



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0092066
(43) 공개일자 2025년06월23일

- | | |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)
H04N 19/597 (2014.01) G06N 3/0455 (2023.01)
G06N 3/0475 (2023.01) H04N 19/124 (2014.01)
H04N 19/70 (2014.01)</p> <p>(52) CPC특허분류
H04N 19/597 (2015.01)
G06N 3/0455 (2023.01)</p> <p>(21) 출원번호 10-2024-0165811</p> <p>(22) 출원일자 2024년11월20일
심사청구일자 없음</p> <p>(30) 우선권주장
1020230181372 2023년12월14일 대한민국(KR)</p> | <p>(71) 출원인
현대자동차주식회사
서울특별시 서초구 현릉로 12 (양재동)
기아 주식회사
서울특별시 서초구 현릉로 12 (양재동)
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)</p> <p>(72) 발명자
강제원
서울특별시 마포구 독막로 145, 104-904
이정경
서울특별시 양천구 목동동로 350 506동 1201호
(뒷면에 계속)</p> <p>(74) 대리인
이철희</p> |
|--|--|

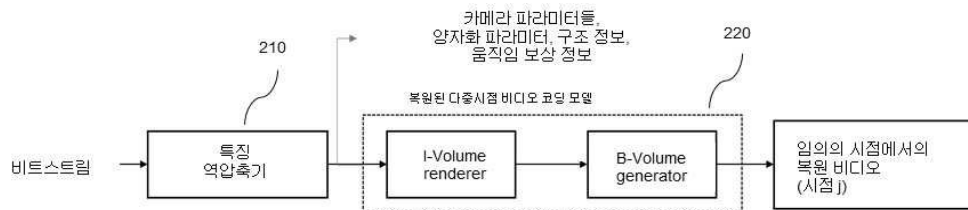
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 동적 3차원 공간 정보를 압축 및 전달하는 방법

(57) 요약

본 실시예는 동적 3차원 공간 정보의 압축을 이용하는 동적 볼륨 코딩방법 및 장치를 개시한다. 본 실시예에서, 볼륨 복호화 장치는 I 볼륨을 표현하는 텐서 특징들, 텐서 생성 모델의 가중치들, 및 부가 정보를 복호화한다. 볼륨 복호화 장치는 텐서 생성 모델의 가중치들, 및 양자화 파라미터와 관련된 정보를 이용하여 텐서 생성 모델의 구조에 관한 정보에 부합하도록 텐서 생성 모델을 복원한다. 볼륨 복호화 장치는 텐서 생성 모델을 이용하여 최인접하는 두 개의 기복호화된 I 볼륨을 기반으로 B 볼륨의 텐서 특징들을 재구성한다. 볼륨 복호화 장치는 현재 시간 및 현재 시점(view point)을 획득한다. 볼륨 복호화 장치는 현재 시간이 I 볼륨의 복호화에 대응하는 경우, I 볼륨의 텐서 특징들에 기초하여 현재 시점의 영상을 생성한다. 볼륨 복호화 장치는 현재 시간이 B 볼륨의 복호화에 대응하는 경우, 재구성된 B 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성한다.

대표도 - 도2



(52) CPC특허분류

G06N 3/0475 (2023.01)

H04N 19/124 (2015.01)

H04N 19/70 (2015.01)

(72) 발명자

신주연

서울특별시 서대문구 성산로24길 26 301호

최정아

경기도 화성시 남양읍 현대연구소로 150

허진

경기도 화성시 남양읍 현대연구소로 150

박승욱

경기도 화성시 남양읍 현대연구소로 150

명세서

청구범위

청구항 1

볼륨 복호화 장치가 수행하는, 다중시점 비디오를 복원하는 방법에 있어서,

비트스트림으로부터 I 볼륨을 표현하는 텐서 특징들, 텐서 생성 모델의 가중치들, 및 부가 정보를 복호화하는 단계, 여기서, 상기 I 볼륨의 텐서 특징들은 분해된 텐서들(decomposed tensors)에 의해 표현되고, 상기 텐서 생성 모델은 B 볼륨의 재구성에 사용되고, 상기 부가 정보는 양자화 파라미터와 관련된 정보, 및 상기 텐서 생성 모델의 구조에 관한 정보를 포함함;

상기 텐서 생성 모델의 가중치들, 및 상기 양자화 파라미터와 관련된 정보를 이용하여 상기 텐서 생성 모델의 구조에 관한 정보에 부합하도록 상기 텐서 생성 모델을 복원하는 단계;

상기 텐서 생성 모델을 이용하여 최인접하는 두 개의 기복호화된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계;

현재 시간 및 현재 시점(view point)을 획득하는 단계; 및

상기 현재 시간이 상기 I 볼륨의 복호화에 대응하는 경우, 상기 I 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계

상기 현재 시간이 상기 B 볼륨의 복호화에 대응하는 경우, 재구성된 B 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계

를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 I 볼륨의 텐서 특징들을 복호화하는 단계는,

비디오 형식으로 패키징된 I 볼륨의 텐서 특징들을 복호화하는 단계; 및

상기 비디오 형식의 I 볼륨의 텐서 특징들에 비디오 복호화 방법을 적용하여, 언패킹된 I 볼륨의 텐서 특징들을 생성하는 단계

를 포함하는, 방법.

청구항 3

제1항에 있어서,

상기 I 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계는,

상기 I 볼륨의 분해된 텐서들에 기초하여 상기 현재 시점에 대응하는 복셀들에서 색상들 및 볼륨 밀도들을 산정하는 단계; 및

상기 색상들 및 상기 볼륨 밀도들에 기초하여 상기 현재 시점의 영상을 생성하는 단계

를 포함하는, 방법.

청구항 4

제1항에 있어서,

상기 부가 정보로서, 상기 I 볼륨 및 상기 B 볼륨을 포함하는 GOV(Group of Volume)의 구조에 관한 정보를 복호화하는 단계를 더 포함하고,

상기 B 볼륨의 텐서 특징들을 재구성하는 단계는,

상기 GOV(Group of Volume)의 구조에 관한 정보에 기초하여 상기 최인접하는 두 개의 I 볼륨의 위치를 결정하는, 방법.

청구항 5

제1항에 있어서,

상기 B 볼륨의 텐서 특징들을 재구성하는 단계는,

상기 최인접하는 두 개의 I 볼륨으로부터 동일한 축을 갖는 텐서들을 획득하는 단계; 및

상기 동일한 축을 갖는 텐서들을 상기 텐서 생성 모델에 입력하여, 생성 텐서를 구성하는 단계를 포함하는, 방법.

청구항 6

제5항에 있어서,

상기 B 볼륨의 텐서 특징들을 재구성하는 단계는,

인접하는 I 볼륨으로부터 상기 생성 텐서와 직교하는 고정 텐서를 획득하는 단계; 및

상기 고정 텐서 및 상기 생성 텐서에 기초하여 상기 B 볼륨의 분해된 텐서들을 재구성하는 단계를 포함하는, 방법.

청구항 7

제5항에 있어서,

상기 텐서 생성 모델은, 인코더와 디코더를 포함하고,

상기 생성 텐서를 구성하는 단계는,

상기 동일한 축을 갖는 텐서들을 상기 인코더에 입력하여 중간 특징을 생성하는 단계;

상기 중간 특징에 상기 현재 시간을 누적하는 단계; 및

상기 현재 시간이 누적된 중간 특징을 상기 디코더에 입력하여 상기 생성 텐서를 구성하는 단계를 포함하는, 방법.

청구항 8

볼륨 부호화 장치가 수행하는, 다중시점 비디오를 부호화하는 방법에 있어서,

현재 시간 및 현재 시점(view point)을 획득하는 단계;

상기 현재 시간이 I 볼륨의 부호화에 대응하는 경우, I 볼륨을 획득하고, 상기 I 볼륨을 분해하여 상기 I 볼륨의 텐서 특징들을 표현하는 저차원의 텐서들을 생성하는 단계;

상기 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 텐서 생성 모델을 이용하여 최인접하는 두 개의 기분해된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계;

상기 I 볼륨의 텐서 특징들을 부호화하는 단계; 및

상기 텐서 생성 모델의 가중치들을 부호화하는 단계를 포함하는, 방법.

청구항 9

제8항에 있어서,

상기 현재 시간이 상기 I 볼륨 또는 상기 B 볼륨의 부호화에 대응하는지에 따라, 상기 I 볼륨의 텐서 특징들 또

는 재구성된 B 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계를 더 포함하는, 방법.

청구항 10

제9항에 있어서,

상기 현재 시점의 영상과 원본 영상 간의 차이에 기초하는 손실함수를 산정하는 단계;

상기 손실함수를 감소시키는 방향으로, 상기 I 볼륨의 텐서 특징들을 업데이트하는 단계; 및

상기 손실함수를 감소시키는 방향으로, 상기 텐서 생성 모델의 가중치들을 업데이트하는 단계를 더 포함하는, 방법.

청구항 11

제8항에 있어서,

상기 I 볼륨 및 상기 B 볼륨을 포함하는 GOV(Group of Volume)의 구조에 관한 정보를 획득하는 단계;

상기 GOV의 구조에 관한 정보에 기초하여 상기 최인접하는 두 개의 I 볼륨의 위치를 결정하는 단계; 및

부가 정보로서, 상기 GOV의 구조에 관한 정보를 부호화하는 단계

를 더 포함하는, 방법.

청구항 12

제8항에 있어서,

상기 B 볼륨의 텐서 특징들을 재구성하는 단계는,

상기 최인접하는 두 개의 I 볼륨으로부터 동일한 축을 갖는 텐서들을 획득하는 단계; 및

상기 동일한 축을 갖는 텐서들을 상기 텐서 생성 모델에 입력하여, 생성 텐서를 구성하는 단계를 포함하는, 방법.

청구항 13

제12항에 있어서,

상기 B 볼륨의 텐서 특징들을 재구성하는 단계는,

인접하는 I 볼륨으로부터 상기 생성 텐서와 직교하는 고정 텐서를 획득하는 단계; 및

상기 고정 텐서 및 상기 생성 텐서에 기초하여 상기 B 볼륨의 분해된 텐서들을 재구성하는 단계를 포함하는, 방법.

청구항 14

제8항에 있어서,

상기 I 볼륨의 텐서 특징들을 부호화하는 단계는,

상기 I 볼륨의 텐서 특징들을 비디오 형식으로 패킹하는 단계; 및

패킹된 I 볼륨의 텐서 특징들에 비디오 부호화 방법을 적용하여 비트스트림을 생성하는 단계를 포함하는 방법.

청구항 15

제8항에 있어서,

상기 텐서 생성 모델의 가중치들을 부호화하는 단계는,
 상기 텐서 생성 모델의 가중치들을 프루닝하는 단계;
 양자화 파라미터와 관련된 정보를 이용하여 프루닝된 가중치들을 양자화하는 단계; 및
 양자화된 가중치들을 부호화하는 단계
 를 포함하는, 방법.

청구항 16

제16항에 있어서,
 상기 양자화 파라미터와 관련된 정보, 및 상기 텐서 생성 모델의 구조에 관한 정보를 포함하는 부가 정보를 부호화하는 단계를 더 포함하는, 방법.

청구항 17

영상 복호화 장치에게 비디오 데이터를 제공하는 방법에 있어서,
 상기 비디오 데이터를 비트스트림으로 부호화하는 단계; 및
 상기 비트스트림을 상기 영상 복호화 장치로 전달하는 단계
 를 포함하고,
 상기 비디오 데이터를 부호화하는 단계는,
 현재 시간 및 현재 시점(view point)을 획득하는 단계;
 상기 현재 시간이 I 볼륨의 부호화에 대응하는 경우, I 볼륨을 획득하고, 상기 I 볼륨을 분해하여 상기 I 볼륨의 텐서 특징들을 표현하는 저차원의 텐서들을 생성하는 단계;
 상기 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 텐서 생성 모델을 이용하여 최인접하는 두 개의 기분해된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계;
 상기 I 볼륨의 텐서 특징들을 부호화하는 단계; 및
 상기 텐서 생성 모델의 가중치들을 부호화하는 단계
 를 포함하는 것을 특징으로 하는, 방법.

발명의 설명

기술 분야

[0001] 본 개시는 동적 3차원 공간 정보의 압축을 이용하는 동적 볼륨 코딩방법 및 장치에 관한 것이다.

배경 기술

[0002] 이하에 기술되는 내용은 단순히 본 실시예와 관련되는 배경 정보만을 제공할 뿐 종래기술을 구성하는 것이 아니다.

[0003] 최근 이미지를 비롯한 각종 데이터의 표현을 신경망 구조로 나타내는 암시적 신경망 표현 모델에 관한 연구가 활발하다. 기존의 3D(Dimensional) 비디오를 표현하기 위해, 다시점 비디오와 깊이 비디오가 함께 취득되거나 라이트필드 비디오, 360도 다시점 비디오 등이 취득된다. 기존의 3D 비디오 표현 방식으로서, 픽셀 위치별 RGB 픽셀값을 갖는 명시적(explicit) 표현 방식이 사용된다. 이러한 명시적 표현 방식을 대체하기 위해, 공간의 픽셀 위치인 (x, y, z) 좌표 및 시점 방향 (θ, ϕ) 으로 구성된 5차원 입력으로부터 (r, g, b) 값과 볼륨 밀도를 출력하는 함수를 신경망으로 표현하는 암시적 신경망 표현(Implicit Neural Representation, INR) 방식이 도입되고 있다. 명시적 표현 방식에 비하여 암시적 신경망 표현 방식은, 임의의 시점에서 좌표에 따라 (r, g, b) 값을 즉각적으로 복원하므로, 기존 디코더 대비 실용적 우수성을 가질 수 있다.

[0004] 암시적 신경 표현 방식 중 3D 비디오의 합성에 사용하는 대표 모델로서 NeRF(Neural Radiance Fields) 기술이 존재한다. 정적인 공간을 표현하는 정적(static) NeRF 기술은, 전술한 5차원 입력 벡터에 위치 임베딩(positional embedding)을 적용하는 전처리를 수행하고, 전처리된 입력 벡터를 간단한 다층신경망(Multilayer Perceptron, MLP)에 입력하여 픽셀값과 볼륨 밀도를 출력한다. 하지만, 작은 크기의 네트워크에 기초하여 대상 3D의 공간 정보를 표현하는 것은 한계가 존재한다. 또한, 시간적으로 변화하는 공간 정보를 표현하는 것이 어려울 수 있다. 전술한 문제를 극복하기 위하여 동적 NeRF 기술이 제안되지만, 네트워크의 크기가 방대해지고, 부호화 효율이 감소될 수 있다. 또한, 동적 NeRF 모델을 위한 전용 코덱의 개발은 종래의 코덱 기술과 호환성 측면에서 실용성이 낮아질 수 있다.

