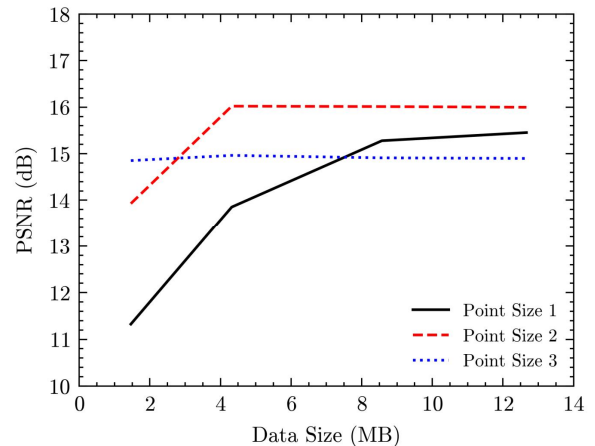


그림 2. 포인트 클라우드에서 포인트 크기에 따른 PSNR
Fig. 2. PSNR according to point size in point cloud

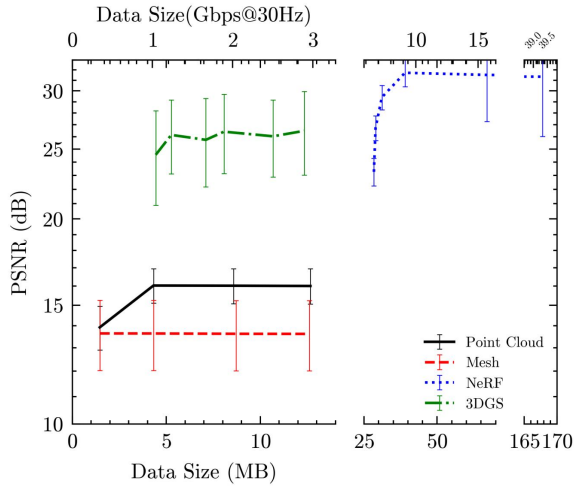


그림 3. 데이터 크기에 따른 시각 품질 (PSNR)
Fig. 3. PSNR according to the data size

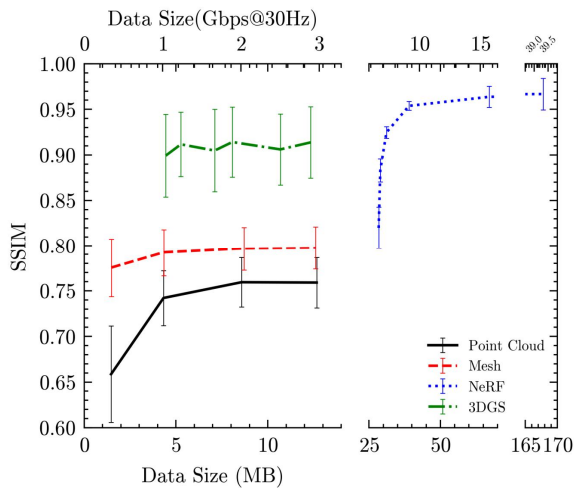


그림 4. 데이터 크기에 따른 시각 품질 (SSIM)
Fig. 4. SSIM according to the data size

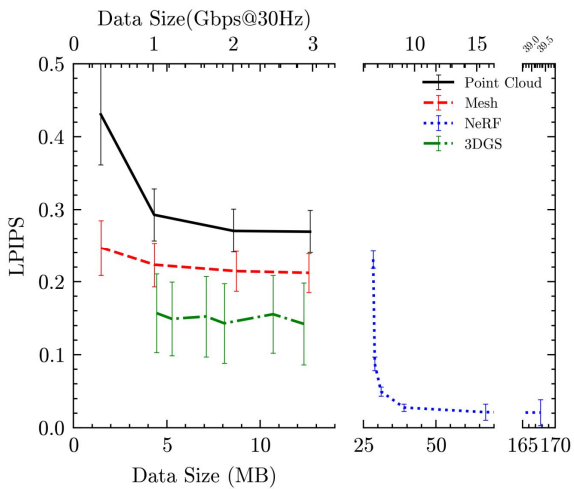


그림 5. 데이터 크기에 따른 시각 품질 (LPIPS)
Fig. 5. LPIPS according to the data size

반면에 포인트의 크기가 크면 이러한 빈 공간을 채워줄 수 있기 때문에 용량(포인트 수/밀도)이 늘어나더라도 품질 차이가 적었다. 이는 포인트 클라우드의 렌더링 방법 또한 품질에 영향을 미칠 수 있다는 것을 보여준다. 다른 3D 표현 방식과의 비교를 위해서는 평가한 세 개의 포인트 크기 중 가장 높은 PSNR을 보인 포인트 크기 2 설정을 사용하였다.

그림 3은 포인트 클라우드, 메쉬, NeRF, 3DGS, 총 네 가지 표현 방식에 대한 데이터 크기-품질 (PSNR) 그래프이다. 그림 4, 그림 5는 각각 데이터 크기-품질(SSIM), 데이터 크기-품질(LPIPS) 그래프이다. 포인트 클라우드의 경우 데이터 크기가 늘어남에 따라 품질 또한 상승했지만, 일정 수준 이상에서는 거의 변화가 없는 모습을 보였다. 