다차원영상기술표준화포럼표준(국문표준)

XDFK_01.0043/R0 제정일 : 2023년 11월 20일

국문 표준명	6자유도 가상현실 영상합성을 위한 복셀 그리드 기반 모델 최적화 기법
영문 표준명	Efficient Voxel Grid-based Model Optimization for 6DoF Video Synthesis
표준초안 작성자	박준형, 류은석(성균관대학교)

XD FORUM
다차원영상기술표준화포럼

서 문

1 표준의 목적

이 표준의 목적은 딥러닝 기반 자유 시점 이미지 합성 기술을 통해 다양한 각도에서 촬영한 몰입형 동영상을 방사형 필드 (Radiance Fields) 로 모델링하여 6 자유도 동영상으로 표현하는 시스템을 제공하는 데 있다. 이때 인접 프레임간 유사정보를 활용하여 복원 속도 및 품질이 향상되고, 압축에 용이한 데이터 구조로 정보를 처리한다.

2 주요 내용 요약

이 표준은 3차원 공간에서 사용자가 바라보는 임의 시점 합성 (Novel View Synthesis, NVS) 을 위한 딥러닝 모델 훈련 시, 복원 품질을 향상 및 모델 사이즈를 압축을 위한 모델 구조 설계 규격과 효율적인 학습 전략을 기술한다. 본 표준은 딥러닝 기반 몰입형 영상 표현 시스템에 적용될 수 있다.

3 인용 표준과의 비교

이 표준은 국제 표준단체 MPEG-Implicit Neural Visual Representation 표준 기술 등을 이용하는 시스템을 위한 별도의 독립적인 표준으로서, MPEG 비디오 시스템 표준과 직접적인 관련성이 없음.

목 차

6자유도 가상현실 영상합성을 위한 복셀 그리드 기반 모델 최적화 기법 (Efficient Voxel Grid-based Model Optimization for 6DoF Video Synthesis)

1 적용 범위

본 표준의 적용 범위는 신경망 기반의 몰입형 영상 표현에서의 모델 및 입출력 데이터를 다루며, 이는 모델 구조 설계, 학습 전략 등을 포함한다. 또한, 본 표준의 구문(Syntax) 및 의미론(Semantics)은 모델 학습 시 초매개변수(Hyperparameter)로 사용되며 이는 별도의 설정 파일로 전달될 수 있다.

2 인용 표준

해당 사항 없음.

3 용어 정의

3.1 임의 시점 합성 (Novel View Synthesis)

기존 2차원 이미지나 3차원 장면에서 새로운 시점의 이미지를 생성하는 기술. 이는 증강현실, 가상현실 등에서 사용되어, 사용자가 3차원 공간을 자유롭게 탐색할 수 있게 한다.

3.2 Structure from Motion

여러 시점의 이미지로부터 3차원 구조를 복원하는 기술. 이 기술은 특성 추출과 매칭을 통해 카메라 매개변수를 복원하고, 이를 바탕으로 3차원 모델을 재구성한다.

3.3 방사형 필드 (Radiance Fields)

3차원 공간에서 빛의 방향과 위치에 따라 색과 밝기를 인코딩하는 함수. 이를 통해 3차원 장면을 표현할 수 있다.

3.4 레이 샘플링 (Ray Sampling)

3차원 그래픽에서 렌더링을 위한 기술로, 가상의 카메라로부터 발사된 레이를 추적하여 물체와의 교차점에서의 색상, 밝기 등을 계산한다. 광원으로부터의 빛의 행동을 시뮬레이션하여 사실적인 이미지를 생성할 수 있게 한다.

3.5 초매개변수 (Hyperparameter)

머신 러닝 모델 학습에 영향을 주는 사용자 설정 값. 이는 학습률, 배치 크기 등을 포함하며, 모델 성능과 학습 방식을 결정한다.

3.6 볼륨 렌더링(Volume Rendering)

3차원 데이터를 2차원 이미지로 변환하는 과정으로, 각 복셀의 색상과 투명도 값을 활용하여 2차원 화면에 3차원 데이터를 정확하게 표현한다. 이 과정에서 렌더링 알고리즘은 복셀 간의 교차 및 오버랩을 계산하여 최종 이미지를 생성한다.