[0005] 따라서, 동적 3D 비디오 부호화 표현 기반의 효율을 향상시키고 동적 3D 비디오 화질을 개선하기 위해, 동적 3D 비디오의 공간 정보를 효율적으로 재구성하는 방안이 고려될 필요가 있다.

발명의 내용

해결하려는 과제

[0006] 본 개시는, 텐서분해 기법에 기초하여 동적 3D 공간 정보를 저차원의 행렬과 벡터로 분할하고, 2D 비디오 코덱에 기초하여 분할된 정보를 압축하며, 복원된 동적 3D 공간 정보에 기초하여 디코더 측에서 임의의 시점의 비디오를 생성하는 동적 볼륨 코딩방법 및 장치를 제공하는 데 목적이 있다.

과제의 해결 수단

[0007] 본 개시의 실시예에 따르면, 볼륨 복호화 장치가 수행하는, 다중시점 비디오를 복원하는 방법에 있어서, 비트스트림으로부터 I 볼륨을 표현하는 텐서 특징들, 텐서 생성 모델의 가중치들, 및 부가 정보를 복호화하는 단계, 여기서, 상기 I 볼륨의 텐서 특징들은 분해된 텐서들(decomposed tensors)에 의해 표현되고, 상기 텐서 생성 모델은 B 볼륨의 재구성에 사용되고, 상기 부가 정보는 양자화 파라미터와 관련된 정보, 및 상기 텐서 생성 모델의 구조에 관한 정보를 포함함; 상기 텐서 생성 모델의 가중치들, 및 상기 양자화 파라미터와 관련된 정보를 이용하여 상기 텐서 생성 모델의 구조에 관한 정보에 부합하도록 상기 텐서 생성 모델을 복원하는 단계; 상기 텐서 생성 모델을 이용하여 최인접하는 두 개의 기복호화된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계; 현재 시간 및 현재 시점(view point)을 획득하는 단계; 및 상기 현재 시간이 상기 I 볼륨의 복호화에 대응하는 경우, 상기 I 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계상기 현재 시간이 상기 B 볼륨의 복호화에 대응하는 경우, 재구성된 B 볼륨의 텐서 특징들에 기초하여 상기 현재 시점의 영상을 생성하는 단계를 포함하는, 방법을 제공한다.

[0008] 본 개시의 다른 실시예에 따르면, 볼륨 부호화 장치가 수행하는, 다중시점 비디오를 부호화하는 방법에 있어서, 현재 시간 및 현재 시점(view point)을 획득하는 단계; 상기 현재 시간이 I 볼륨의 부호화에 대응하는 경우, I 볼륨을 획득하고, 상기 I 볼륨을 분해하여 상기 I 볼륨의 텐서 특징들을 표현하는 저차원의 텐서들을 생성하는 단계; 상기 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 텐서 생성 모델을 이용하여 최인접하는 두 개의 기분해된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계; 상기 I 볼륨의 텐서 특징들을 부호화하는 단계; 및 상기 텐서 생성 모델의 가중치들을 부호화하는 단계를 포함하는, 방법을 제공한다.

[0009] 본 개시의 다른 실시예에 따르면, 영상 복호화 장치에게 비디오 데이터를 제공하는 방법에 있어서, 상기 비디오 데이터를 비트스트림으로 부호화하는 단계; 및 상기 비트스트림을 상기 영상 복호화 장치로 전달하는 단계를 포함하고, 상기 비디오 데이터를 부호화하는 단계는, 현재 시간 및 현재 시점(view point)을 획득하는 단계; 상기 현재 시간이 I 볼륨의 부호화에 대응하는 경우, I 볼륨을 획득하고, 상기 I 볼륨을 분해하여 상기 I 볼륨의 텐서 특징들을 표현하는 저차원의 텐서들을 생성하는 단계; 상기 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 텐서 생성 모델을 이용하여 최인접하는 두 개의 기분해된 I 볼륨을 기반으로 상기 B 볼륨의 텐서 특징들을 재구성하는 단계; 상기 I 볼륨의 텐서 특징들을 부호화하는 단계; 및 상기 텐서 생성 모델의 가중치들을 부호화하는 단계를 포함하는 것을 특징으로 하는, 방법을 제공한다.

발명의 효과

[0010] 이상에서 설명한 바와 같이 본 실시예에 따르면, 텐서분해 기법에 기초하여 동적 3D 공간 정보를 저차원의 행렬과 벡터로 분할하고, 2D 비디오 코덱에 기초하여 분할된 정보를 압축하며, 복원된 동적 3D 공간 정보에 기초하여 디코더 측에서 임의의 시점의 비디오를 생성하는 동적 볼륨 코딩방법 및 장치를 제공함으로써, 동적 3D 비디오

부호화 효율 및 동적 3D 비디오 화질을 개선하는 것이 가능해지는 효과가 있다.

[0011] 또한 본 실시예에 따르면, 동적 3D 공간 정보를 효율적으로 압축 및 복원하는 동적 볼륨 코딩방법 및 장치를 제 공함으로써, UHD(Ultra High Definition) 비디오, 게임 방송, 360도 동영상 스트리밍, VR/AR(Virtual Reality/Augmented Reality) 영상, 온라인 강의 등과 같은 다양한 콘텐츠에서 비트율 감소에 기초하여 네트워크 의 부담을 경감시키고, 영상 재생 가능한 기기에 대한 에너지 소비를 절감하는 것이 가능해지는 효과가 있다.

도면의 간단한 설명

[0012] 도 1은 본 개시의 일 실시예에 따른, 다중시점 비디오 코딩 기반 동적 볼륨 부호화 장치를 나타내는 블록도이다.

도 2는 본 개시의 일 실시예에 따른, 다중시점 비디오 코딩 기반 동적 볼륨 복호화 장치를 나타내는 블록도이다.

도 3은 딥러닝 기반 신경망으로서 전방향 네트워크를 나타내는 예시도이다.

도 4는 본 개시의 일 실시예에 따른 콘볼루션 레이어의 연산을 나타내는 예시도이다.

도 5는 SISR(Single Image Super Resolution) 네트워크를 나타내는 예시도이다.

도 6은 SISR에 이용되는 잔차 블록을 나타내는 예시도이다.

도 7은 NeRF(Neural Radiance Fields)의 임의의 시점의 추출과 관련된 좌표계들을 나타내는 예시도이다.

도 8은 본 개시의 일 실시예에 따른 GOV(Group of Volume)를 나타내는 예시도이다.

도 9a 및 도 9b는 본 개시의 일 실시예에 따른 텐서 분해를 나타내는 예시도이다.

도 10은 본 개시의 일 실시예에 따른 텐서 생성 모델을 나타내는 예시도이다.

도 11은 본 개시의 다른 실시예에 따른 텐서 생성 모델을 나타내는 예시도이다.

도 12a 내지 도 12c는 본 개시의 일 실시예에 따른, B 볼륨의 생성을 나타내는 예시도이다.

도 13은 본 개시의 일 실시예에 따른, 신경망 압축 방식의 부호화 파이프라인을 나타내는 예시도이다.

도 14는 본 개시의 일 실시예에 따른, 신경망 압축 방식의 복호화 파이프라인을 나타내는 예시도이다.

도 15는 본 개시의 일 실시예에 따른, 볼륨 부호화 장치가 다중시점 비디오를 부호화하는 방법을 나타내는 순서 도이다.

도 16은 본 개시의 일 실시예에 따른, 볼륨 복호화 장치가 다중시점 비디오를 복원하는 방법을 나타내는 순서도 이다.

발명을 실시하기 위한 구체적인 내용

[0013] 이하, 본 개시의 일부 실시예들을 예시적인 도면을 참조하여 상세하게 설명한다. 각 도면의 구성요소들에 참조 부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부 호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 실시예들을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 실시예들의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0014] 본 실시예는, 텐서분해 기법에 기초하여 동적 3D 공간 정보를 저차원의 행렬과 벡터로 분할하고, 2D 비디오 코 텍에 기초하여 분할된 정보를 압축하며, 압축된 정보로부터 복원된 동적 3D 공간 정보에 기초하여 디코더 측에 서 임의의 시점의 비디오를 생성하는 동적 볼륨 코딩방법 및 장치를 제공한다.

[0015] 도 1은 본 개시의 일 실시예에 따른, 다중시점 비디오 코딩 기반 동적 볼륨 부호화 장치를 나타내는 블록도이다.

[0016] 도 1의 예시에서 다중시점 비디오 코딩 기반 동적 볼륨 부호화 장치(이하, '볼륨 부호화 장치')는 다중시점에서 동적 3D 비디오를 생성하는 것을 학습하도록 다중시점 비디오 코딩 모델을 생성(즉, 트레이닝)하고, 학습된 다 중시점 비디오 코딩 모델을 압축하여 비트스트림을 생성하며, 생성된 비트스트림을 전송한다. 볼륨 부호화 장치는 다중시점 비디오 코딩 모델을 포함하는 다중시점 비디오 부호화부(110) 및 특정 압축기(feature compressor,

120)를 포함한다. 여기서, 본 개시에 따른 볼륨 부호화 장치에 포함되는 구성요소가 반드시 이에 한정되는 것은 아니다. 볼륨 부호화 장치는 다중시점 비디오 코딩 모델을 트레이닝하기 위한 트레이닝부(미도시)를 추가로 구비하거나, 외부의 트레이닝부와 연동되는 형태로 구현될 수 있다.

[0017] 볼륨 부호화 장치의 각 구성요소는 하드웨어 또는 소프트웨어로 구현되거나, 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, 각 구성요소의 기능이 소프트웨어로 구현되고 마이크로프로세서가 각 구성요소에 대응하는 소프트웨어의 기능을 실행하도록 구현될 수도 있다.

[0018] 볼륨 부호화 장치는 부호화된 비디오 데이터의 비트스트림을 비일시적인 기록매체에 저장하거나 통신 네트워크를 이용하여 다중시점 비디오 코딩 기반 동적 볼륨 복호화 장치로 전송할 수 있다.

[0019] 도 2는 본 개시의 일 실시예에 따른, 다중시점 비디오 코딩 기반 동적 볼륨 복호화 장치를 나타내는 블록도이다.

[0020] 도 2의 예시에서 다중시점 비디오 코딩 기반 동적 볼륨 복호화 장치(이하, '볼륨 복호화 장치')는 비트스트림을 역압축(decompression)하여 다중시점 비디오 코딩 모델을 복원하고, 복원된 다중시점 비디오 코딩 모델을 이용하여 다중시점에서 비디오를 생성한다. 볼륨 복호화 장치는 특징 역압축기(feature decompressor, 210), 및 복원된 다중시점 비디오 코딩 모델을 포함하는 다중시점 비디오 복호화부(220)를 포함한다.

[0021] 도 1에 예시된 볼륨 부호화 장치와 마찬가지로, 볼륨 복호화 장치의 각 구성요소는 하드웨어 또는 소프트웨어로 구현되거나, 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, 각 구성요소의 기능이 소프트웨어로 구현되고 마이크로프로세서가 각 구성요소에 대응하는 소프트웨어의 기능을 실행하도록 구현될 수도 있다.

[0022] 이하, 볼륨 및 3D 공간은 호환적으로 사용되고, 동적 볼륨, 4D 시공간, 3D 비디오, 및 다중시점 비디오는 호환적으로 사용된다.

[0023] 이하, 도 1 및 도 2에 예시된 구성요소들을 설명하기 전에, 본 개시에 사용되는 딥러닝 관련 요소 기술들을 설명한다.

[0024] I-1. MLP(Multilayer Perceptron)

[0025] 다층신경망(Multilayer Perceptron, MLP)은 도 3의 예시와 같이, 다수의 뉴런들과 뉴런들을 연결하는 에지(edge)들로 구성이 된다. MLP는 입력 레이어(input layer)와 출력 레이어(output layer) 외에, 하나 이상의 은닉 레이어(hidden layer)를 포함할 수 있다. 하나의 은닉 레이어는 하나 이상의 은닉 유닛(hidden unit, 또는 은닉 노드(hidden node))를 포함한다. 또한, 각 에지에는 상이한 가중치(weight)가 할당될 수 있다. 또한, 하나의 계층에서 다음 계층으로 출력 값들이 전파되는 과정에서, 활성화 함수(activation function)가 사용될 수 있다. 대표적인 활성화 함수로서, 시그모이드(sigmoid) 함수, 탄젠트 쌍곡선(tangent hyperbolic) 함수, 또는 Relu(Rectified linear unit) 함수가 사용될 수 있다.

[0026] 도 3의 예시와 같이, 전방향 네트워크 구조를 갖는 2-레이어 MLP는 입력 벡터 $\mathbf{x}$ 를 중간층의 은닉 유닛을 통과시켜 출력 벡터 $\mathbf{y}$ 를 제공한다. 전술한 전방향 네트워크 구조에서 출력 레이어에 위치하는 임의의 노드값은 수학식 1과 같이 표현될 수 있다.

수학식 1

$$y_k(\mathbf{x}, \mathbf{W}) = \sigma\left(\sum_{j=0}^M w_{kj}^{(2)} \cdot h\left(\sum_{i=0}^D w_{ji}^{(1)} \cdot x_i\right)\right)$$

[0027]

[0028] 수학식 1에서, $\mathbf{W}^1$ 과 $\mathbf{W}^2$ 는 가중치 매트릭스로서 각각 입력 벡터와 은닉 유닛들을 연결하는 에지들과 은닉 유닛들과 출력 벡터를 연결하는 에지들에 대응한다. $w_{ji}^{(1)}$ 및 $w_{kj}^{(2)}$ 는 각각 $\mathbf{W}^1$ 과 $\mathbf{W}^2$ 의 구성요소를 나타낸다. y_k 는 출력 벡터 $\mathbf{y}$ 의 구성요소를 나타낸다. x_0 및 z_0 는 각각 입력 레이어와 은닉 레이어의 바이어스 유닛을 나타낸다. 따라서, D는 입력 벡터 $\mathbf{x}$ 의 차원이고, M은 은닉 유닛들의 개수를 나타내며, K는 출력 벡터 $\mathbf{y}$ 의 차원이다. 또한, h와 σ 는 각각 은닉 레이어 및 출력 레이어에 적용되는 활성화 함수이다. 한편, 은닉 레이어와 출력 레이어의 출력

에 반드시 활성화 함수가 인가되는 것은 아니다. 활성화 함수가 사용되지 않는 경우, 수식식 1은 W^1 과 W^2 간의 가중치 매트릭스 곱에 해당하는 연산을 포함한다.