이는 콘텐츠에 따라, 혹은 사용자 위치와 콘텐츠 사이의 거리에 따라 적절한 포인트 밀도나 개수를 설정하면 유사한 시각적 품질을 얻을 수 있다 것을 의미한다. 메쉬의 경우 포인트 및 면의 수가 늘어날 때 SSIM과 LPIPS가 소폭 개선되었지만, PSNR는 거의 변화를 보이지 않았다. 이는 메쉬를 이용하는 경우 낮은 용량의 모델을 이용하더라도 품질에서 큰 손해를 보지 않을 수 있다는 것을 의미한다. NeRF는 가장 높은 품질을 보여주었지만, 전체적으로 데이터 크기가 매우 크게 나타났다. 또한 일정 용량 이상에서는 PSNR은 거의 변화가 없었고, SSIM과 LPIPS만 소폭 개선되었다. 3DGS는 모든 크기에 대해 거의 비슷한 품질 수치를 보여주었다. 포인트 클라우드와 메쉬보다는 많이 증가하는 수치를 보여주었고, NeRF보다는 낮은 수치들을 보여주었다. 용량은 NeRF보다는 훨씬 적은 크기로, 용량 대비 품질에서는 가장 좋은 모습을 보여주었다.

2. 패킷 손실의 영향

표 5는 각 표현 방식 별로 0.1%, 1%, 10% 패킷 손실이 일어나는 환경에서 시각 품질 (PSNR) 변화를 보여준다. 포인트 클라우드의 경우에는 포인트 간 연결이 없고, 각 포인트의 위치 및 색 정보만 불러올 수 있다면 렌더링을 할 수 있기 때문에, 손실률이 증가함에 따라 품질이 감소하지만, 모델을 로드할 수 있었다. 손실률에 따라서 일부 포인트들이 잘못된 위치에 렌더링이 되거나 잘못된 색을 가지는 것을 확인할 수 있다. 다만 저장 형식에 대한 정보를 담고 있는 헤더가 손실될 경우에는 로드가 불가능할 것으로 보인다. 메쉬의 경우에는 포인트