3.7 복셀 (Voxel)

3차원 공간에서의 픽셀로, 3차원 공간의 한 지점의 정보를 표현한다.

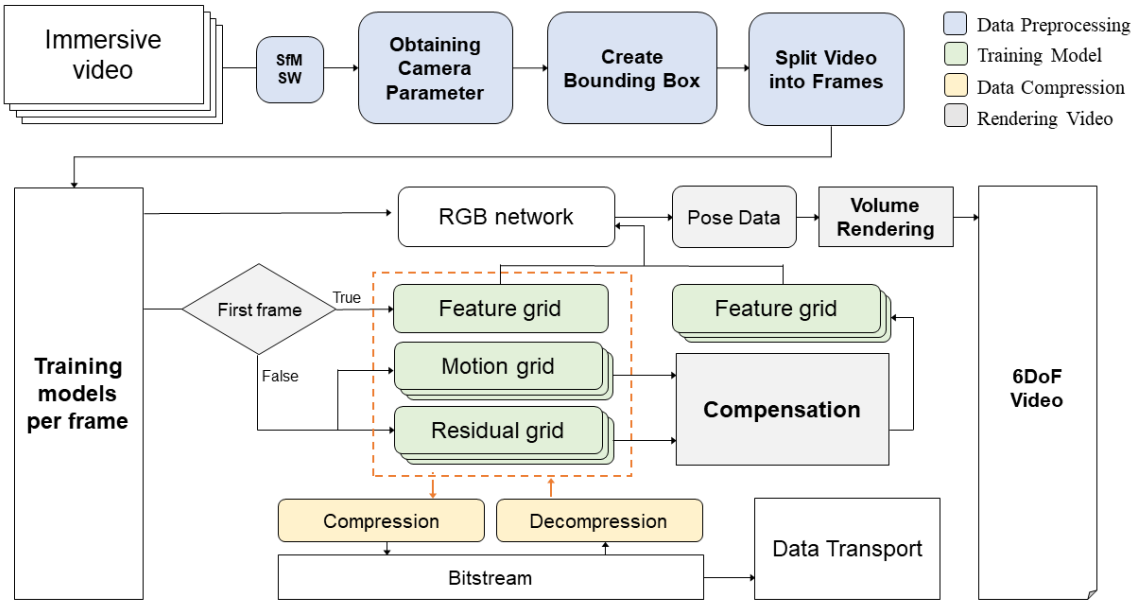
4 약어

MLP	Multi Layer Perceptron
NeRF	Neural Radiance Fields
NVS	Novel View Synthesis
SfM	Structure from Motion

5 복셀 그리드 기반 방사형 필드 재구성 모델 최적화 기법

5.1 프레임간 유사 정보를 사용한 6 자유도 가상현실 영상합성 프레임워크

신경 복사 필드 (NeRF) 기술은 다양한 각도에서 촬영된 2차원 이미지들을 활용하여 3차원 공간 내 임의의 좌표의 색상과 밀도 값을 근사하는 모델을 학습하는 딥러닝 기반의 방법론이다. 본 표준에서 제시하는 시스템은 해당 기술의 발전 동향을 반영하여 3차원 공간의 특성 값을 복셀 그리드 구조로 모델링하고 이를 직접 최적화하는 방식을 사용한다. 이는 기존 NeRF 기술에서 높은 계산 비용을 요구하는 신경망 구조로 인해 실시간 장면 복원이 어려웠던 문제를 개선한 효율적인 방식이다. 이러한 설계 구조를 통해 3차원 장면 복원이 보다 신속하게 이루어질 수 있다.



(그림 5-1) 복셀 그리드 기반 6 자유도 영상 합성 시스템 구조도

(그림 5-1)은 복셀 그리드를 기반으로 한 6 자유도 영상 합성 시스템의 구조를 보여준다. 이 시스템은 다수의 카메라를 이용해 얻은 몰입형 영상을 입력으로 받아, 임의의 사용자 시점에서 6 자유도 동영상을 출력한다. 전체 시스템은 데이터 전처리, 모델 학습, 데이터 압축, 영상 복원의 네 단계로 구성된다.