[0029] 트레이닝용 데이터 및 레이블들에 기초하여 가중치들을 산정하는 과정을 트레이닝(training)이라 한다. 트레이닝은 일반적으로 역전파(back propagation) 알고리즘에 의한 SGD(stochastic gradient descent) 알고리즘을 이용하여 수행될 수 있다. 트레이닝에 따라 산정된 가중치들을 이용하여 전방향(feed-forward)으로 출력을 계산하는 과정을 추론 또는 검증(test)이라 한다.

[0030] 이하, 다층신경망, MLP, 또는 전방향 네트워크는 호환적으로 사용될 수 있다.

[0031] I-2 CNN(Convolutional Neural Network)

[0032] CNN은 복수의 컨볼루션 레이어(convolution layer)와 풀링 레이어(pooling layer)로 구성된 신경망을 지칭하고, 영상 처리에 가장 적합한 것으로 알려진 딥러닝 기술이다. 컨볼루션 레이어는 다수의 커널(kernel) 또는 필터(filter)를 이용하여 특징맵(feature map, 또는 '특징'과 호환적으로 이용)을 추출한다. 이때 필터를 구성하는 커널 계수(kernel coefficient)가 학습 과정에서 결정되는 파라미터이다.

[0033] CNN의 컨볼루션 레이어 중에서 입력에 가까운 전단 레이어는 선, 점, 또는 면과 같은 단순하고 낮은(lower) 수준의 영상 특징에 반응하는 특징맵을 추출하고, 출력에 가까운 후단 계층은 텍스처(texture), 사물 일부(object parts) 등 높은(higher) 수준에 반응하는 특징맵을 추출한다.

[0034] 도 4는 본 개시의 일 실시예에 따른 컨볼루션 레이어의 연산을 나타내는 예시도이다.

[0035] 컨볼루션 레이어는 컨볼루션 연산을 이용하여 입력 영상으로부터 특징맵을 생성한다. 도 4의 예시에는, 커널 크기가 3×3 인 커널(또는 필터)이 표현되어 있다. 커널 크기는 커널 사이즈(kernel size) 또는 필터 사이즈(filter size)로도 지칭된다. 커널은 가중치(weight)로도 불리우는 커널 파라미터(kernel parameter 또는 filter parameter)를 갖는다. 도 4에 예시된 커널은 총 9 개의 커널 파라미터를 갖는다. 커널 파라미터는 초기에 임의의 값으로 설정되고, 트레이닝(learning)을 기반으로 그 값이 업데이트될 수 있다.

[0036] 컨볼루션 레이어는 입력 영상에서 커널 사이즈만큼의 블록을 이용하여 컨볼루션 연산을 수행한다. 이때, 입력 영상 내의 커널 사이즈만큼의 블록을 윈도우(window)로 지칭한다.

[0037] 입력 영상에 대한 필터링을 래스터 스캔 순서(raster-scan order)로 수행할 때, 윈도우의 이동 크기를 스트라이드(stride)라고 한다. 도 4의 예시에서, 스트라이드는 1이다. 만약 스트라이드를 2로 설정하면, 윈도우를 2 샘플씩 띄워서 컨볼루션 연산을 수행하고, 결과적으로 특징맵의 가로와 세로의 크기는 입력 영상의 가로와 세로의 크기의 반이 된다.

[0038] 전술한 바와 같이, 하나의 컨볼루션 레이어는 다수의 필터를 포함할 수 있다. 필터의 개수(number of filter/filters) 또는 커널의 개수(number of kernel/kernels)를 채널(channel)이라고 한다. 즉, 채널의 개수는 필터의 개수와 동일하다. 또한 필터의 개수는 특징맵의 차원(dimension)의 크기를 결정한다.

[0039] 패딩(padding)은 컨볼루션 연산을 수행하기 전, 입력 데이터 주변을 특정값으로 채워서 입력 데이터를 확장하는 방법을 나타낸다. 패딩은 주로 출력 데이터의 공간적(spatial) 크기를 조절하기 위해 사용된다. 패딩 시 이용되는 값은 하이퍼파라미터에 의해 결정될 수 있으나, 주로 제로패딩(zero-padding)이 이용된다. 패딩을 사용하지 않을 경우, 컨볼루션 레이어를 거칠 때마다 출력 데이터의 공간적 크기가 감소하여, 경계의 정보들이 사라지는 문제가 발생할 수 있다. 따라서, 이러한 문제를 방지하기 위해, 패딩이 사용된다. 즉, 컨볼루션 레이어의 출력 데이터와 입력 데이터의 공간적 크기를 동일하게 맞추기 위해 패딩이 사용될 수 있다.

[0040] 디컨볼루션 레이어는 컨볼루션 레이어와 반대되는 연산을 수행한다. 디컨볼루션 레이어는 입력인 특징맵으로부터 원하는 데이터 영상을 출력으로 생성한다.

[0041] 풀링 레이어는 컨볼루션 레이어에 의해 생성된 특징맵을 서브샘플링(sub sampling)하는 과정인 풀링을 수행한다. 풀링 레이어는 2×2 크기의 윈도우를 이용하여 출력 결과가 입력의 가로와 세로 각각의 절반이 되도록 샘플을 선택한다. 즉, 풀링 레이어는 2×2 영역을 샘플 하나로 집약하여 입력 영상 또는 입력 특징맵의 크기를 축소하기 위해 이용된다.

[0042] 풀링 레이어의 반대 개념은 언풀링(unpooling) 레이어로 정의한다. 언풀링 레이어는 풀링 레이어와 반대로 차원을 확대하는 역할을 하고, 주로 디컨볼루션 레이어 이후에 사용된다.

- [0043] 콘볼루션 인코더(convolutional encoder)-디코더(decoder) 구조는 콘볼루션 레이어들과 디콘볼루션 레이어들의 짝(pair)으로 구성되는 네트워크 구조이다. 콘볼루션 인코더는 콘볼루션 레이어와 풀링 레이어로 구성되어, 입력 영상으로부터 특징맵(또는 특징 벡터)을 출력한다. 콘볼루션 인코더의 최종 출력 벡터를 잠재 벡터(latent vector)로도 지칭한다. 콘볼루션 디코더(convolutional decoder)는 디콘볼루션 레이어와 업샘플링 레이어로 구성되어, 특징맵 또는 잠재 벡터로부터 출력 영상을 생성한다.
- [0044] 콘볼루션 인코더-디코더의 입력과 출력은 애플리케이션과 네트워크의 목적에 따라 다양하게 설정될 수 있다. 예컨대, 입력과 출력은 옵티컬 플로우 맵(optical flow map), 돌출 맵(saliency map), 영상 프레임(image frame) 등일 수 있다.
- [0045] 도 5는 SISR 네트워크를 나타내는 예시도이다.
- [0046] CNN이 적용되는 일 예로서 SISR(Single Image Super Resolution)이 있다. SISR 네트워크는 저해상도 입력 영상으로부터 고해상도 이미지를 출력으로 생성한다. SISR 네트워크는 도 5에 예시된 바와 같이, 다수의 콘볼루션 레이어들을 포함할 수 있다. 각 콘볼루션 레이어는 ReLU(Rectified Linear Unit)과 같은 활성화 함수(activation function)를 포함한다. SISR 네트워크의 파라미터들은 생성되는 SR(Super Resolution) 영상이 GT(Ground Truth)에 근접하도록 트레이닝될 수 있다.
- [0047] CNN을 이용한 SR 방식들은 깊이를 증가시켜(예컨대, 콘볼루션 레이어의 개수를 증가시키는 것을 의미함) SR 성능을 향상시킬 수 있다. 깊이의 증가에 따라 발생할 수 있는 학습에서의 과적합(overfitting) 문제를 극복하기 위해, 스킵 연결(skip connection) 및 잔차 학습(residual learning)을 수행할 수 있는 잔차 블록(residual block)이 SISR 네트워크에 이용될 수 있다. 잔차 블록은, 도 6에 예시된 바와 같이, 입력 특징 x_i 에 콘볼루션 연산을 적용하는 경로 외에 스킵 경로를 포함한다. 또한, 잔차 블록은 출력 x_{i+1} 을 생성 시, 학습 효율에 기초하여 콘볼루션 연산을 적용하는 경로 또는 스킵 경로를 선택할 수도 있다. 도 6의 예시에서, 잔차 블록은 BN(Batch Normalization) 레이어를 포함한다.
- [0048] 일 예로서, EDSR(Enhanced Deep residual networks for SISR)은 잔차 블록들을 연속적으로 연결하여 깊이를 증가시킴으로써, 네트워크의 성능을 증가시킨다. 다른 예로서, VDSR(Accurate Image Super-Resolution Using Very Deep Convolutional Networks)은 VGG(Visual Geometry Group) 네트워크 기반의 CNN 모델로서, 잔차 프레임 최종 출력에 더해 주는 방식인 잔차신호 기반 학습(residual learning)을 사용한다. VDSR은 네트워크의 맨 마지막에 잔차 신호를 추가하여, 입력 신호에 잔차 신호를 더한다.
- [0049] **I-3. 암시적 신경망 표현(implicit neural representation) 모델**
- [0050] 최근 이미지를 비롯한 각종 데이터의 표현을 신경망 구조로 나타내는 암시적 신경망 표현(Implicit Neural Representation, INR) 모델에 관한 연구가 활발하다. 기존의 비디오 표현 방식은 픽셀 위치별 RGB 픽셀값을 갖는 명시적(explicit) 표현 방식을 사용한다. 이러한 명시적 표현 방식을 대체하기 위해 영상의 픽셀 위치인 (x, y) 좌표에서 (r, g, b) 값을 생성하는 함수를 신경망으로 표현하는 암시적 신경망 표현 모델이 도입되고 있다. 명시적 표현 방식과 비교하여 암시적 신경망 표현 모델은 이미지의 해상도와 관계 없이 사용될 수 있다.
- [0051] 암시적 신경망 표현 모델을 이용하는 3차원 비디오 코딩의 일 예로서, NeRF(Neural Radiance Fields) 기술이 존재한다.
- [0052] 다시점 비디오의 다수 시점들(view point)에서 촬영한 이미지들을 학습함으로써, NeRF는 임의의 위치와 방향에서 바라본 대상의 모습을 합성한다. 카메라가 바라보는 방향의 레이 정보(ray information)를 샘플링한 3D 위치 (x, y, z) 및 시점 방향(viewing direction) (θ, ϕ) 로 구성된 5차원 좌표를 입력받은 후, NeRF는 입력 좌표에 대응하는 RGB 값 (r, g, b) 및 볼륨 밀도(volume density) σ 를 출력한다. 여기서, 3D 위치 (x, y, z) 는 월드 좌표계의 위치이고, 시점 방향을 구성하는 θ, ϕ 는 각각 방위각과 고도를 나타낼 수 있다.
- [0053] 기존 암시적 신경망 표현 방식과 대비하여 고해상도의 복잡한 장면을 더 잘 표현하기 위해 NeRF는 다음과 같은 개선점을 갖는다. NeRF는 위치 인코딩(positional encoding)을 이용하여 위치 정보를 임베딩한 후, 수학적 2와 같이 임베딩된 인덱스를 입력으로 사용한다.

수학식 2

$$\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \dots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p))$$

[0054]

[0055]

수학식 2의 예시와 같이, 각 공간 좌표는 2L 차원의 벡터 $\gamma(p)$ 로 매핑될 수 있다. 3D의 입력 좌표가 그대로 입력되는 것을 대신하여, sin, cos함수를 이용하여 각 공간 좌표는 2L 차원의 벡터 $\gamma(p)$ 로 매핑될 수 있다. 임베딩된 공간 인덱스를 NeRF 모델의 트레이닝에 이용함으로써, NeRF 모델은 고주파 변이를 포함하는 데이터를 더 잘 예측할 수 있다.

[0056]

NeRF는 암시적 신경망 표현 모델로서, 다수의 전연결 레이어(Fully-connected Layer, FC)들을 포함하는 다층신경망(Multi-layer Perceptron, MLP) 네트워크를 사용한다. NeRF는 고차원의 입력 좌표에 대응하는 RGB 값 (r, g, b) 및 볼륨 밀도 σ 를 출력함으로써 임의의 장면(scene)을 표현한다. 이후, RGB 값 (r, g, b) 및 볼륨 밀도에 기초하는 볼륨 렌더링 기법에 따라, 장면을 지나는 임의의 레이 r의 색상 $C(r)$ 이 렌더링될 수 있다. $C(r)$ 은 미분가능(differentiable)하므로, GT(Ground Truth)와 함께, NeRF의 트레이닝에 이용된다. 이때, 각 5차원 좌표에 대해 GT로서, 시점 방향 (θ, ϕ)을 따라 월드 좌표 (x,y,z)를 촬영한 영상이 사용된다.

[0057]

한편, 렌더링 효율을 향상시키기 위해 NeRF는 단일 MLP를 사용하는 것을 대신하여, 2 개의 망, 즉 coarse 네트워크 및 fine 네트워크를 포함한다. NeRF는 계층적인 샘플링(stratified sampling)을 이용하여 N_c 개의 위치를 샘플링한 후, N_c 개의 위치에서 coarse 네트워크의 출력을 산정한다. coarse 네트워크의 출력으로부터 fine 네트워크에 적용하기 위한 더 많은 샘플들을 생성하기 위해, N_c 개의 coarse 네트워크 출력에 기초하여 복합 색상 $C_c(r)$ 은 수학식 3과 같이 표현된다.