정보와 함께 면을 이루기 위한 포인트 간 연결 정보를 함께 담고 있어 해당 부분이 손상되는 경우 모델을 로드할 수 없었다. 포인트 클라우드와 마찬가지로 헤더가 손상될 경우에는 로드가 불가능한 것으로 보인다. NeRF 모델은 로드가 가능한 경우와 불가능한 경우가 모두 있었는데, 역시 헤더 등 체크포인트 내 중요 정보가 손상되지 않는다면 대부분 신경망의 가중치 혹은 해시맵과 관련된 자료일 것으로 예상되어 로드가 가능할 것으로 보인다. 본 실험에서는 손실률 0.1%에서 NeRF 모델 로드에 성공했지만, 항상 로드에 성공할 것이라는 보장을 할 수 없다는 것을 의미한다. 3DGS의 경우에는 포인트 클라우드와 유사하게 각 가우시안에 대한 정보를 담고 있고, 메쉬처럼 면을 구성하기 위한 연결 정보를 담고 있지는 않기 때문에 헤더를 제외한 부분이 손상되더라도 렌더링이 가능했다. 하지만 일부 가우시안이 잘못된 위치, 크기, 색으로 렌더링이 되어 전체 콘텐츠를 방해하며 크게 시각 품질을 떨어뜨리는 모습을 확인할 수 있다. SSIM 및 LPIPS 지표에서도 손실이 늘어남에 따라 품질이 떨어지는 유사한 양상을 보였다. 표 6은 손실률에 따른 SSIM의 변화를 보여주며, PSNR과 마찬가지로 손실이 발생할 경우 SSIM 값이 감소하여 시각 품질이 떨어졌음을 확인할 수 있다. 표 7은 손실률에 따른 LPIPS의 변화를 보여주며, 손실이 발생할 경우 값이 증가해 시각 품질이 떨어졌음을 확인할 수 있다. 손실이 일어날 수 있는 환경에서 안정적으로 3D 콘텐츠를 로드하기 위해서는 적절한 우선순위 설정, 패킷화, 재전송, 대응 방식을 고려할 필요가 있다.

표 5. 손실률에 따른 PSNR 변화

Table 5. PSNR difference according to the loss rate

Loss Rate	Point Cloud	Mesh	NeRF	3DGS
0.1%	-0.000	Load Failed	-0.075	-0.08
1%	-0.035	Load Failed	Load Failed	-17.041
10%	-0.228	Load Failed	Load Failed	-19.132

표 6. 손실률에 따른 SSIM 변화

Table 6. SSIM difference according to the loss rate

Loss Rate	Point Cloud	Mesh	NeRF	3DGS
0.1%	-0.000	Load Failed	-0.001	-0.000
1%	-0.005	Load Failed	Load Failed	-0.151
10%	-0.031	Load Failed	Load Failed	-0.223

표 7. 손실률에 따른 LPIPS 변화

Table 7. LPIPS difference according to the loss rate

Loss Rate	Point Cloud	Mesh	NeRF	3DGS
0.1%	0.000	Load Failed	0.000	0.000
1%	0.013	Load Failed	Load Failed	0.256
10%	0.031	Load Failed	Load Failed	0.335

3. 접근성

본 항목에서는 각 표현 방식을 사용하는데 있어 접근 용이성에 대해 서술한다. 기존 표현 방식인 포인트 클라우드와 메쉬의 경우, PC, 모바일, 웹 등 대부분의 플랫폼에서 지원할 정도로 접근성이 높다. Blender [20], Maya [21] 등 여러 3D 소프트웨어와 Unity [22], Unreal Engine [23]과 같은 게임 엔진에서 대부분 로드가 가능하며, 웹 환경에서도 Three.js [24] 등의 라이브러리를 이용해 WebGL 기반 렌더링이 가능하다. 반면, NeRF의 경우 신경망 기반 표현 방식으로 기존 표현 방식들과 렌더링 방식이 크게 다르고, 렌더링을 위해 신경망 구조와 가중치를 모두 알아야 하기 때문에 구현 방식이나 사용 기술에 따라라도 다르다는 점이 접근성을 낮게 만든다. NeRF 기술과 관련된 연구는 활발하게 이루어지고 있어 관련하여 공개된 코드들과 툴들이 많이 존재하지만, 설치 및 사용이 복잡하여 일반 사용자에게 서비스로 제공하기 위해서는 이런 문제들이 해결되어야 한다. 또한 여러 형태의 신경망, 기술을 읽을 수 있는 표준 표현 방식 혹은 API가 정의될 필요가 있다. 일부 툴 [15]에서는 NeRF로 훈련된 모델을 기반으로 하여 고품질의 포인트 클라우드, 메쉬의 형태로 내보낼 수 있도록 지원을 하고 있으며, NeRF 모델의 대체제로 이용이 가능하다.

