

다차원영상기술표준화포럼표준(국문표준)

XDFK\_01.0056/R0

제정일 : 2024년 12월 03일

국문 표준명	동적 3D 가우시안 스플래팅 모델 구성 요소의 압축 기법
영문 표준명	Compression Methods for Components of Dynamic 3D Gaussian Splatting Model
표준초안 작성자	최재열, 박준형, 정종범, 김영규, 류은석(성균관대학교)



본 문서에 대한 저작권은 다차원영상기술표준화포럼에 있으며, 이 문서의 전 체 또는 일부에 대하여 상업적 이익을 목적으로 하는 무단 복제 및 배포를 금합니다.

Copyright© XDF(2019). All Rights Reserved.

# 서 문

## 1 표준의 목적

이 표준의 목적은 고품질 중간 시점 합성과 고속 렌더링을 지원하는 3차원 재구성 모델인 4D 가우시안 스플래팅 (4D Gaussian splatting, 4D-GS)을 압축하는 기술을 제안함으로써 저장/전송의 효율화를 달성하는 것이다. 4D-GS를 구성하는 개별 요소들에 대해 맞춤형으로 압축 모듈을 제공하고 응용 소프트웨어 개발에 필요한 프레임워크를 제공하는 것 또한 이 표준의 목적이다.

## 2 주요 내용 요약

다수의 시점으로 구성된 몰입형 영상으로부터 기계학습을 통해 학습된 3D 가우시안 스플래팅 (3D Gaussian splatting, 3DGS) 기반 동적 3차원 재구성 모델을 효율적으로 압축하기 위한 압축 모듈을 제공한다. 첫 번째로 피처 임베딩 (embedding)이 포함된 4차원 신경 복셀 (voxel)을 2차원으로 분할하여 양자화 한 후 비디오 코덱으로 부호화하는 모듈이며, 두 번째는 렌더링 시 영향력을 적게 가지는 3D 가우시안을 프루닝 (pruning)하여 불필요한 가우시안을 제외하는 모듈이다. 두 가지 방법을 이용해 동적 3DGS 모델에 대한 파일 크기를 감소시켜 효율적 저장 및 전송이 가능하다.

## 3 인용 표준과의 비교

이 표준은 MPEG 비디오 시스템 표준 및 특정 신경 복사 필드 모델 또는 특정 비디오 코덱 표준과 관련성이 없는 범용의 표준임.

## 목 차

1	적용 범위	1
2	인용 표준	1
3	용어 정의	1
4	약어	1
5	동적 3D 가우시안 스플래팅 모델 구성 요소의 압축 기법	2
5.1	4D 가우시안 스플래팅을 통한 동적 장면 부호화	2
5.2	4차원 신경 복셀을 활용한 위치적 특성 임베딩의 부호화 및 복호화	3
5.3	가우시안 프루닝 기술을 사용한 표준 가우시안 네트워크 압축	4
부록 I	필요성	5
부록 II-1	지식재산권 요약서 정보	6
II-2	시험인증 관련 사항	7
II-3	본 표준의 연계(family) 표준	8
II-4	참고 문헌	9
II-5	영문표준 해설서	10
II-6	표준의 이력	11

# 동적 3D 가우시안 스플래팅 모델 구성 요소의 압축 기법 (Compression Methods for Components of Dynamic 3D Gaussian Splatting Model)

## 1 적용 범위

본 표준의 적용 범위는 3DGS[1] 기반의 동적 장면 표현시 압축에서의 입출력이며, 이는 구체적으로 동적 3DGS 모델에서의 4차원 신경망의 부호화와 복호화 과정 및 가우시안 프루닝 과정을 포함한다.

## 2 인용 표준

해당 사항 없음.

## 3 용어 정의

### 3.1 방사형 필드 (Radiance Fields)

공간상에서 빛의 색상과 밝기를 출력하기 위해 공간상에서 입자의 위치와 방향을 부호화한 함수.

### 3.2 임의 시점 합성 (Novel View Synthesis)

존재하지 않는 시점에서의 영상을 합성하는 기술. 다중 시점에서 이산적으로 촬영된 영상이 존재할 때 중간 시점의 영상을 취득하기 위해 사용됨.