수학식 3

$$C_c(r) = \sum_{i=0}^{N_c} w_i c_i$$

[0058]

[0059]

수학식 3에서, c_i 는 N_c 개의 위치에서 산정된 coarse 네트워크의 색상이다. w_i 는 대응하는 가중치로서 볼륨 밀도 및 샘플링된 위치들 간의 거리에 의존하는 값이다. 역변환 샘플링(inverse transform sampling)을 이용하여 전술한 복합 색상으로부터 높은 가중치를 갖는 N_f 개의 위치를 샘플링한 후, NeRF는 N_c+N_f 개의 위치에서 fine 네트워크의 출력을 산정하고, 산정된 fine 네트워크의 출력에 기초하는 색상 $C_f(r)$ 를 렌더링할 수 있다. 이때, coarse 네트워크 및 fine 네트워크의 출력에 따른 색상과 GT 간 MSE(Mean Square Error) 손실함수를 이용하여, NeRF가 트레이닝될 수 있다.

[0060]

전술한 바와 같이, 이미지 내의 모든 픽셀들을 월드 좌표계로 재구성하기 위한 학습 과정을 이용하여 NeRF는 임의의 시점에서 렌더링된 결과를 출력한다. 따라서 이미지 좌표계 (x, y)의 한 점으로 투영되는 모든 점들을 월드 좌표계로 변환하는 과정이 필요하므로, NeRF의 입력은 월드 좌표계에서의 인덱스와 더불어 월드 좌표계로의 투영에 필요한 카메라 파라미터들을 포함한다.

[0061]

카메라 파라미터들을 추출하는 과정은, 도 7의 예시와 같이 이미지 내 픽셀들을 초점거리 $Z=-1$ 를 갖는 좌표계 [X, Y, Z]를 거쳐서 월드 좌표계 [U, V, W]로 변환한다. 이때, 카메라에 대한 사전 정보가 필수적이며, 사전 정보로서 카메라 파라미터들은 내부(internal) 파라미터들과 외부(external) 파라미터들을 포함한다. 카메라 내부 파라미터들은 초점거리(focal length), 이미지 평면상의 주점(principal point), 및 비대칭 계수(skew coefficient)를 포함한다, 카메라 내부 파라미터들에 따라 대상은 카메라 좌표계 [X, Y, Z]로 투영될 수 있다. 카메라 좌표계로 투영된 정보를 월드 좌표계 [U, V, W]로 변환하기 위해, 월드 좌표계 내에서 카메라가 어디에 위치하는지에 대한 평행이동(translation) 정보, 및 어디를 바라보고 있는지에 대한 회전(rotation) 정보가 필요하다. 카메라 외부 파라미터들은 전술한 평행이동(translation) 정보 및 회전 정보를 포함한다. 영상을 촬영

할 때 회전 행렬(rotation matrix) R , 및 평행이동 벡터(translation vector) t 가 결합된 행렬 $[R|t]$ 가 월드 좌표계에서 이미지 좌표계로의 투영계산에 사용된다. 따라서, 카메라 외부 파라미터들은 결합 행렬 $[R|t]$ 의 역행렬을 산정하여 추출될 수 있다.

[0062] 한편, NeRF의 입력인 5차원 좌표는 추출된 카메라 파라미터들로부터 유도될 수 있다. 카메라 파라미터들은 전술한 바와 같이, 촬영된 2D 영상 및 2D 영상에 대응하는 월드 좌표계 내의 위치들에 기초하여 유도될 수 있다. GT로 사용하기 위한 영상들을 촬영 시, 카메라 내부 파라미터들은 고정될 수 있다. 카메라 내부 파라미터들이 고정된 채로 사용되는 경우, 5차원 좌표는 카메라 외부 파라미터들로부터 유도될 수 있다.

[0063] 또는, 영상 촬영 시 사용되었던 카메라 파라미터들을 사용하는 방법 외에 colmap과 같이 다시점 이미지들의 특징점을 추출하여 월드 좌표계 내에서 카메라의 위치와 방향에 대한 정보를 재구성하는 방법도 이용될 수 있다.

[0064] 다른 예로서, 그리드 기반 정적 NeRF는 암시적 신경망 표현 외에 명시적 표현을 함께 사용하는 하이브리드 NeRF이다. 그리드 기반 정적 방법은, 3D 볼륨의 신속한 재구성을 위해 복셀(voxel) 그리드를 사용한다. 하지만, 복셀 기반 방법은 복셀 당 특징(예를 들어, 색상 및 볼륨 밀도)을 저장해야 하므로, 대용량의 메모리를 필요로 한다. 메모리 효율성을 높이기 위해, 특징 그리드 또는 특징 텐서(tensor)의 구성 요소를 분해(decomposition)하는 방법이 사용된다. 일반적으로 널리 사용되는 분해 방법은 벡터행렬(Vector-Matrix, VM) 분해 또는 CP(Canonical Polyadic) 분해인데, 두 분해 방법은 행렬 특이값 분해(Singular Value Decomposition, SVD)의 일반화로 볼 수 있다.

[0065] 그리드 기반 정적 방법은, 전술한 NeRF와 유사하게, 명시적으로 색상을 표현하는 텐서와 볼륨 밀도를 표현하는 텐서를 트레이닝하고, 대상 시점을 구성하는 색상들과 볼륨 밀도들을 적분하여 대상 픽셀의 색상을 산정한다. 이때, 볼륨 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더(shader)에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH(Spherical Harmonics) 함수로 구현될 수 있다.

[0066] 그리드 기반 정적 NeRF는, 정적 NeRF와 비교하여 트레이닝 시간이 감소한다는 장점을 가지나, 여전히 동적 볼륨에 대해 대처하지 못한다는 단점을 갖는다.

[0067] 다른 예로서, 동적 NeRF(D-nerf)는 전술한 NeRF에 시간 차원을 추가한다. 기존 NeRF는, 시간을 고려하지 않은 정적 영상에서 공간정보를 학습하므로, 움직이는 영상에 대해 큰 오차를 발생시킨다. 시간을 고려하기 위해, 동적 NeRF는 (x, y, z, t) 를 입력으로 사용하고, 4D 시공간을 암시적으로 표현하는 신경망을 트레이닝한다. 암시적 신경망은 두 개의 모듈을 포함하는데, 첫 번째 모듈은 기준 시간(예를 들어, $t=0$)에서의 표준 공간과 시간 t 에서의 공간 간의 차이 $(\delta x, \delta y, \delta z)$ 를 산정한다. 두 번째 모듈은, 기존의 NeRF와 유사하게, $(x+\delta x, y+\delta y, z+\delta z)$ 및 시점 방향 (θ, ϕ) 의 5차원 정보로부터 RGB 값 (r, g, b) 및 볼륨 밀도를 생성한다. 이후, RGB 값 (r, g, b) 및 볼륨 밀도에 기초하는 볼륨 렌더링 기법에 따라, 대상 픽셀의 색상이 산정될 수 있다.

[0068] 동적 NeRF는 정적 NeRF와 유사하게 트레이닝에 많은 시간이 소요된다는 단점을 갖는다.

[0069] 다른 예로서, 그리드 기반 동적 NeRF는 암시적 신경망 표현 외에 명시적 표현을 함께 사용하는 하이브리드 동적 NeRF이다. 그리드 기반 동적 방법은, 시간을 고려한 볼륨, 즉, 4D 시공간의 신속한 재구성을 위해 복셀 그리드를 사용한다. 예컨대, 4D 시공간 그리드는 각 좌표 축 쌍(예를 들어, XY, ZT)을 포함하는 여섯 개의 특징 평면으로 분해된다. 여기서, 특징은 색상 및 볼륨 밀도를 명시적으로 표현한다. 각 특징 평면에 4D 시공간의 점을 투영하고, 6 개의 평면에 투영된 결과를 융합함으로써, 4D 시공간의 점에 대해 특징 벡터가 재구성될 수 있다. 그리드 기반 동적 방법은 색상 및 볼륨 밀도를 표현하는 6 개의 평면을 트레이닝하고, 대상 시점을 구성하는 색상들과 볼륨 밀도들을 적분하여 대상 픽셀의 색상을 산정한다. 이때, 볼륨 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH 함수로 구현될 수 있다.

[0070] 그리드 기반 동적 NeRF는 동적 NeRF와 비교하여 트레이닝 시간이 감소한다는 장점을 가지나, 4D 시공간을 구성하는 볼륨들 간의 연관성을 적절하게 활용하지 못한다는 단점을 갖는다.

[0071] 이하, 동적 NeRF와 비교하여 트레이닝 시간을 감소시키면서도, 그리드 기반 동적 NeRF와 비교하여 볼륨들 간의 연관성을 효율적으로 활용할 수 있는 동적 볼륨 코딩 방법을 기술한다.

[0072] 이하, 다중시점 비디오 코딩 모델과 다중시점 모델은 호환적으로 사용될 수 있다.

[0073] 이하, 특징맵과 특징은 호환적으로 사용될 수 있다.

[0074] 이하, 볼륨 밀도와 밀도는 호환적으로 사용될 수 있다.

[0075] 이하, 복셀은 3차원 그리드를 구성하는 요소를 나타내고, 픽셀은 2차원 평면을 구성하는 요소를 나타낸다.

[0076] II. 본 개시에 따른 실시예들

[0077] 도 1에 예시된 볼륨 부호화 장치는 다중시점을 갖는 비디오를 이용하여 다중시점 비디오 부호화부(110) 내 다중시점 비디오 코딩 모델을 트레이닝함으로써, 다중시점 모델로 하여금 다양한 시점들의 동적 3D 비디오를 생성할 수 있도록 한다. 이때, 다중시점 모델의 입력으로는 4D 시공간 좌표인 (x, y, z, t) 및 시점 방향 (θ, ϕ) 를 나타내는 정보가 사용되고, GT(Ground Truth)로는 대상 픽셀에 대응하는 (r, g, b) 값이 사용된다. 여기서, t 는 동적 3D 비디오 내 볼륨에 대응하는 시간을 나타내며, (x, y, z) 는 월드 좌표계에서의 3D 좌표를 나타낸다. 전술한 바와 같이, 카메라 파라미터에 기초하여 $(x, y, z), (\theta, \phi)$ 와 이미지 평면에서의 대상 픽셀 간의 변환이 정의될 수 있다.

[0078] 본 개시에 따른 다중시점 모델은 4D 시공간에 대한 명시적 표현과 암시적 신경망 표현을 모두 포함할 수 있다. 트레이닝부는 다중시점 모델의 출력과 GT 간 차이를 이용하여 명시적 표현을 나타내는 특징들 및 암시적 신경망 표현을 나타내는 파라미터들을 업데이트함으로써, 다중시점 모델로 하여금 다양한 시점 및 다양한 시간에서 영상을 생성하는 방법을 학습하도록 한다. 학습이 완료된 다중시점 모델은 특징들 및 파라미터들에 기초하여, 동적 3D 비디오에 내재된 4D 시공간 정보를 표현할 수 있다.

[0079] 특징 압축기(120)는 트레이닝된 다중시점 모델의 특징들 및 파라미터들을 압축하여 비트스트림을 생성한다. 볼륨 부호화 장치는 생성된 비트스트림을 볼륨 복호화 장치로 전송한다. 볼륨 부호화 장치는 부가 정보로서, 다중시점 모델의 구조정보, 양자화 파라미터, 카메라 파라미터들, 및 움직임 보상에 대한 정보를 포함할 수 있다.

[0080] 이하, 기존 비디오의 압축 방법에서 사용하는 예측 부호화를 활용하는 다중시점 모델을 기술한다.

[0081] 도 8은 본 개시의 일 실시예에 따른, GOV(Group of Volume)를 나타내는 예시도이다.

[0082] 다중시점 모델은 GOV(Group of Volume)를 사용하여 4D 시공간 정보를 표현하고 부호화한다. GOV는 I(Intra) 볼륨과 B(Bidirectional) 볼륨을 포함한다. I 볼륨과 B 볼륨을 사용함으로써, 6차원의 입력 중에서, 시간 t 가 처리될 수 있다. 도 8과 같이, I 볼륨은 참조 없이 해당 시간에 대한 정적 3D 공간 정보를 나타낼 수 있다. B 볼륨은 인접한 두 I 볼륨을 참조하여 생성될 수 있다.

[0083] 부가 정보로서 움직임 보상에 관련된 정보는 전술한 GOV의 구조와 관련된 정보일 수 있다. GOV의 구조와 관련된 정보는, GOV의 크기, I 볼륨의 위치(즉, I 볼륨의 시간 정보) 등을 포함할 수 있다.

[0084] B 볼륨은 인접한 두 I 볼륨을 참조하므로, 볼륨 부호화 장치는 GOV 내에서 I 볼륨들을 먼저 부호화된 후, B 볼륨들을 부호화할 수 있다.

[0085] 부호화 측면에서 다중시점 모델은, 도 1에 예시된 바와 같이, I 볼륨 분해기 및 B 볼륨 생성기를 포함한다.

[0086] I 볼륨은, B 볼륨과 대비하여 큰 정보량을 갖는다. I 볼륨의 개수를 조절하거나, 그리드의 해상도/채널 수를 늘리거나 줄임으로써, 특징들의 크기, 및 특징들의 크기에 따른 비트스트림의 크기가 결정될 수 있다. 또한, 재구성된 볼륨의 복원력에 관한 트레이드오프(trade-off)가 조정될 수 있다. 예를 들어, I 볼륨의 개수가 증가하면, 비트스트림의 크기가 증대되고 복원력이 증가한다.

[0087] 일 예로서, 트레이닝부는 다양한 GOV 구조에 대해 I 볼륨과 B 볼륨을 학습하고, 최적의 성능을 제공하는 GOV 구조를 확정할 수 있다. 볼륨 부호화 장치는 움직임 보상에 관련된 부가 정보로서, GOV의 구조와 관련된 정보를 볼륨 복호화 장치에게 전달할 수 있다.

[0088] 도 9a 및 도 9b는 본 개시의 일 실시예에 따른, 텐서 분해를 나타내는 예시도이다.

[0089] I 볼륨은 다른 볼륨을 참조하지 않으므로, 하나의 완전한 3D 공간의 복셀 그리드를 구성할 수 있는 텐서들을 포함한다. 메모리 효율성을 높이기 위해, 도 9a 및 도 9b와 같이, I 볼륨 분해기는 기존의 텐서 분해 방법을 기초하여 I 볼륨의 그리드를 축소하고, 축소된 그리드를 이용하여 I 볼륨을 명시적으로 표현할 수 있다.