표 8. 각 표현 방식 별 장단점 비교

Table 8. Advantages and Disadvantages of each 3D representations

Type	Advantages and Disadvantages
Point Cloud	+ Simple, easy to use. Supported by many existing libraries and frameworks. Available even if loss occurs. Easy to edit (Beneficial for adaptive streaming). - Low visual quality when there are few points or low density. High visual quality requires high bandwidth.
Mesh	+ Simple, easy to use. Supported by many existing libraries and frameworks. Low bandwidth. Relatively easy to edit. - Low visual quality. Unavailable if loss occurs.
NeRF	+ High visual quality (photorealistic). A lot of research is currently ongoing (Disadvantages are likely to be resolved in the future). - Requires training. High bandwidth. Unavailable if loss occurs. Low FPS. Limited or no support by existing libraries and frameworks. Hard to edit.
3DGS	+ High visual quality (photorealistic) with lower bandwidth. High FPS. A lot of research is currently ongoing. Editable. - Requires training. Available even if loss occurs, but quality can be greatly reduced. Limited or no support by existing libraries and frameworks.

3DGS의 경우에는 명시적인 표현 방식으로 NeRF보다는 접근성이 높으며, Blender, Unity와 같은 3D 프로그램을 위한 플러그인 [25], 웹에서 로드하기 위한 뷰어 [26, 27] 등이 존재하지만, 기존 방식과 다른 형태의 렌더링 방식을 요구하기 때문에 기존 표현 방식들에 비해서는 접근성이 떨어진다.

4. 압축 알고리즘

본 항목에서는 높은 대역폭 및 저장 공간을 요구하는 3D 콘텐츠의 미가공 (Raw) 데이터를 더욱 효율적으로 전송하기 위한 압축 알고리즘의 측면에서 각 표현 방식을 비교한다. 압축 과정은 3D 콘텐츠의 용량을 크게 줄이는 과정 중 하나로, 많게는 대역폭 사용량을 기존 대비 1% 이하로 줄일 수 있다. 포인트 클라우드와 메쉬는 오랜 기간 연구되어 압축 알고리즘이 다양하게 존재한다. 포인트 클라우드의 경우에는 Draco [28], PCL [29]에서 지원하는 옥트리 (Octree) 기반 압축 방식, GROOT [30]와 같이 개선된 트리 기반 방식, MPEG의 PCC 표준이 존재한다 (V-PCC [31], G-PCC [32]). 메쉬 역시 Draco와 같은 라이브러리를 사용하여 압축이 가능하며, 특히 glTF [33] 포맷은 3D 모델 및 장면을 효율적으로 전송하기 위한 포맷으로, 기하 정보 뿐만 아니라 변형이 쉽다는 메쉬의 장점을 살려 애니메이션 등 움직임 정보까지 담을 수 있도록 설계되었다. NeRF의 경우 NeRF 기술에 대한 연구와 함께 압축을 위한 연구 [34, 35], 혹은 작은 모델로 효율적으로 학습하기 위한 연구 [36]도 다양하게 이

루어지고 있다. 3DGS의 경우 2023년에 제안된 기술로, 압축 방식에 관한 연구는 많지 않지만, 기존 포인트 클라우드, 메쉬를 저장하기 위해 사용한 파일과 구조 및 형식 면에서 유사하기 때문에 이를 응용할 수 있을 것으로 예상된다.

5. 기타 고려 요소

본 항목에서는 각 표현 방식을 사용하기 위해 추가로 고려해야 하는 요소들에 대한 설명을 작성한다.

먼저 렌더링 속도이다. 기본적으로 렌더링 속도는 GPU 성능에 의존하는데, 기존 표현 방식인 포인트 클라우드와 메쉬는 데이터 크기에 따라 다르지만, 일반적으로 빠른 렌더링 속도를 보여주며, 3DGS 역시 빠른 렌더링 속도를 특징으로 하고 있다. 그에 반해 대부분의 NeRF 방식들은 상대적으로 느린 렌더링 속도를 가지고 있다. 다만 이러한 한계를 극복하기 위해 FastNeRF [37] 등 빠른 렌더링 속도를 위한 연구도 활발히 이루어지고 있다.