### 3.3 3차원 재구성 (3D Reconstruction)

특정 물체나 공간에 대한 입체적 정보를 디지털로 모델링한 형태. 해당 3차원 범위 내에서 2차원 임의 시점 합성이 가능하다는 점에서 가상 현실 제작 시 사용됨.

### 3.4 움직임 기반 구조 (Structure from Motion)

동일한 객체를 중첩되도록 촬영한 여러 시점의 이미지로부터 3차원 구조를 복원하는 기술. 특성 추출과 함께 사용하여 촬영된 이미지들의 카메라 매개변수를 추정 시 사용됨.

### 3.5 3차원 가우시안 스플래팅 (3D Gaussian Splatting)

다수의 독립적인 3차원 가우시안 분포를 활용하여 3차원 재구성을 수행하는 기계 학습 기반 방법론. 초당 30 프레임 이상의 속도의 임의 시점 합성을 지원함.

### 3.6 초매개변수 (Hyperparameter)

기계학습시 학습 과정을 설정하기 위해 학습 전 지정되는 매개변수. 모델의 학습 및 일반화 성능에 영향을 미치며 학습률, 에포크 수, 활성화 함수 등이 해당함.

### 3.7 자유 시점 비디오 (Free-Viewpoint Video)

시청자가 능동적으로 시청하고자 하는 시점을 선택할 수 있는 몰입형 비디오 시스템. 2D 비디오와 비교하여 가상 시점 합성 및 3차원 재구성 등 복잡한 연산이 요구됨.

### 3.8 구면 조화 함수 (Spherical Harmonics)

구의 표면에서 구면 좌표계에 대해 정의되는 함수의 집합. 컴퓨터 그래픽스 분야에서 시야각에 따라 달라지는 색상 변화를 표현하기 위해 사용됨.

## 4 약어

3DGS	3D Gaussian Splatting
4DGS	4D Gaussian Splatting
HMD	Head-Mounted Display
NeRF	Neural Radiance Fields