[0090] 텐서 분해의 일 실시예로서, 도 9a의 예시를 이용하여, 벡터행렬(Vector-Matrix, VM) 분해에 기초하여 I 볼륨의

특징맵을 축소하는 방법을 기술한다.

[0091] I 볼륨 분해기는, 도 9a에 예시된 바와 같이, VM 분해를 이용하여, I 볼륨의 특징맵을 벡터와 행렬의 선형 조합으로 분해할 수 있다. 여기서, 벡터는 랭크 1 텐서에 해당하고, 행렬은 랭크 2 텐서에 해당한다. I 볼륨의 특징맵, 즉 그리드는 $n_1 \times n_2 \times n_3$ 크기를 갖는 것으로 가정한다. 즉, I 볼륨의 특징맵은 수학식 4에 나타낸 바와 같이 $P_1+P_2+P_3$ (여기서, P_1, P_2, P_3 는 자연수) 개의 벡터/행렬의 조합에 기초하여 근사될 수 있다.

수학식 4

$$F \approx \sum_{p=1}^{P_1} v_{x,p} \circ M_{yz,p} + \sum_{p=1}^{P_2} v_{y,p} \circ M_{xz,p} + \sum_{p=1}^{P_3} v_{z,p} \circ M_{xy,p}$$

[0092]

[0093] 수학식 4에서, F는 I 볼륨의 특징맵을 나타내고, 연산자 $\circ$ 는 외적을 나타낸다. $v_{x,i}(1 \leq i \leq P_1)$, $v_{y,i}(1 \leq i \leq P_2)$, $v_{z,i}(1 \leq i \leq P_3)$ 는 각각 n_1 , n_2 , n_3 차원을 갖는 벡터를 나타낸다. $M_{yz,i}(1 \leq i \leq P_1)$, $M_{xz,i}(1 \leq i \leq P_2)$, $M_{xy,i}(1 \leq i \leq P_3)$ 는 각각 $n_2 \times n_3$, $n_1 \times n_3$, $n_1 \times n_2$ 차원을 갖는 행렬을 나타낸다.

[0094] 볼륨 부호화 장치는, $M_{yz,i}(1 \leq i \leq P_1)$, $M_{xz,i}(1 \leq i \leq P_2)$, $M_{xy,i}(1 \leq i \leq P_3)$ 의 행렬들을 부호화할 수 있다. 또한, 볼륨 부호화 장치는, 수학식 5에 나타낸 바와 같이, 크기가 같은 랭크 1 텐서들끼리 결합(concatenation)하여, 예컨대, 랭크 2 텐서들 A, B 및 C를 생성한 후, 랭크 2 텐서들을 부호화한다.

수학식 5

$$A = [v_{x,1}, v_{x,2}, \dots, v_{x,P_1}]$$

$$B = [v_{y,1}, v_{y,2}, \dots, v_{y,P_2}]$$

$$C = [v_{z,1}, v_{z,2}, \dots, v_{z,P_3}]$$

[0095]

[0096] 수학식 5에서, A, B, C는 각각 $n_1 \times P_1$, $n_2 \times P_2$, $n_3 \times P_3$ 크기를 갖는 랭크 2 텐서를 나타낼 수 있다.

[0097] VM 분해에 따라, 원본 특징맵의 데이터 크기 $n_1 \times n_2 \times n_3$ 는 ' $(n_1 + n_2 \times n_3)P_1 + (n_2 + n_1 \times n_3)P_2 + (n_3 + n_1 \times n_2)P_3$ '로 감축된다.

[0098] 일 예로서, I 볼륨의 색상 및 밀도에 VM 분해가 적용된 경우, $P_1+P_2+P_3$ 개의 행렬, 수학식 5에 따른 A, B, C는 I 볼륨의 채널을 구성할 있다.

[0099] 텐서 분해의 다른 실시예로서, 도 9b의 예시를 이용하여, CP(Canonical Polyadic) 분해에 기초하여 I 볼륨의 특징맵을 축소하는 방법을 기술한다.

[0100] I 볼륨 분해기는, 도 9b에 예시된 바와 같이, CP 분해를 이용하여, I 볼륨의 특징맵을 랭크 1 텐서들, 즉 벡터들의 선형 조합으로 분해할 수 있다. 여기서, I 볼륨의 특징맵은 $n_1 \times n_2 \times n_3$ 크기를 갖는 것으로 가정한다. 즉, I 볼륨의 특징맵은 수학식 6에 나타낸 바와 같이 P(여기서, P는 자연수) 개의 랭크 1 텐서들에 기초하여 근사될 수 있다.

수학식 6

$$F \approx \sum_{p=1}^P v_{x,p} \circ v_{y,p} \circ v_{z,p}$$

[0101]

[0102] 수학식 6에서, F는 I 볼륨의 특징맵을 나타내고, 연산자 $\circ$ 는 외적을 나타낸다. $v_{x,i}$, $v_{y,i}$, $v_{z,i}$ ($1 \leq i \leq P$)는 각각 n_1 , n_2 , n_3 크기를 갖는 랭크 1 텐서를 나타낸다.

[0103] 볼륨 부호화 장치는, 수학식 7에 나타난 바와 같이, 크기가 같은 랭크 1 텐서들끼리 결합하여, 예컨대, A, B 및 C를 생성한 후, 이들을 부호화한다.

수학식 7

$$A = [v_{x,1}, v_{x,2}, \dots, v_{x,P}]$$

$$B = [v_{y,1}, v_{y,2}, \dots, v_{y,P}]$$

$$C = [v_{z,1}, v_{z,2}, \dots, v_{z,P}]$$

[0104]

[0105] 수학식 7에서, A, B, C는 각각 $n_1 \times P$, $n_2 \times P$, $n_3 \times P$ 크기를 갖는 랭크 2 텐서를 나타낼 수 있다.

[0106] CP 분해에 따라, 원본 특징맵의 데이터 크기 $n_1 \times n_2 \times n_3$ 는 ' $(n_1 + n_2 + n_3)P$ '로 감축된다.

[0107] 일 예로서, I 볼륨의 색상 및 밀도에 CP 분해가 적용된 경우, 수학식 7에 따른 A, B, C는 I 볼륨의 채널을 구성할 있다.

[0108] I 볼륨은 GOV의 길이를 결정한다. I 볼륨은 동적 볼륨 스트림의 어디에도 위치할 수 있으며 키프레임 탐색을 위해 사용될 수 있다. 그리드 기반 정적 방법과 동일하게, 다중시점 모델은 명시적으로 색상을 표현하는 텐서 및 밀도를 표현하는 텐서를 트레이닝한다. 여기서, 텐서들은 전술한 바와 같이, 분해된 텐서들을 나타낸다. 다중시점 모델은, I 볼륨의 분해된 텐서들에 기초하여 대상 시점에 대응하는 복셀들에서 색상들과 밀도들을 산정하고, 산정된 색상들과 밀도들을 적분하여 대상 픽셀의 색상을 산정한다(즉, 대상 시점에 대응하는 비디오 또는 영상을 렌더링한다). 이때, 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더(shader)에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH(Spherical Harmonics) 함수로 구현될 수 있다.

[0109] 일 예로서, 전술한 GOV 내 각 I 볼륨에 대해 텐서 분해가 수행될 수 있다.

[0110] B 볼륨 생성기는 현재 시간에서 가장 인접한 I 볼륨 두 개를 참조하여 양방향 예측에 따라 B 볼륨을 생성한다. 즉, B 볼륨 생성기는 현재 B 볼륨보다 시간적으로 선행하는 I 볼륨과 시간적으로 후행하는 I 볼륨을 활용한다. B 볼륨을 많이 포함할수록 전체 동적 볼륨에 대한 용량은 감소하지만, I 볼륨과 대비하여 렌더링된 영상의 화질이 저하될 수 있다. 학습 가능한 신경망에 기초하여 기존의 기분해된 I 볼륨의 정보를 사용하여 B 볼륨을 표현하는 텐서들(색상 및 볼륨 밀도에 대한 텐서들)을 생성될 수 있다. 여기서, 텐서들은 분해된 텐서들을 나타낸다. B 볼륨 생성기는 생성된 텐서들에 기초하여 B 볼륨을 재구성할 수 있다.

[0111] 재구성된 B 볼륨을 이용하여, 다중시점 모델은 대상 시점에 대응하는 복셀들에서 B 볼륨의 색상들과 밀도들을 산정하고, 산정된 색상들과 밀도들을 적분하여 대상 픽셀의 색상을 산정한다(즉, 대상 시점에 대응하는 비디오 또는 영상을 렌더링한다). 이때, B 볼륨의 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, B 볼륨의 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH 함수로

구현될 수 있다.

- [0112] B 볼륨은 하나 이상의 축에 대한 고정 텐서 및 생성 텐서를 사용한다. 여기서 생성 텐서는 고정 텐서에 직교(orthogonal)하는 축에 존재할 수 있다. 즉, 생성 텐서와 고정 텐서에 기초하여 시점에 맞는 볼륨이 생성될 수 있다. 생성 텐서는 대상 시점에 대해 생성된 텐서를 의미한다. 해당 텐서는 전술한 신경망, 즉, B 볼륨 텐서 생성 모델(이하, '텐서 생성 모델')에 의해 생성될 수 있다.
- [0113] 고정 텐서는 생성 텐서가 아닌 텐서를 의미한다. 고정 텐서로서, 모든 시점에 대해 공유되는 하나의 학습 가능한 텐서가 사용될 수 있다. 또는, 인접한 I 볼륨에서 B 볼륨의 생성 텐서에 직교하는 축에 대응하는 텐서가 고정 텐서로 사용될 수 있다. 따라서, 인접한 I 볼륨에서 B 볼륨의 고정 텐서와 동일한 축을 갖는 텐서가 고정 텐서로 사용될 수 있다.
- [0114] 텐서 생성 모델은 임의 시간의 공간 정보에 대한 생성 텐서를 추론하기 위해 사용된다. 텐서 생성 모델은 인접한 I 볼륨들의 동일한 축에서 동일한 모양 갖는 텐서 2 개를 입력으로 사용한다. 예컨대, 동일한 축을 갖는 벡터 2 개를 사용하거나 동일한 축을 갖는 평면 텐서 2 개가 사용된다. 텐서 생성 모델의 인코더와 디코더는 MLP, CNN 등과 같은 학습 가능한 신경망으로 구성될 수 있다. 색상과 밀도를 위한 각 텐서에 대해 동일한 파라미터의 신경망이 사용되거나, 색상과 밀도에 대한 상이한 신경망이 사용될 수 있다. 학습 가능한 신경망은 다음과 같은 모듈 중 하나 이상의 조합으로 이루어져 있다.
- [0115] 학습 가능한 신경망 기반 인코더/디코더는, MLP 레이어, 컨볼루션 레이어, 디포머블 컨볼루션 레이어 등으로 구성되는 신경망으로서, N 개의 레이어를 연속적으로 포함할 수 있다. MLP 레이어, 컨볼루션 레이어, 디포머블 컨볼루션 레이어 중 하나만 사용되거나, 2 개 이상의 조합을 사용하여 인코더 또는 디코더가 구성될 수 있다. 여기서, 디포머블 컨볼루션 레이어에서는, 컨볼루션 커널의 그리드에 2D오프셋을 더함으로써, 컨볼루션 커널이 다양한 형태로 변환된다.
- [0116] 텐서 생성 모델은, 인코더를 지난 특징에 시간적 정보를 채널 기준으로 누적한 후, 시간적 정보가 누적된 특징을 디코더에 전달한다. 시간 정보는, 시간에 대해 정규화된 비디오 시퀀스에서 해당 시간의 좌표 값을 나타낸다. 이때, 채널 기준으로 누적하기 위해, 시간적 정보는, 특징과 동일한 크기의 높이와 너비를 가진 행렬로 생성된다. 전술한 방법에 따라, 시간에 대한 연속적인 B 볼륨 텐서가 생성될 수 있다.
- [0117] 텐서 생성 모델은, 인코더를 지난 특징에 시점 정보를 채널 기준으로 누적한 후, 시점 정보가 누적된 특징을 디코더에 전달한다. 시점 정보는, 시점에 대해 정규화된 비디오 시퀀스에서 해당 시점의 좌표 값을 나타낸다. 이때, 채널 기준으로 누적하기 위해, 시점 정보는, 특징과 동일한 크기의 높이와 너비를 가진 행렬로 생성된다. 전술한 방법에 따라, 시점에 대한 연속적인 B 볼륨 텐서가 생성될 수 있다.
- [0118] 텐서 생성 모델은, 인코더를 지난 특징을 이용하여 입력 텐서에 대한 플로우를 생성할 수 있다. 텐서 생성 모델은, 생성한 플로우를 이용하여 입력 텐서 또는 (인코더의) 출력 특징을 와핑(warping)함으로써, 입력 텐서들 간의 움직임 정보를 파악할 수 있다.
- [0119] 도 10 및 도 11은 전술한 바와 같은 텐서 생성 모델의 나타낸다. 도 10에서, 텐서 생성 모델은 벡터들을 입력으로 사용하고, 도 11에서, 텐서 평면들을 입력으로 사용한다. 도 10 및 도 11에서, 시간적 좌표(temporal coordinate)는 시간 정보 또는 시점 정보를 나타내는 행렬이다. 도 11에서, 텐서 생성 모델은 입력 텐서에 대한 플로우를 생성한다.
- [0120] 일 예로서, 텐서 생성 모델로서, 전술한 GOV 내 B 볼륨들에 대해 공통적으로 하나의 모델이 이용될 수 있다. 다른 예로서, 텐서 생성 모델로서, 다수의 GOV들에 대해 공통적으로 하나의 모델이 이용될 수 있다.
- [0121] 볼륨 부호화 장치는 다양한 텐서 생성 모델의 구조를 트레이닝하고, 최적의 성능을 제공하는 구조를 볼륨 복호화 장치에게 전달할 수 있다.
- [0122] 도 12a 내지 도 12c는 본 개시의 일 실시예에 따른, B 볼륨의 생성을 나타내는 예시도이다.
- [0123] B 볼륨 생성기는, 전술한 바와 같이, I 볼륨의 하나 이상의 축에 대한 고정 텐서, 및 고정 텐서에 직교하는 축에 대한 생성 텐서를 사용한다. 텐서 생성 모델은, 전술한 바와 같이, 인접한 I 볼륨의 텐서를 1 개 이상 사용하여 대상 시간에 대한 생성 텐서를 생성한다. 이때, 인접한 I 볼륨의 텐서는 고정 텐서에 직교한다. 일 예로서, 도 12a에서는 고정된 XY, YZ, XZ 평면, 및 고정 평면에 직교하는 생성 Z, X, Y 벡터를 이용하여 B 볼륨이 재구성된다.