다음은 사전 처리 시간이다. NeRF와 3DGS 기반 방식들은 3D 콘텐츠를 만들기 위해 모델 학습을 요구하며, 이에 따라 실시간 스트리밍, 화상회의 등 일부 어플리케이션에서는 적합하지 않을 수 있다. 다만 이 항목 역시 Instant-NGP [16] 등 적은 학습으로도 3D 복원을 가능하게 하는 연구가 많이 제안되고 있다.

다양한 네트워크 상황에 적응하기 위하여 2D 영상에서는 MDC [38], SVC [39] 등과 같은 방식으

로 다중 품질로의 변환을 통해 적응형 스트리밍을 구현할 수 있었다. 포인트 클라우드, 메쉬, 3DGS와 같은 명시적인 표현 방식은 공간상에서 빠르게 샘플링을 통해 품질 조정이 가능한데 반해, NeRF와 같은 암시적인 표현 방식은 이러한 샘플링이 어렵다. NeRF를 스트리밍에 이용하기 위해서 2D 영상에 적용한 방법론을 적용하는 연구들이 진행되고 있다 [40, 41].

마지막으로는 사용자 시선에 따른 처리이다. ViVo [42], GROOT [30] 등 여러 기존 분류메트릭 영상 스트리밍 관련 연구들이 사용자 시선 내의 콘텐츠만 전송하여 대역폭을 절약한다. 3D 공간에서 유저의 위치와 3D 데이터의 위치를 알기 때문에 가능한 것으로, 포인트 클라우드, 메쉬, 3DGS와 같이 명시적인 표현 방식에서 빠르게 가능하지만, NeRF와 같은 암시적인 표현 방식에서는 모호한 부분이 존재한다. 한 콘텐츠에 대해 하나의 신경망이 아닌 위치에 따라 여러 개의 신경망을 활용하는 연구 [43]를 활용할 수 있다.

표 8은 본 장에서 설명한 네트워크 측면에서의 여러 고려 요소들에 대해 각 방식 별로 장단점인 부분들을 정리한 표이다.

V. 한계점

본 연구에서는 네 가지 3D 콘텐츠 표현 방식에 대한 분석 및 관련 연구, 기술들을 제시하였다. 하지만 몇 가지 한계점이 존재하는데, 분석을 위해 같은 이미지 셋에 대한 3D 모델을 구축하였지만, NeRF와 3DGS의 경우 품질과 용량에 영향을 줄 수 있는 요소가 너무 많다. NeRF 혹은 3DGS를 기반으로 하는 각 방식마다 사용하는 신경망의 크기, 데이터 처리 및 저장 방식 등이 모두 다르며, 학습을 위한 파라미터 세팅에 따라서도 용량이 크게 달라질 수 있다. 본 연구에서는 NeRF와 3DGS 각각 한 가지 방식만을 사용하여 평가하였고, 파라미터 결정을 위해서도 다른 요소들보다 용량에 주는 파라미터들을 주로 고려하였다. 이에 따라 다른 세팅으로 3D 모델을 제작하는 경우에는 용량-품질 간 관계가 본 연구에서의 결과와 크게 달라질 수 있다. 또한 본 연구에서는 간단한 콘텐츠 중 하나인 lego 씬을 대상으로 연구를 수행하였지만, 더욱 크고 복잡하거나, 실제 환경을 보여주는 콘텐츠의 경우에는 다른 양상을 보일 수 있다. 이러한 다양한 장면들과 방식을 고려하기 위해서는

Nerfbaselines[10]처럼 다양한 모델과 데이터에 대한 단일화된 API를 만들어 제공하는 방식이 가능하다.