데이터 전처리 단계에서는 모델 학습에 필요한 데이터를 생성하며, 이때 SfM 소프트웨어를 이용해 입력 영상의 카메라 매개변수를 추정한다. 다음으로, 시퀀스 크기에 맞춰 레이 샘플링 경계를 설정하고, 동영상을 프레임 단위의 이미지 파일로 분할하여 학습을 위한 데이터를 준비한다.

모델 학습 단계에서는 추출된 프레임 이미지와 카메라 정보를 활용하여 복셀 그리드 형태로 방사형 필드를 모델링한다. 이 과정에서, 좌표 기반의 작은 MLP를 통해 각 시점의

색상 정보를 디코딩한다. 이렇게 하면, 전체 방사형 필드를 직접 모델링하지 않고도 효율적으로 처리할 수 있다. 특히, 본 시스템은 프레임 간의 학습을 진행하며, 프레임 간 유사 정보, 즉 잔차 정보와 모션 정보를 복셀 그리드 형태로 추가적으로 모델링한다.
이러한 방식의 확장은 다양한 프레임 간의 유사성을 고려하여 첫 프레임을 기반으로 이후 프레임들을 순차적으로 복원한다. 잔차 정보와 모션 정보는 대부분의 영역에서 변화가 없거나 유사하기 때문에 희소한 구조를 형성한다. 이러한 구조는 높은 압축률을 달성하며, 특징 정보를 효율적으로 보관할 수 있게 한다. 이는 기존 동적 방사형 필드 모델들이 각 프레임의 모든 특징값을 보관하였기에 발생했던 비효율성을 해결하며, 더욱 효율적인 처리를 가능케 한다.
데이터 압축 단계에서는 모델링 된 복셀 그리드를 압축하여 좌표 기반의 MLP와 함께 재생 장치로 전송한다. 렌더링 단계에서는 잔차 정보와 모션 정보 그리드를 활용하여 현재 프레임의 특징 정보를 복원하고, 출력할 시점의 자세 데이터를 입력 받아 볼륨 렌더링을 진행하여 6자유도 영상을 출력한다.

5.2 데이터 형식에 따른 모델 변형

5.2.1 복셀 그리드 구조 변형

3차원 공간을 복셀 그리드로 모델링할 때, 입력되는 몰입형 영상이 배경이 없는 객체 데이터인지, 아니면 배경과 전경을 모두 포함한 실사 데이터인지에 따라 모델의 형태를 변형해야 한다. 객체 데이터의 경우, 모델은 특정 대상의 정보 학습에 중점을 둔다. 3차원 공간을 큐브 형태의 복셀 그리드로 나누어 모델링하고, 마스킹 정보를 활용하여 객체가 할당되지 않은 공간의 학습률을 조정함으로써 3차원 공간을 더욱 간결하고 효율적으로 표현할 수 있다.
장면 데이터는 전경과 배경 정보를 모두 포함하므로, 장면의 각 부분을 세밀하게 표현해야 한다. 따라서 3차원 공간을 큐브 형태가 아닌, 여러 겹의 2차원 평면 그리드로 나타낸다. 이러한 구조를 바탕으로 레이 샘플링 시 각 2차원 그리드를 개별적으로 샘플링하면 복원의 정밀도가 향상된다. 이러한 구조 변형을 통해 실사 데이터의 전반적인 정보와 디테일을 보다 정확하게 모델링할 수 있다.

5.2.2 학습 전략 변형

데이터의 특성에 따라 학습 전략이 유연하게 변형될 수 있어야 한다. 방사형 필드를 표현하는 신경망을 학습할 때, 계층적으로 동작하는 초기 네트워크 (Coarse Network)와 상세 네트워크 (Fine Network)를 사용한다. 초기 네트워크가 먼저 학습되어 초기 매개변수를 설정하고, 이 결과는 상세 네트워크의 입력으로 사용된다. 상세 네트워크는 이를 기반으로 보다 정교한 모델링을 수행한다. 이러한 구조는 모델이 전반적인 구조를 빠르게 파악하고, 이후에 세부 정보를 조정하며 학습하도록 돕는다.