## 5 동적 3D 가우시안 스플래팅 모델 구성요소의 압축 기법

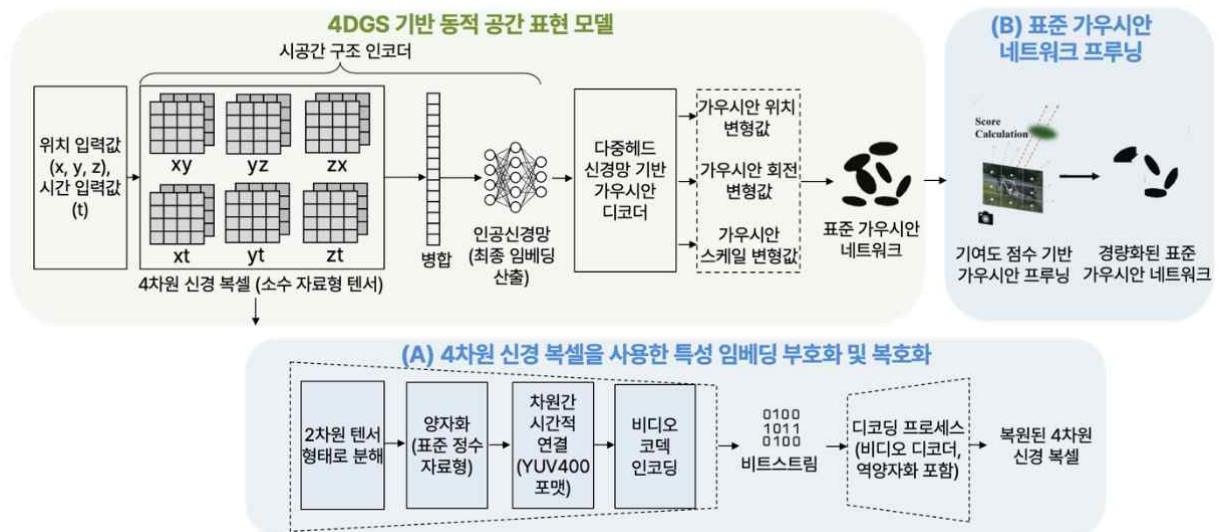
### 5.1 4D 가우시안 스플래팅을 통한 동적 장면 부호화

3DGS 기술은 기계학습 기반으로 여러 시점에서 취득된 이미지 집합을 통해 3차원 구조를 복원할 수 있다. 움직이는 물체를 포함한 동적 장면을 복원하기 위해 3DGS의 여러 응용 방법론이 존재하며, 그 중 4DGS[2] 기반 모델은 모든 시간대의 좌표를 표준 (canonical) 시간대의 좌표로 대응시키는 것으로 이를 해결한다. 표준 시간대의 좌표로 대응시키기 위해 임의의 시간  $t$ 에 해당하는 좌표를 가장 일치하게 표현할 수 있는 표준 시간  $t'$ 의 좌표로 변환하는 네트워크가 주로 사용된다. 이 때 정밀한 표현을 위해 4차원 좌표  $(x, y, z, t)$ 를 고차원의 학습 가능한 피쳐 임베딩 평면으로 매핑한다. 이러한 고차원 피쳐 평면으로의 대응은 3DGS 기반 3차원 재구성 모델 뿐 아니라 K-Planes, HexPlane 등의 NeRF 기반 3차원 재구성 모델에도 사용되는 기법에 해당한다. 매핑된 특성 임베딩은 아다마르 곱 (hadamard product) 연산 및 연결 (concatenate) 연산이 수행되어 최종 피쳐 임베딩으로 부호화할 수 있다. 즉, 입력 좌표로부터 최종 임베딩을 얻기 위해서 고차원 피쳐 평면을 개별적으로 저장해야 한다.

이후 디코더 단계에서는 임베딩으로부터 표준 (canonical) 좌표계의 3D 가우시안을 추론

하는 과정이 시행된다. 디코더 과정 또한 경사하강법 기반의 기계학습 방식으로 학습되며, 4DGS 기반 모델의 부호기와 복호기는 최종적인 렌더링 결과물과 비교하여 손실 함수를 계산하는 식으로 동시에 학습된다. 이는 종단 간 (end-to-end) 학습과정이라고 볼 수 있으며, 학습이 끝난 후 HMD 등에 응용하여 임의 시점 합성을 수행할 때에도 전체 모델이 필요로 함을 의미한다.

(그림 5-1)은 4DGS 기반 3차원 재구성 모델의 구조도와 그 중 본 표준이 다루는 압축 구조도를 제시한다. 본 표준은 4차원 피쳐 임베딩 평면에 대해 기존 동영상 코덱 기반으로 압축을 수행하고 이를 복호화하는 과정까지 포함하는 파이프라인을 정의한다. 그리고 4DGS 기반 모델을 구성하는 표준 (canonical) 3DGS 모델을 경량화 하기 위한 목적으로 기어도 점수 기반의 가우시안 프루닝 (pruning)을 변형 좌표계에 적용할 수 있도록 한다. 또한 본 표준은 초매개변수를 통해 상황에 따라 피쳐 임베딩의 압축률과 프루닝 (pruning) 비율 등을 설정할 수 있는 포맷을 포함한다.