VI. 결 론

본 연구에서는 네트워크 측면에서 3D 콘텐츠 표현 방식에 대한 이해를 돕기 위한 분석 및 관련 연구를 제시하였다. 같은 3D 콘텐츠에 대해 포인트 클라우드, 메쉬, NeRF, 3DGS, 총 네 가지 표현 방식으로 제작한 후, 서비스 품질에 영향을 주는 대역폭, 패킷 손실 등의 측면에서 분석을 수행하고, 압축 알고리즘, 적응형 스트리밍 등 서비스 및 네트워크 품질 관리 측면의 몇 가지 고려 요소들에 대한 관련 연구와 이해를 제시하였다. 향후 연구로는 여러 3D 콘텐츠에 대해 전송과 더불어 실제로 렌더링을 수행하는 시스템을 구축하여 더욱 종합적인 비교 및 사용자 평가를 수행할 것이다.

References

- [1] Prescient & Strategic Intelligence, Extended Reality Market Research Report (2022), Retrieved Nov., 14, 2024, from <https://www.psmarketresearch.com/market-analysis/extended-reality-xr-market-insights>.
- [2] International Data Corporation, Slower Growth for AR/VR Headset Shipments in 2023 but Strong Growth Forecast Through 2027, According to IDC (2023), Retrieved Nov., 14, 2024, from <https://www.idc.com/getdoc.jsp?containerId=prUS50511523>
- [3] Mildenhall, Ben, et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis." in European Conference on Computer Vision (pp. 405 - 421). 2020.
- [4] Kerbl, Bernhard, et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". ACM Transactions on Graphics 42, no.4 (2023).
- [5] Yili Jin, et al. "From Capture to Display: A Survey on Volumetric Video." arXiv preprint arXiv:2309.05658 (2023).
- [6] Yili Jin, et al. "A Networking Perspective of Volumetric Video Service: Architecture, Opportunities and Case Study". IEEE Network (2024): 1-1.
- [7] Bangya Liu, and Suman Banerjee. "SwinGS: Sliding

- Window Gaussian Splatting for Volumetric Video Streaming with Arbitrary Length." arXiv preprint arXiv:2409.07759 (2024).
- [8] Sambugaro, Zeno, et al. "3D reconstruction methods in industrial settings: a comparative study for COLMAP, NeRF and 3D Gaussian Splatting." in *ITALIA 2024* (2024).
- [9] Liu, Kaiyan, et al. "Toward next-generation volumetric video streaming with neural-based content representations." In *Proceedings of the 1st ACM Workshop on Mobile Immersive Computing, Networking, and Systems*, pp. 199-207. 2023.
- [10] Kulhanek, Jonas, and Torsten Sattler. "NerfBaseline: Consistent and Reproducible Evaluation of Novel View Synthesis Methods." arXiv preprint arXiv:2406.17345 (2024).
- [11] COLMAP, Retrieved Nov., 14, 2024, <https://github.com/colmap/colmap>.
- [12] Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4104-4113. 2016.
- [13] Schönberger, Johannes L., et al. "Pixelwise view selection for unstructured multi-view stereo." In *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III* 14, pp. 501-518. Springer International Publishing, 2016.
- [14] Zhou, Qian-Yi, et al. "Open3D: A modern library for 3D data processing." arXiv preprint arXiv:1801.09847 (2018).
- [15] Tancik, Matthew, et al. "Nerfstudio: A modular framework for neural radiance field development." . In *ACM SIGGRAPH 2023 Conference Proceedings* (pp. 1 – 12).2023.
- [16] Muller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding". *ACM transactions on graphics (TOG)* 41, no.4 (2022): 1 – 15.
- [17] Brunet, Dominique, et al. "On the mathematical properties of the structural similarity index". *IEEE Transactions on Image Processing* 21, no.4 (2011): 1488 – 1499.
- [18] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." . In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 586 – 595).2018.
- [19] torchmetrics, Retrieved Nov., 14, 2024, <https://github.com/Lightning-AI/torchmetrics>.
- [20] Blender, Retrieved Nov., 14, 2024, <https://www.blender.org>.
- [21] Maya, Retrieved Nov., 14, 2024, <http://www.autodesk.co.kr/products/maya/overview>.
- [22] Unity, Retrieved Nov., 14, 2024, <https://unity.com>.
- [23] Unreal Engine, Retrieved Nov., 14, 2024, <https://www.unrealengine.com>.
- [24] Three.js, Retrieved Nov., 14, 2024, <https://threejs.org>.
- [25] UnityGaussianSplatting, Retrieved Nov., 14, 2024, <https://github.com/aras-p/UnityGaussianSplatting>.
- [26] Polycam Gaussian Splatting Viewer, Retrieved Nov., 14, 2024, <https://poly.cam/tools/gaussian-splatting>.
- [27] 3D Gaussian Splatting for Three.js, Retrieved Nov., 14, 2024, <https://github.com/mkkellogg/GaussianSplats3D>.
- [28] Draco 3D Data Compression, Retrieved Nov., 14, 2024, <https://google.github.io/draco>.
- [29] Radu Bogdan Rusu, and Steve Cousins. "3D is here: Point Cloud Library (PCL)." . In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
- [30] Lee, Kyungjin, et al. "GROOT: a real-time streaming system of high-fidelity volumetric videos." . In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking* (pp. 1 – 14).2020.
- [31] MPEG V-PCC, Retrieved Nov., 14, 2024, <https://www.iso.org/standard/83535.html>.
- [32] MPEG G-PCC, Retrieved Nov., 14, 2024, <https://www.iso.org/standard/78990.html>.
- [33] glTF 2.0 Specification, Retrieved Nov., 14, 2024, <https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html>.
- [34] Girish, Sharath, et al. "Shacira: Scalable hash-grid compression for implicit neural representations." . In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 17513 – 17524).2023.