- [0124] 텐서 분해 방식으로 VM 분해 기법이 사용되는 경우, 평면 또는 벡터가 고정될 수 있다. CP 분해 기법이 사용되는 경우, 각각의 축에 해당하는 하나 이상의 벡터가 고정될 수 있다. 도 12a 내지 도 12c에서는, 생성 텐서에 기초하여 B 볼륨이 재구성된다. 도 12a 및 도 12b에서는 VM 분해 기법에 기초하여 B 볼륨이 재구성되고, 도 12c에서는 CP 분해 기법에 기초하여 B 볼륨이 재구성된다. 도 12a에서는, I 볼륨에서 획득된 고정 텐서 평면 (XY, YZ, XZ), 및 생성 텐서 벡터 간의 외적에 기초하여 대상 시간에 대한 B 볼륨이 재구성된다. 도 12b에서는, I 볼륨에서 획득된 고정된 텐서 벡터 (X, Y, Z), 및 생성 텐서 평면 간의 외적에 기초하여 대상 시간에 대한 B 볼륨이 재구성된다. 도 12c에서는, I 볼륨에서 획득된 고정된 텐서 벡터들, 및 생성 텐서 벡터 간의 외적에 기초하여 대상 시간에 대한 B 볼륨이 재구성된다.
- [0125] 전술한 바와 같이, 다중시점 모델은 재구성된 B 볼륨에 기초하여, 대상 시점에 대응하는 복셀들에서 B 볼륨의 색상들과 밀도들을 산정하고, 산정된 색상들과 밀도들을 적분하여 대상 픽셀의 색상을 산정한다(즉, 대상 시점에 대응하는 비디오 또는 영상을 렌더링한다).
- [0126] 도 2에 예시된 볼륨 복호화 장치는 특징 역압축기(210)를 이용하여 비트스트림으로부터 다중시점 모델의 특징들 및 파라미터들을 복호화한다. 여기서, 다중시점 모델의 특징들은 I 볼륨의 분해된 텐서 특징들을 나타낼 수 있다. 다중시점 모델의 파라미터들은, B 볼륨의 생성 텐서를 구성하는 텐서 생성 모델과 관련하여 텐서 생성 모델의 파라미터들을 나타낼 수 있다.
- [0127] 볼륨 복호화 장치는 부가 정보로서 다중시점 모델의 구조에 관한 정보, 양자화 파라미터, 카메라 파라미터 및 움직임 보상에 대한 정보를 복호화한다. 예컨대, 볼륨 복호화 장치는 움직임 보상에 관련된 부가 정보로서, GOV의 구조와 관련된 정보를 복호화할 수 있다.
- [0128] 다중시점 비디오 복호화부(220)는 복원된 다중시점 모델을 이용하여 다중시점에서 비디오를 생성한다. 볼륨 부호화 장치에서와 같이, I 볼륨과 B 볼륨을 포함하는 GOV를 활용함으로써, 다중시점 모델은 동적 3D 비디오의 시간을 처리할 수 있다. 볼륨 복호화 장치는 임의의 시점 j 를 획득한 후, 임의의 시점을 다중시점 모델에 입력하여 임의의 시점에 대응하는 2D 영상을 생성할 수 있다. I 볼륨과 B 볼륨에 대응하는 2D 영상을 연결함으로써, 볼륨 복호화 장치는 임의의 시점에 대응하는 2D 비디오를 생성할 수 있다.
- [0129] 이때, 임의의 시점은 입력 범위 내에 존재해야 한다. 예컨대, 임의의 시점이 입력 범위 내에 존재하지 않는 경우, 볼륨 복호화 장치는 시점에 대응하는 영상을 제공하지 않을 수 있다. 전술한 바와 같이, 카메라 파라미터에 기초하여, (x, y, z) , (θ, ϕ) 와 이미지 평면에서의 픽셀 간의 변환이 정의될 수 있다. (x, y, z) , (θ, ϕ) 는 t 를 제외한 다중시점 모델의 입력을 나타낸다.
- [0130] 복호화 측면에서 다중시점 모델은, 도 2에 예시된 바와 같이, I 볼륨 렌더러와 B 볼륨 생성기를 포함한다.
- [0131] I 볼륨 렌더러는 복원된 I 볼륨 텐서 특징들(예컨대, 색상 및 밀도를 표현하는 텐서들)을 활용한다. 여기서, 복원된 I 볼륨 텐서 특징들은 분해된 텐서 특징들일 수 있다. 복원된 I 볼륨 텐서 특징들에 기초하여, I 볼륨 렌더러는 임의의 시점에 대응하는 복셀들에서 색상들과 밀도들을 산정하고, 산정된 색상들과 밀도들을 적분하여 대상 픽셀의 색상을 산정한다(즉, 대상 시점에 대응하는 비디오 또는 영상을 렌더링한다). 이때, 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH 함수로 구현될 수 있다.
- [0132] 복호화 측면에서 B 볼륨 생성기의 동작은, 전술한 부호화 측면의 B 볼륨 생성기의 동작과 유사하다. B 볼륨 생성기는 현재 시간에서 가장 인접한 복원 I 볼륨 두 개를 참조하여 양방향 예측에 따라 B 볼륨을 생성한다. B 볼륨 생성기는 복원된 텐서 생성 모델을 이용하여 생성 텐서를 구성하고, 생성 텐서와 고정 텐서를 활용하여 B 볼륨을 표현하는 텐서들(색상 및 볼륨 밀도에 대한 텐서들)을 재구성한다. 여기서, 재구성된 B 볼륨 텐서 특징들은 분해된 텐서 특징들일 수 있다. 재구성된 B 볼륨 텐서 특징들에 기초하여, B 볼륨 생성기는 임의의 시점에 대응하는 복셀들에서 색상들과 밀도들을 산정하고, 산정된 색상들과 밀도들을 적분하여 대상 픽셀의 색상을 산정한다(즉, 대상 시점에 대응하는 비디오 또는 영상을 렌더링한다). 이때, B 볼륨의 밀도는 추가적인 단계 없이 사용될 수 있다. 광역 특성을 보강하기 위해, B 볼륨의 색상을 작은 크기의 MLP에 입력함으로써, 암시적 표현이 색상에 추가될 수 있다. 또한, MLP의 출력과 시점 방향을 셰이더에 입력함으로써, 색상에 시점 방향이 반영될 수 있다. 셰이더는 MLP 또는 SH 함수로 구현될 수 있다.
- [0133] 볼륨 부호화 장치와 볼륨 복호화 장치 간의 전송 정보는, 다중시점 모델과 관련하여, I 볼륨을 구성하는 텐서 특징들(색상 및 볼륨 밀도), B 볼륨 생성기 내 텐서 생성 모델의 파라미터들(즉, 가중치들), 렌더링에 활용되는

MLP의 가중치들을 포함할 수 있다. 볼륨 부호화 장치는 복호화 동작을 지원하기 위한 부가 정보를 볼륨 부호화 장치에게 전송할 수 있다.

[0134] 이하, I 볼륨을 구성하는 텐서 특징들을 압축하는 방법을 기술한다.

[0135] 일 예로서, 기존의 2D 비디오 코덱을 이용하여 텐서 특징들이 압축될 수 있다.

[0136] I 볼륨에 대한 인코딩 과정은 텐서 특징 패킹(tensor feature packing)이다. 인코딩 과정은, $1 \times C \times H \times W$ 차원인 평면 특징맵들을 $1 \times 1 \times pH \times pW$ 차원의 평면으로 공간적으로 패킹한다. C는 평면 특징맵의 채널의 개수를 나타내고, H와 W는 그리드의 해상도를 나타낸다. pH 및 pW는 평면 패킹 후의 높이와 너비를 나타낸다. 일 예로서, XY, YZ 및 ZX 평면의 색상 및 밀도 평면이 각각 [48, 12, 12], [16, 4, 4] 차원의 공간으로 생성된 경우, 인코딩 과정은 pW 및 pH를 다음과 설정할 수 있다. C가 48 인 경우, $pW = 8 \times W$, $pH = 6 \times H$ 로 설정될 수 있다. C가 16 인 경우, $pW = 4 \times W$, $pH = 4 \times H$ 로 설정될 수 있다. C가 12 인 경우, $pW = 4 \times W$, $pH = 3 \times H$ 로 설정될 수 있다. C가 4 인 경우, $pW = 2 \times W$, $pH = 2 \times H$ 로 설정될 수 있다.

[0137] 인코딩 과정은 패킹된 특징을 m 비트로 양자화(이 때, m은 임의의 양의 정수)함으로써, YUV 형식으로 변환한다. 이때, 최소값과 최대값이 디코딩 과정에게 신호될 수 있다.

[0138] 인코딩 과정은 평면 특징으로 변환된 YUV를, 기존의 영상 부호화 방법에 따라 All intra, Random access, 또는 Low delay 구성으로 부호화하여 비트스트림을 생성할 수 있다.

[0139] I 볼륨에 대한 디코딩 과정은, 비트스트림을 디코딩한다. 디코딩 과정은 YUV 형식을 텐서 유형으로 변환하고, 전달된 최소값과 최대값을 사용하여 변환된 텐서 특징을 32비트 부동소수점(float32)로 역양자화한다. 디코딩 과정은 공간적으로 패킹된 $1 \times 1 \times pH \times pW$ 차원 평면을 $1 \times C \times H \times W$ 차원의 평면 특징맵들로 언패킹한다.

[0140] 다른 예로서, 학습된 이미지 압축(Learning Image Compression, LIC) 모델을 이용하여 텐서 특징들이 압축될 수 있다.

[0141] 인코딩 과정은, 전술한 2D 비디오 코덱을 위한 텐서 특징 패킹과 동일한 방법에 따라, 특징 패킹을 진행한다. 색상 및 밀도 평면 특징맵들의 차원은 $R^{N \times C \times H \times W}$ 이다. 여기서, N은 3이고, XY, YZ, ZX 평면을 나타낸다. C는 평면 특징맵의 채널의 개수를 나타내고, H와 W는 그리드의 해상도를 나타낸다.

[0142] 인코딩 과정은, 학습된 이미지 압축 모델(예컨대, 압축 모델의 인코더)을 사용하여 평면 특징맵들을 채널별로 압축한다. 기존의 LIC 모델은 3 채널 RGB 이미지에 기초하여 트레이닝되므로, 평면 특징맵을 바로 압축할 수 없다. 채널의 평면 특징맵을 채널로 복사 또는 확장한 후, 기존의 LIC 모델이 사용될 수 있다. 또는, 평면 특징의 채널 1 개를 인코딩하는 1 채널 LIC 모델이 사용될 수 있다.

[0143] 1 채널 LIC 모델의 경우, 일 예로서, 1 채널 특징맵의 LIC 압축을 여러 차례 적용하는 적용하는 방식이 사용될 수 있다. $R^{N \times C \times H \times W}$ 차원의 색상 및 밀도 평면 특징맵을 1 채널씩 분리한 후 채널별로 LIC 모델이 적용될 수 있다.

[0144] 또는, 전술한 공간적 패킹에 따른 1 채널 특징맵에 LIC 압축을 한번 적용하는 방법이 사용될 수 있다. 전술한 방식과 동일하게 $R^{N \times C \times H \times W}$ 차원의 색상 및 밀도 평면 특징맵을 공간적으로 패킹하여 $R^{N \times 1 \times pH \times pW}$ 차원의 1 채널 특징맵으로 변환한 후, 변환된 특징맵에 LIC 압축이 적용된다.

[0145] 디코딩 과정은, 학습된 이미지 압축 모델(예컨대, 압축 모델의 디코더)을 사용하여, 압축된 특징맵으로부터 원래 사이즈의 $R^{N \times C \times H \times W}$ 차원의 I 볼륨 특징맵을 복원한다.

[0146] 다른 예로서, 신경망 압축 방식(Neural Network Compression, NNC)을 이용하여 텐서 특징들이 압축될 수 있다. I 볼륨의 텐서 특징의 성능을 최대한 유지하면서 텐서 특징의 크기, 연산량을 감소시키고, 다양한 딥러닝 프레임워크들 간에 상호 운용 가능한 형태로 텐서 특징을 압축하기 위해, NNC가 사용될 수 있다.

[0147] NNC의 인코딩 과정은 도 13에 예시된 부호화 파이프라인에 의해 수행될 수 있다. 부호화 파이프라인은 모델 프루너(model pruner, 1320), 모델 양자화부(model quantizer, 1330), 및 가중치 부호화부(weight encoder, 1340)의 전부 또는 일부를 포함할 수 있다.

[0148] 모델 프루너(1320)은 I 볼륨의 텐서 특징을 프루닝을 이용하여 단순화한다. 예컨대, 프루닝은 기설정된 임계치보다 작은 가중치(예컨대, 텐서 특징 값)를 영으로 설정한다. 모델 양자화부(1330)은 프루닝이 적용된 가중치들

(예컨대, 임계치 이상의 텐서 특징 값)을 양자화한다. 가중치 부호화부(1340)은 양자화된 가중치들에 엔트로피 부호화를 적용하여 가중치들의 비트스트림을 생성한다.