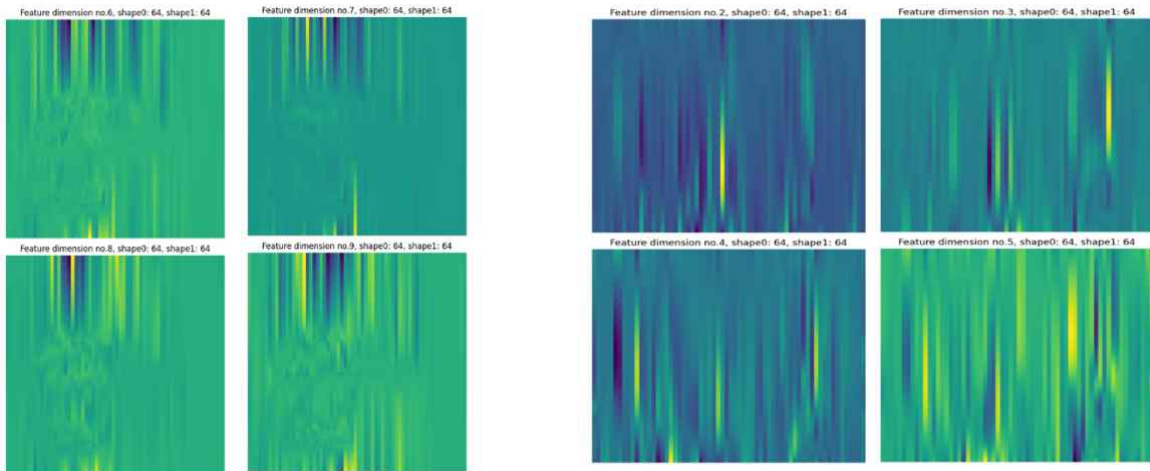


(그림 5-1) 4DGS 기반 동적 3차원 재구성 모델과 표준에서 제시하는 압축 파이프라인 구조도

## 5.2 4차원 신경 복셀을 사용한 위치적 특성 임베딩의 부호화 및 복호화

본 절에서는 4DGS 기반 3차원 재구성 모델 파일 용량의 약 30%가량을 차지하는 4차원 신경 복셀을 부호화하는 방법을 정의한다. 5.1절에 설명되었듯이 4차원 신경 복셀은 위치와 시간 정보의 임베딩을 저장하는 역할을 하며 학습 가능한 4차원 텐서 (tensor) 형태로 구성된다. 텐서의 구조 (shape)는  $[6 \times L, H, \text{높이}, \text{너비}]$ 이며, 이 때 L 은 레벨의 개수, 첫 차원이  $6 \times L$  인 이유는 x, y, z, t 4 가지 요소를 가지고 두 개씩 조합할 수 있는 경우의 수가 6이기 때문이다. 본 표준에서는 해당 4차원 텐서를 다수의 2차원 행렬로 변환하여 처리한다. 그 이유는 높이와 너비로 정의되는 2차원 행렬에 대해서는 이미지 또는 비디오 코덱으로 압축 가능하기 때문이다. 4차원 텐서를 다수의 2차원 텐서로 변환한 이후 양자화 과정을 진행한다. 기존 32비트 소수점으로 정의된 데이터를 초매개변수

설정 파일에 따라 8비트 또는 16비트로 전환할 수 있다. 자료형을 전환하여 별도의 YUV 파일로 저장하게 되는데, 이 때 각 Y 컴포넌트만 존재하는 YUV400 포맷을 사용한다. 4DGS 모델에서  $6 \times L \times H$  개의 평면을 출력할 수 있는데, H 개의 평면을 시간적으로 연결하여 화면간 예측 (inter coding) 을 적용한 비디오 코딩을 진행하게 된다. 이처럼 4차원 신경 복셀에 대해 비디오 부호화를 적용하는 이유는 신경 복셀을 분해한 2차원 텐서를 시각화한 (그림 5-2)를 보면 확인할 수 있는데 자세한 설명은 다음과 같다: (1) 특성 평면에서 인접한 위치에 존재하는 특성의 값의 지역적 유사성, (2) H 차원에 걸쳐 시간적으로 연결했을 때 인접한 차원의 유사한 위치에 분포하는 특성 값의 유사성. 해당 비디오 코딩 과정을 거친 이후 bit 확장자의 압축된 비트스트림을 획득 가능하며, 복호화 과정에서 필요한 자료형의 비트수, 너비값, 높이값 등의 메타데이터는 파일명에 포함하여 기록한다.



(그림 5-2) 분해된 특성 평면 값의 시각화 (좌: mirror 시퀀스, 우: coffee\_martini 시퀀스)

### 5.3 가우시안 프루닝 기술을 사용한 표준 가우시안 네트워크 압축

본 절에서는 4DGS 기반 3차원 재구성 모델 파일 용량의 약 50% 가량을 차지하는 표준 가우시안 네트워크 상에서 불필요한 가우시안을 프루닝하는 방법을 정의한다. (그림5-1) 의 (B) 부분이 전체 파이프라인에서 해당 과정의 역할을 나타낸다. LightGaussian [3] 에 소개된 기법으로, 전역적 기여도 점수 (global significance score) 에 따른 가우시안 프루닝 알고리즘이 존재한다. 전역적 기여도 점수는 가우시안 네트워크 내 개별 가우시안 분포마다 모든 학습 뷰 (training view)에 대해 광선 (ray) 구성시 가우시안이 교차하는지 여부를 계산한 후 만약 교차한다면 불투명도와 3D 볼륨을 곱하여 점수를 산출한다. 가우시안 네트워크 내 가우시안들을 전역적 기여도 점수 기반으로 정렬한 후 설정 파일에 명시된 초매개변수 설정값에 따라 일정 개수 또는 비율에 대한 프루닝을 지원한다.