- [35] Li, Lingzhi, et al. "Compressing volumetric radiance fields to 1 mb." . In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4222 - 4231).2023.
- [36] Barron, Jonathan T, et al. "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields." . In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5855 - 5864).2021.
- [37] Garbin, Stephan J, et al. "Fastnerf: High-fidelity neural rendering at 200fps." . In Proceedings of the IEEE/CVF international conference on computer vision (pp. 14346 - 14355).2021.
- [38] Goyal, Vivek K. "Multiple description coding: Compression meets the network". IEEE Signal processing magazine 18, no.5 (2001): 74 - 93.
- [39] Schwarz, Heiko, et al. "Overview of the scalable video coding extension of the H. 264/AVC standard," in IEEE Transactions on circuits and systems for video technology, vol. 17, no. 9, pp. 1103-1120, 2007.
- [40] Cho, Junwoo, et al. "Streamable neural fields." . In European Conference on Computer Vision (pp. 595 - 612).2022.
- [41] Yang, Guo-Wei, et al. "Recursive-nerf: An efficient and dynamically growing nerf". IEEE Transactions on Visualization and Computer Graphics 29, no.12 (2022): 5124 - 5136.
- [42] Han, Bo, et al. "ViVo: Visibility-aware mobile volumetric video streaming." . In Proceedings of the 26th annual international conference on mobile computing and networking (pp. 1 - 13).2020.
- [43] Reiser, Christian, et al. "Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps." . In Proceedings of the IEEE/CVF international conference on computer vision (pp. 14335 - 14345).2021.

유 상 우 (Sangwoo Ryu)



2020 포항공과대학교, 컴퓨터공학과 학사
 2020 ~ 포항공과대학교, 인공지능대학원 석/박사 통합과정
 <관심분야> 2D/3D 비디오 스트리밍, 영상회의, 인공지능

홍 원 기 (James Won-Ki Hong)



2020~현재 포스텍 크립토티블록체인연구센터 공동센터장
 2019-2023 포스텍 교육혁신센터장
 2012~2014 KT 종합기술원 원장
 2008~2012 포스텍 정보전자융

합공학부장

2008~2010 포스텍 컴퓨터공학과 주임교수
 2007~2010 포스텍 정보통신연구소 연구소장
 2007~2011 포스텍 정보통신대학원장
 1995~현재 포스텍 컴퓨터공학과 교수
 1992~1995 Univ. of Western Ontario, 연구교수
 <관심분야> AI 기반 네트워크관리, SDN/NFV, 무크기반 교육, 블록체인 및 암호화폐, Vmeeting 화상회의시스템