- [0149] 특징 압축 성능을 향상시키기 위해, 가중치 부호화(또는, 엔트로피 부호화)가 적용된 특징에, 후처리 기술(예컨대, 특징을 고정한 상태에서 바이어스(bias) 값을 미세조정하여 특징을 업데이트함)이 적용될 수 있다.
- [0150] 디코딩 과정은 도 14에 예시된 복호화 파이프라인에 의해 수행될 수 있다. 복호화 파이프라인은 가중치 복호화부(weight decoder, 1410), 모델 역양자화부(model dequantizer, 1420), 및 모델 복원부(model reconstructor, 1430)의 전부 또는 일부를 포함할 수 있다.
- [0151] 가중치 복호화부(1410)는 가중치들(예컨대, I 볼륨의 텐서 특징 값들)의 비트스트림에 엔트로피 부호화를 적용하여 양자화된 가중치들을 생성한다. 모델 역양자화부(1420)는 양자화된 가중치들을 역양자화하여 프루닝이 적용된 가중치들을 생성한다. 모델 복원부(1430)는 프루닝이 적용된 가중치들에 기초하여 I 볼륨을 복원한다.
- [0152] 한편, 전술한 신경망 압축 방식은, B 볼륨의 생성 텐서와 관련하여, 텐서 생성 모델의 파라미터들의 압축에 적용될 수 있다. 또는, 전술한 신경망 압축 방식은, 볼륨 부호화 장치 내 다른 신경망(예컨대, 렌더링 과정에서 사용되는 MLP들)의 파라미터들의 압축에도 적용될 수 있다.
- [0153] 특징을 압축하기 위한 최적의 네트워크 구조 및 압축 방식에 관한 부가 정보는 NNC 표준에서 정의한 비트스트림 형태/단위(예를 들어, NNR(Neural Network compression and Representation) unit)를 이용하여 전달될 수 있다.
- [0154] 일 예로서, 전술한 부가 정보를 전달하기 위하여 다음의 선택 요소들을 포함하는 동적 볼륨 IPS(INR Parameter Set)가 정의된다.
- [0155] 신경망 구조와 관련된 정보는, 모델(예컨대, 텐서 생성 모델)을 구성하는 MLP와 CNN의 크기를 정의하는 선택, 모델을 구성하고 있는 연결(edge)의 구조를 정의하는 선택, 모델을 구성하는 파라미터들의 형식 및 정밀도(precision)를 정의하는 선택, 모델을 구성하는 MLP와 CNN의 크기 및 개수에 관한 상한과 하한을 정의하는 선택 등을 포함한다.
- [0156] 카메라 파라미터들과 관련된 정보는, 전술한 바와 같은 카메라 내부 파라미터들 및 카메라 외부 파라미터들을 포함한다.
- [0157] 양자화 파라미터와 관련된 정보는, 2D 비디오 코덱의 양자화 파라미터, 2D 비디오 코덱/LIC 기반 압축 시, 양자화되기 전의 특징의 최소값과 최대값, 2D 비디오 코덱/LIC 기반의 특징 패킹을 정의하는 선택, NNC의 양자화 파라미터. NNC의 양자화 파라미터의 사용 가능한 크기의 상한과 하한을 정의하는 선택 등을 포함한다.
- [0158] 움직임 보상에 대한 정보는 GOV의 구조를 정의하는 선택을 포함한다.
- [0159] 입력 범위와 관련된 정보는 4D 시공간 좌표 및 시점 방향의 상한과 하한을 정의하는 선택을 포함한다. 전술한 바와 같이, 4D 시공간 좌표는 (x, y, z, t) 를 나타낸다.
- [0160] 한편, IPS는 무손실 압축된 후, 전달될 수 있다.
- [0161] 이하, 도 15 및 도 16의 도시를 이용하여, 동적 볼륨, 즉, 동적 다중시점 비디오를 부호화하거나 복원하는 방법을 기술한다.
- [0162] 도 15는 본 개시의 일 실시예에 따른, 볼륨 부호화 장치가 동적 다중시점 비디오를 부호화하는 방법을 나타내는 순서도이다.
- [0163] 볼륨 부호화 장치는 현재 시간 및 현재 시점(view point)을 획득한다(S1500).
- [0164] 볼륨 부호화 장치는 I 볼륨 및 B 볼륨을 포함하는 GOV(Group of Volume)의 구조에 관한 정보를 획득할 수 있다. 볼륨 부호화 장치는 GOV를 활용함으로써, 동적 3D 비디오의 시간을 처리할 수 있다.
- [0165] 볼륨 부호화 장치는 현재 시간이 I 볼륨의 부호화에 대응하는 경우, I 볼륨을 획득하고, I 볼륨을 분해하여 I 볼륨의 텐서 특징들을 표현하는 저차원의 텐서들을 생성한다(S1502).
- [0166] 볼륨 부호화 장치는 도 9a 또는 도 9b와 같이, 벡터행렬 분해 또는 CP(Canonical Polyadic) 분해에 기초하여 I 볼륨을 분해할 수 있다.
- [0167] 볼륨 부호화 장치는 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 텐서 생성 모델을 이용하여 최인접하는 두

개의 기본해된 I 볼륨을 기반으로 B 볼륨의 텐서 특징들을 재구성한다(S1504).

- [0168] 볼륨 부호화 장치는 GOV의 구조에 관한 정보에 기초하여 최인접하는 두 개의 I 볼륨을 결정할 수 있다. 볼륨 부호화 장치는 최인접하는 두 개의 I 볼륨으로부터 동일한 축을 갖는 텐서들을 획득할 수 있다. 볼륨 부호화 장치는 동일한 축을 갖는 텐서들을 텐서 생성 모델에 입력하여, 도 10 또는 도 11의 예시와 같이, 생성 텐서를 구성할 수 있다.
- [0169] 일 예로서, 텐서 생성 모델은, 도 10의 예시와 같이, 인코더와 디코더를 포함할 수 있다. 텐서 생성 모델은, 동일한 축을 갖는 텐서들을 인코더에 입력하여 중간 특징을 생성하고, 중간 특징에 현재 시간을 누적할 수 있다. 텐서 생성 모델은, 현재 시간이 누적된 중간 특징을 디코더에 입력하여 생성 텐서를 구성할 수 있다.
- [0170] 볼륨 부호화 장치는 인접하는 I 볼륨으로부터 생성 텐서와 직교하는 고정 텐서를 획득할 수 있다. 볼륨 부호화 장치는 도 12a 내지 도 12c의 예시와 같이, 고정 텐서 및 생성 텐서에 기초하여 B 볼륨의 분해된 텐서들을 재구성할 수 있다.
- [0171] 볼륨 부호화 장치는 I 볼륨의 텐서 특징들을 부호화한다(S1506).
- [0172] 일 예로서, 볼륨 부호화 장치는 I 볼륨의 텐서 특징들을 비디오 형식으로 패키징할 수 있다. 볼륨 부호화 장치는 패키징된 I 볼륨의 텐서 특징들에 비디오 부호화 방법을 적용하여 비트스트림을 생성하고, 생성된 비트스트림을 볼륨 복호화 장치에게 전송할 수 있다.
- [0173] 다른 예로서, 볼륨 부호화 장치는 I 볼륨의 텐서 특징들을 패키징할 수 있다. 볼륨 부호화 장치는 패키징된 I 볼륨의 텐서 특징들에 딥러닝 기반 압축 모델을 적용하여 비트스트림을 생성하고, 생성된 비트스트림을 볼륨 복호화 장치에게 전송할 수 있다.
- [0174] 다른 예로서, 볼륨 부호화 장치는 부호화 파이프라인에 기초하는 신경망 압축 방식을 이용하여 I 볼륨의 텐서 특징들을 압축할 수 있다.
- [0175] 볼륨 부호화 장치는 텐서 생성 모델의 가중치들을 부호화한다(S1508).
- [0176] 일 예로서, 볼륨 부호화 장치는 부호화 파이프라인에 기초하는 신경망 압축 방식을 이용하여 텐서 생성 모델의 가중치들을 부호화할 수 있다. 볼륨 부호화 장치는 텐서 생성 모델의 가중치들을 프루닝할 수 있다. 볼륨 부호화 장치는 양자화 파라미터를 이용하여 프루닝된 가중치들을 양자화하고, 양자화된 가중치들을 부호화할 수 있다.
- [0177] 한편, 볼륨 부호화 장치는 현재 시간이 I 볼륨의 부호화에 대응하는 경우, 카메라 파라미터들 및 I 볼륨의 텐서 특징들에 기초하여 현재 시점의 영상을 생성할 수 있다. 전술한 바와 같이, 카메라 파라미터들에 기초하여 "볼륨 상의 좌표/시점" 과 이미지 평면에서의 픽셀 간의 변환이 정의될 수 있다.
- [0178] 볼륨 부호화 장치는 I 볼륨의 분해된 텐서들에 기초하여 현재 시점에 대응하는 복셀들에서 색상들 및 볼륨 밀도들을 산정할 수 있다. 볼륨 부호화 장치는 색상들 및 볼륨 밀도들에 기초하여 현재 시점의 영상을 생성할 수 있다.
- [0179] 또한, 볼륨 부호화 장치는 현재 시간이 B 볼륨의 부호화에 대응하는 경우, 카메라 파라미터들 및 재구성된 B 볼륨의 텐서 특징들에 기초하여 현재 시점의 영상을 생성할 수 있다.
- [0180] 볼륨 부호화 장치는 B 볼륨의 분해된 텐서들에 기초하여 현재 시점에 대응하는 복셀들에서 색상들 및 볼륨 밀도들을 산정할 수 있다. 볼륨 부호화 장치는 색상들 및 볼륨 밀도들에 기초하여 현재 시점의 영상을 생성할 수 있다.
- [0181] 볼륨 부호화 장치는 현재 시점의 영상과 GT(Ground Truth)에 기초하여 I 볼륨의 텐서 특징들 및 텐서 생성 모델을 트레이닝할 수 있다. 여기서, GT는 I 볼륨 또는 B 볼륨에 대응하는 원본 영상들을 포함할 수 있다.
- [0182] 볼륨 부호화 장치는 현재 시점의 영상과 원본 영상 간의 차이에 기초하는 손실함수를 산정할 수 있다. 볼륨 부호화 장치는 손실함수를 감소시키는 방향으로, I 볼륨의 텐서 특징들을 업데이트하고, 텐서 생성 모델의 가중치들을 업데이트할 수 있다.
- [0183] 일 예로서, 볼륨 부호화 장치는 부가 정보로서, GOV의 구조에 관한 정보를 부호화할 수 있다.
- [0184] 또한, 볼륨 부호화 장치는 양자화 파라미터와 관련된 정보, 텐서 생성 모델의 구조에 관한 정보, 및 카메라 파

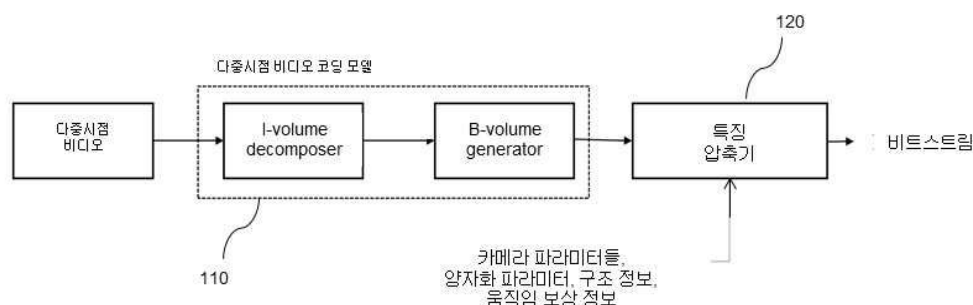
라미터들을 포함하는 부가 정보를 부호화할 수 있다.

- [0185] 도 16은 본 개시의 일 실시예에 따른, 볼륨 복호화 장치가 동적 다중시점 비디오를 복원하는 방법을 나타내는 순서도이다.
- [0186] 볼륨 복호화 장치는 비트스트림으로부터 I 볼륨을 표현하는 텐서 특징들, 텐서 생성 모델의 가중치들, 및 부가 정보를 복호화한다(S1600).
- [0187] 여기서, I 볼륨의 텐서 특징들은 분해된 텐서들(decomposed tensors)에 의해 표현된다. I 볼륨의 분해된 텐서들은, 벡터행렬 분해 또는 CP(Canonical Polyadic) 분해에 의해 생성된 저차원 텐서들일 수 있다.
- [0188] 텐서 생성 모델은 B 볼륨의 재구성에 사용될 수 있다.
- [0189] 부가 정보는 양자화 파라미터와 관련된 정보, 텐서 생성 모델의 구조에 관한 정보, 및 카메라 파라미터들을 포함할 수 있다.
- [0190] 볼륨 복호화 장치는 다음과 같이 I 볼륨의 텐서 특징들을 복호화할 수 있다.
- [0191] 일 예로서, 볼륨 복호화 장치는 비디오 형식으로 패키징된 I 볼륨의 텐서 특징들을 복호화할 수 있다. 볼륨 복호화 장치는 비디오 형식의 I 볼륨의 텐서 특징들에 비디오 복호화 방법을 적용하여, 언패킹된 I 볼륨의 텐서 특징들을 생성할 수 있다.
- [0192] 다른 예로서, 볼륨 복호화 장치는 딥러닝 기반 압축 모델의 인코더에 의해 압축된 I 볼륨의 텐서 특징들을 복호화할 수 있다. 볼륨 복호화 장치는 압축된 I 볼륨의 텐서 특징들에 딥러닝 기반 압축 모델의 디코더를 적용하여, 역압축된 I 볼륨의 텐서 특징들을 생성할 수 있다.
- [0193] 다른 예로서, 볼륨 복호화 장치는 복호화 파이프라인에 기초하는 신경망 압축 방식을 이용하여 I 볼륨의 텐서 특징들을 복호화할 수 있다.
- [0194] 또한, 볼륨 복호화 장치는 부가 정보로서, I 볼륨 및 B 볼륨을 포함하는 GOV(Group of Volume)의 구조에 관한 정보를 복호화할 수 있다. 볼륨 복호화 장치는 GOV를 활용함으로써, 동적 3D 비디오의 시간을 처리할 수 있다.
- [0195] 볼륨 복호화 장치는 텐서 생성 모델의 가중치들, 및 양자화 파라미터에 관한 정보를 이용하여 텐서 생성 모델의 구조에 관한 정보에 부합하도록 텐서 생성 모델을 복원한다(S1602).
- [0196] 일 예로서, 볼륨 복호화 장치는 복호화 파이프라인에 기초하는 신경망 압축 방식을 이용하여 텐서 생성 모델의 양자화된 가중치들을 복호화할 수 있다. 볼륨 복호화 장치는 양자화된 가중치들을 역양자화하여 프루닝이 적용된 가중치들을 생성할 수 있다. 볼륨 복호화 장치는 역양자화된 가중치들을 이용하여 텐서 생성 모델을 생성할 수 있다.
- [0197] 볼륨 복호화 장치는 텐서 생성 모델을 이용하여 최인접하는 두 개의 기복호화된 I 볼륨을 기반으로 B 볼륨의 텐서 특징들을 재구성한다(S1604).
- [0198] 일 예로서, 볼륨 복호화 장치는 GOV의 구조에 관한 정보에 기초하여 최인접하는 두 개의 I 볼륨의 위치를 결정할 수 있다.
- [0199] 볼륨 복호화 장치는 최인접하는 두 개의 I 볼륨으로부터 동일한 축을 갖는 텐서들을 획득할 수 있다. 볼륨 복호화 장치는 동일한 축을 갖는 텐서들을 텐서 생성 모델에 입력하여, 도 10 또는 도 11의 예시와 같이, 생성 텐서를 구성할 수 있다.
- [0200] 일 예로서, 텐서 생성 모델은, 도 10의 예시와 같이, 인코더와 디코더를 포함할 수 있다. 텐서 생성 모델은, 동일한 축을 갖는 텐서들을 인코더에 입력하여 중간 특징을 생성하고, 중간 특징에 현재 시간을 누적할 수 있다. 텐서 생성 모델은, 현재 시간이 누적된 중간 특징을 디코더에 입력하여 생성 텐서를 구성할 수 있다.
- [0201] 볼륨 복호화 장치는 인접하는 I 볼륨으로부터 생성 텐서와 직교하는 고정 텐서를 획득할 수 있다. 볼륨 복호화 장치는 도 12a 내지 도 12c의 예시와 같이, 고정 텐서 및 생성 텐서에 기초하여 B 볼륨의 분해된 텐서들을 재구성할 수 있다.
- [0202] 볼륨 복호화 장치는 현재 시간 및 현재 시점(view point)을 획득한다(S1606).
- [0203] 볼륨 복호화 장치는 현재 시간이 I 볼륨의 복호화에 대응하는 경우, 카메라 파라미터들 및 I 볼륨의 텐서 특징들에 기초하여 현재 시점의 영상을 생성한다(S1608).