객체 데이터의 경우, 초기 네트워크 학습 과정에서 대상 객체의 위치와 형태를 대략적으로 파악하는 것이 중요하다. 이 단계에서, 모델은 객체의 주요 특성과 구조에 대한 이해를 바탕으로 뒷 단계의 상세한 학습을 위한 기반을 마련한다. 마스킹 정보를 활용하여 학습률을 최적화하고, 객체가 없는 영역은 학습에서 배제하여 학습의 효율성을 높인다.
장면 데이터의 경우, 전경과 배경의 모든 정보가 중요하므로, 초기 네트워크 학습 단계를 생략하고 바로 각 2차원 평면 그리드를 상세하게 학습한다. 여러 겹의 2차원 평면 그리드가 독립적으로 샘플링되어 학습되며, 각 겹은 장면의 다양한 면을 세밀하게 표현한다. 이러한 접근 방식을 통해, 장면의 전반적인 정보와 디테일이 정확하게 복원되며, 복원의 정밀도가 향상된다.

부 록 Ⅰ

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

필요성

Ⅰ.1 본 표준의 필요성

6 자유도 영상은 방대한 양의 데이터를 포함하고 있기에 처리와 관리에 대한 문제가 따른다. 최근에는 신경 복사 필드 (Neural Radiance Fields, NeRF)와 같은 신경망 기반 방법론이 제안되었다. NeRF는 연속적인 공간 좌표를 RGB와 밀도를 가진 방사형 필드로 출력하는 다층 퍼셉트론 (Multi-Layer Perceptron, MLP)으로 3차원 공간을 표현한다.
신경망 기반 표현은 상당한 연산량을 필요로 해 많은 컴퓨팅 자원과 긴 훈련 시간이 요구된다. 이 문제를 해결하기 위해, 다양한 외부 데이터를 통합하여 사용하는 복합 구조가 제안되었다. 명시적 특성 볼륨 (Feature Volume)을 활용한 최근의 기술들은 정적 장면에 대해 효과적인 성능 향상을 달성했다.
동적 방사형 필드를 표현하기 위해, 각 프레임을 독립적인 정적 장면으로 보고 각각에 정적 방법론을 적용할 수 있다. 그러나 이러한 방식은 시간에 따른 일관성 정보를 고려하지 않아, 시퀀스가 길어질수록 비효율적이고 복원 품질이 저하된다.
동적 방사형 필드를 표현하는 최신 기술들은 변하지 않는 표준 (Canonical) 특성 공간 (Feature Space)을 유지하고, 각 프레임을 이 표준 공간에 맞게 왜곡하여 재구성하는 방식을 사용한다. 그러나 이러한 표준 공간에 의존하는 방식은 동영상 내부의 큰 움직임이나 토폴로지 변화를 잘 표현하지 못하며, 시퀀스의 길이가 길어질수록 왜곡 매핑에 필요한 연산량이 증가하여 훈련 오버헤드가 증가한다.
본 표준에서 제시하는 시스템은 인접한 타임스탬프 사이의 방사형 필드의 잔차를 특성 공간에서 명시적으로 모델링한다. 이 방법은 글로벌 좌표 기반의 소형 MLP를 활용하여, 연속적인 시공간의 특성 공간의 입력으로부터 방사형 필드의 색상 값과 밀도 값을 근사해 출력한다. 또한, 훈련과 추론에서 높은 효율성을 유지하기 위해, 명시적인 복셀 그리드 표현을 사용하여 특성 공간을 모델링한다.
첫 번째 키 프레임에서 모델 학습을 통해 얻어진 그리드 볼륨은 초기 특성 볼륨으로 사용되며, 각 후속 프레임은 저해상도의 모션 그리드와 잔차 그리드로 표현된다. 이 설계 방식의 핵심은 두 개의 그리드를 사용하여 후속 특성 그리드를 연속적으로 획득할 수 있어, 글로벌 표준 공간의 도입 없이 모든 프레임을 표현할 수 있다는 것이다. 모션 그리드와 잔차 그리드는 기존 특성 공간에 비해 압축에 적합하므로, 대상이 되는 6 자유도 동영상이 빠른 움직임을 포함하거나, 많은 수의 프레임을 포함하더라도 효과적으로 표현할 수 있다.

부 록 II-1

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

지식재산권 확약서 정보

II-1.1 지식재산권 확약서(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

해당 사항 없음.

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	2023.11.20	제정 XDFK.01.0043/R0	-	운영위원회