## 부 록 I

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 필요성

#### 1.1 본 표준의 필요성

6자유도 가상현실 영상의 합성을 지원하는 기술 중 3D Gaussian splatting (3DGS) 은 고품질 영상 합성 및 FHD 해상도의 영상 기준 초당 30fps 이상의 실시간 렌더링을 지원한다는 점에서 차세대 실감미디어 기술로 주목받고 있다. 인공지능망 가중치를 통해 공간을 표현하는 신경 복사 필드 (neural radiance fields, NeRF) 기반 방법과 비교해서 광선 추적 (ray tracing) 대신 투영 (projection) 을 통해 임의 시점 합성을 제공함으로써 렌더링 차원에서 고속화를 제공한다. 정적인 장면의 재구성을 지원하는 기존 3DGS를 넘어 움직이는 물체가 존재하는 장면의 재구성까지 지원하는 4D Gaussian splatting (4DGS) 모델이 최근 제안되었다. 하지만 인공지능망을 활용한 암시적 표현의 비중이 적고 명시적으로 3차원 공간상의 타원체를 나타내는 가우시안 함수를 정의하고 개별 가우시안마다 회전, 구면 조화 함수의 계수 등을 저장한다는 점에서 매우 큰 저장 공간을 차지하게 된다. 따라서 3DGS 기반 동적 재구성 모델의 압축의 필요성이 대두된다.

이러한 저장 공간의 문제를 해결하기 위해, 4DGS 기반 모델을 구성하는 학습 가능한 4차원 신경 복셀과 표준 3DGS 네트워크를 압축할 수 있다. 본 표준에서 제시하는 시스템은 x, y, z축과 동시에 시간적 방향성 (t축)을 포함하여 구성할 수 있는 6개의 3차원 복셀을 2차원으로 분할하여 이를 기존 비디오 코덱을 활용하여 부호화한다. 이를 사용한다면 시퀀스의 길이가 증가하더라도 초매개변수로 사전 설정된 t축의 길이에 맞추어 유연하게 부호화할 수 있고, 기존 개발된 비디오 코덱을 상황에 맞추어 선택적으로 적용 가능하여 압축할 수 있다. 또한, 본 표준에서는 4DGS 모델에 공통적으로 적용되는 표준 3DGS 네트워크를 압축하는 모듈을 제시한다. 이를 통해 3DGS 네트워크 상 불필요한 가우시안을 유동적으로 제거할 수 있다.

동적 3DGS의 구성 요소 각각에 따른 모듈화된 3DGS 압축 표준은 소프트웨어 개발 과정을 효율화 함으로써 가상현실 구현을 위한 3DGS 기반 어플리케이션 개발의 활성화 효과를 나타낼 것으로 기대된다.

## 부 록 II-1

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 지식재산권 협약서 정보

#### II-1.1 지식재산권 협약서(1)

- 해당 사항 없음

## 부 록 II-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 시험인증 관련 사항

해당 사항 없음.

## 부 록 II-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 본 표준의 연계(family) 표준

해당 사항 없음.

## 부 록 II-4

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 참고 문헌

- [1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” ACM Transactions on Graphics, vol. 42, no. 4, 2023.
- [2] G. Wu et al., “4d gaussian splatting for real-time dynamic scene rendering,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20310–20320, 2024.
- [3] Z. Fan, K. Wang, K. Wen, Z. Zhu, D. Xu, and Z. Wang, “LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS,” Mar. 29, 2024.

## 부 록 II-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 영문표준 해설서

해당 사항 없음.

## 부 록 II-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

### 표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2024.12.03	제정 XDFK_01.0056/R0	-	운영위원회