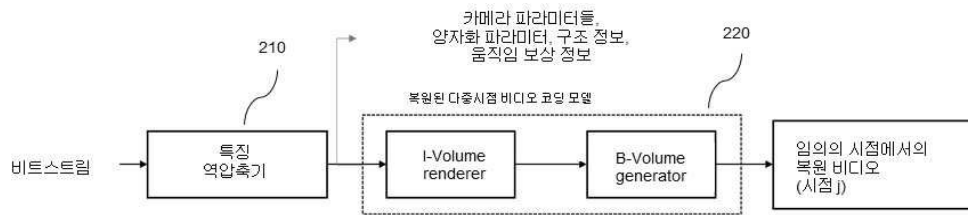
- [0204] 볼륨 복호화 장치는 I 볼륨의 분해된 텐서들에 기초하여 현재 시점에 대응하는 복셀들에서 색상들 및 볼륨 밀도들을 산정할 수 있다. 볼륨 복호화 장치는 색상들 및 볼륨 밀도들에 기초하여 현재 시점의 영상을 생성할 수 있다.
- [0205] 전술한 바와 같이, 카메라 파라미터들에 기초하여 "볼륨 상의 좌표/시점" 과 이미지 평면에서의 픽셀 간의 변환이 정의될 수 있다.
- [0206] 볼륨 복호화 장치는 현재 시간이 B 볼륨의 복호화에 대응하는 경우, 카메라 파라미터들 및 재구성된 B 볼륨의 텐서 특징들에 기초하여 현재 시점의 영상을 생성한다(S1610).
- [0207] 볼륨 복호화 장치는 B 볼륨의 분해된 텐서들에 기초하여 현재 시점에 대응하는 복셀들에서 색상들 및 볼륨 밀도들을 산정할 수 있다. 볼륨 복호화 장치는 색상들 및 볼륨 밀도들에 기초하여 현재 시점의 영상을 생성할 수 있다.
- [0208] 볼륨 복호화 장치는 I 볼륨과 B 볼륨에 대응하는 2D 영상을 연결함으로써, 임의의 시점에 대응하는 2D 비디오를 생성할 수 있다.
- [0209] 본 명세서의 흐름도/타이밍도에서는 각 과정들을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 본 개시의 일 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것이다. 다시 말해, 본 개시의 일 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 개시의 일 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 흐름도/타이밍도에 기재된 순서를 변경하여 실행하거나 각 과정들 중 하나 이상의 과정을 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 흐름도/타이밍도는 시계열적인 순서로 한정되는 것은 아니다.
- [0210] 이상의 설명에서 예시적인 실시예들은 많은 다른 방식으로 구현될 수 있다는 것을 이해해야 한다. 하나 이상의 예시들에서 설명된 기능들 혹은 방법들은 하드웨어, 소프트웨어, 펌웨어 또는 이들의 임의의 조합으로 구현될 수 있다. 본 명세서에서 설명된 기능적 컴포넌트들은 그들의 구현 독립성을 특히 더 강조하기 위해 "...부(unit)" 로 라벨링되었음을 이해해야 한다.
- [0211] 한편, 본 실시예에서 설명된 다양한 기능들 혹은 방법들은 하나 이상의 프로세서에 의해 관독되고 실행될 수 있는 비일시적 기록매체에 저장된 명령어들로 구현될 수도 있다. 비일시적 기록매체는, 예를 들어, 컴퓨터 시스템에 의하여 관독가능한 형태로 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 예를 들어, 비일시적 기록매체는 EPROM(erasable programmable read only memory), 플래시 드라이브, 광학 드라이브, 자기 하드 드라이브, 솔리드 스테이트 드라이브(SSD)와 같은 저장매체를 포함한다.
- [0212] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.
- [0213]

도면

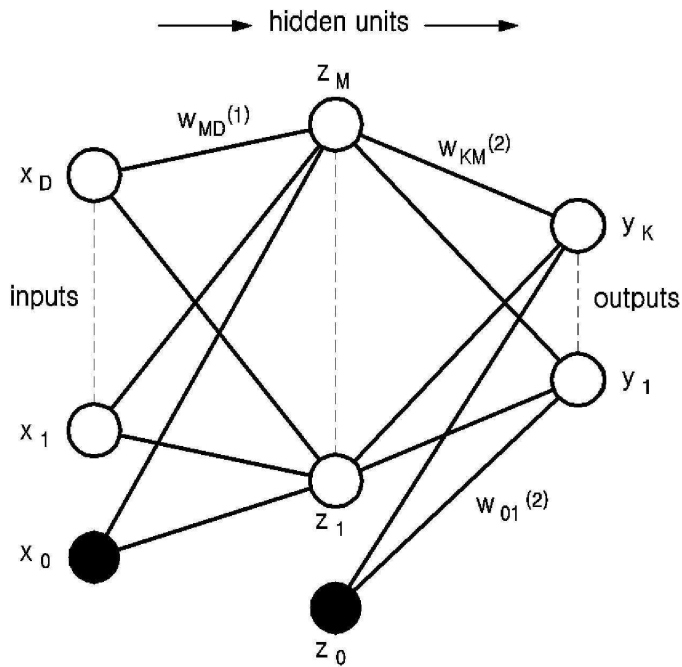
도면1



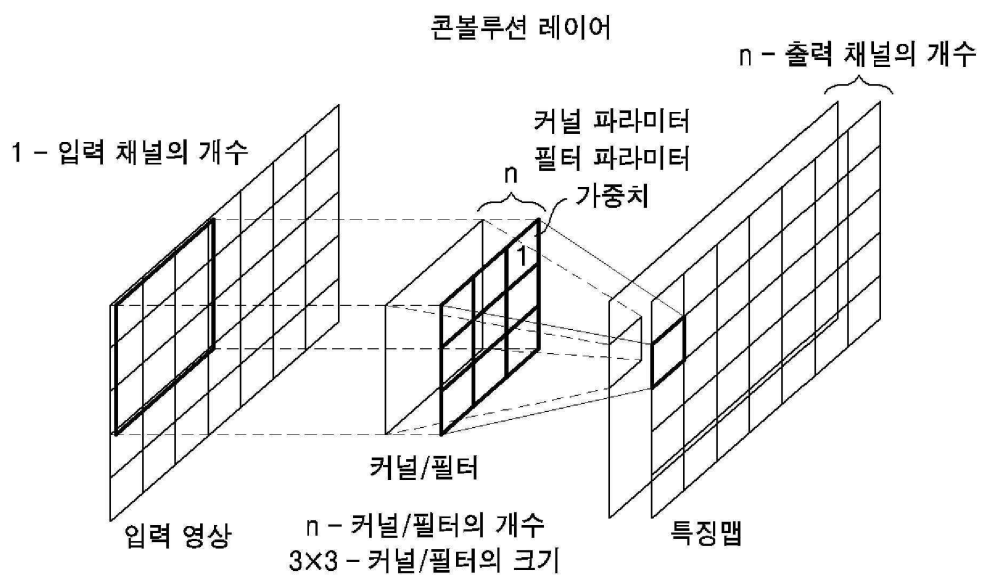
도면2



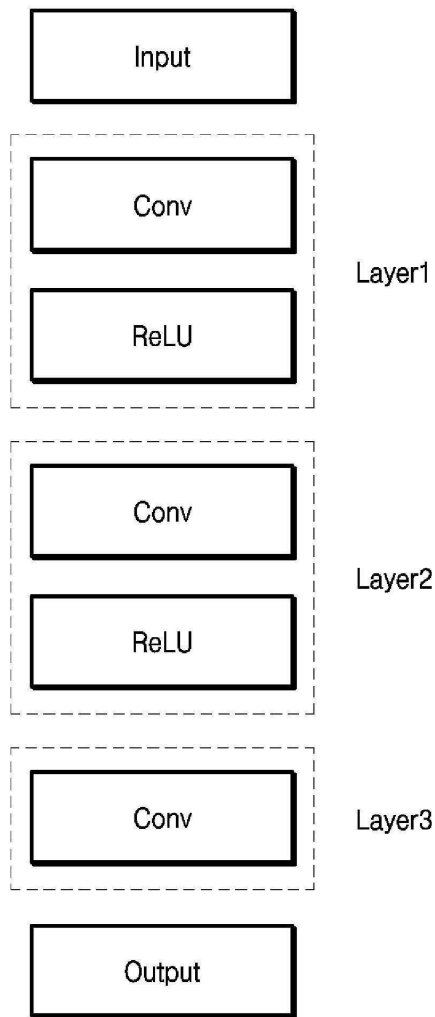
도면3



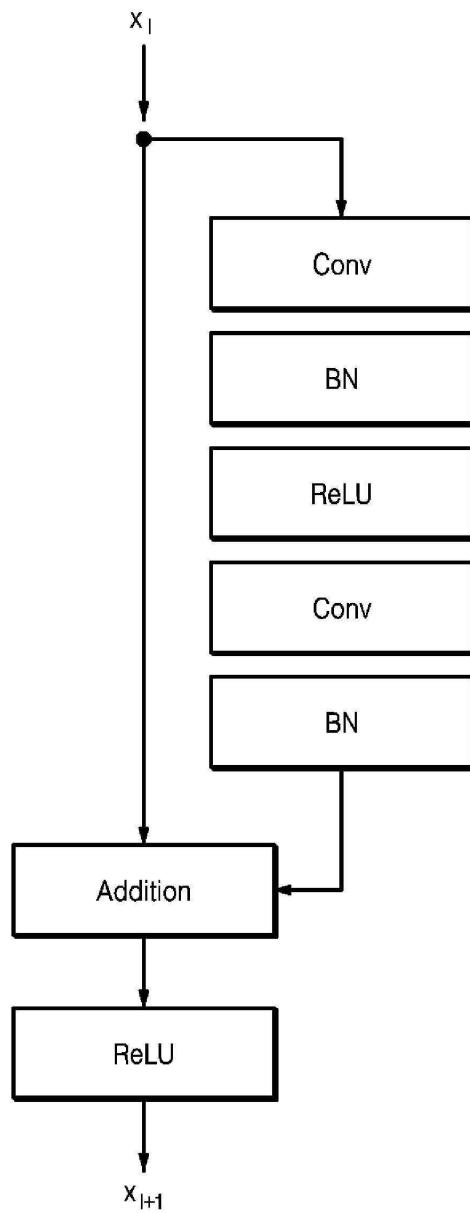
도면4



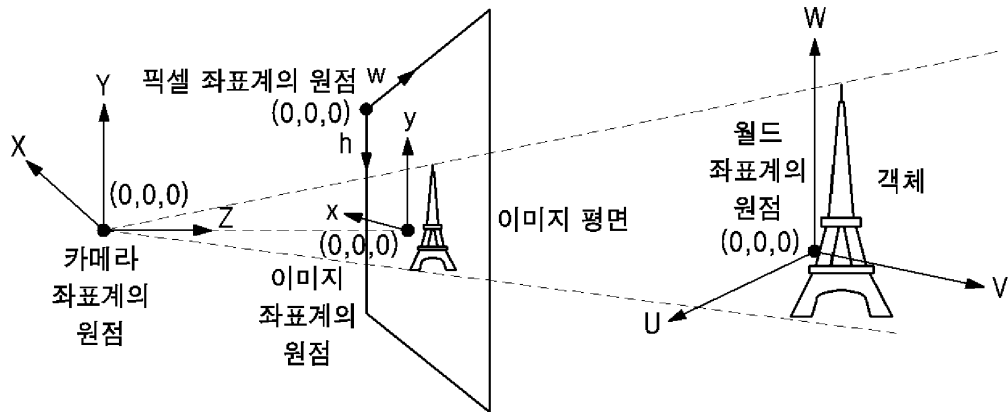
도면5



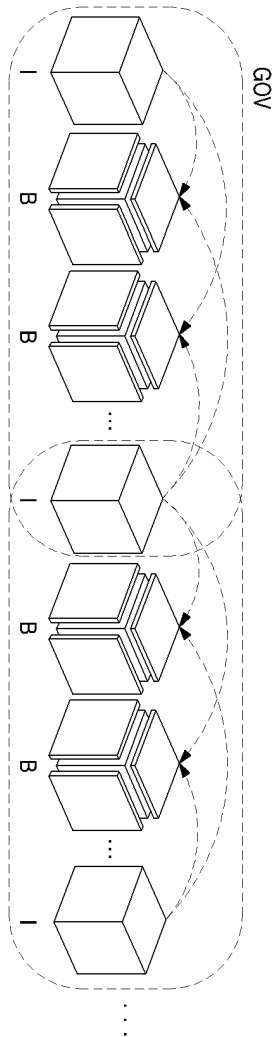
도면6



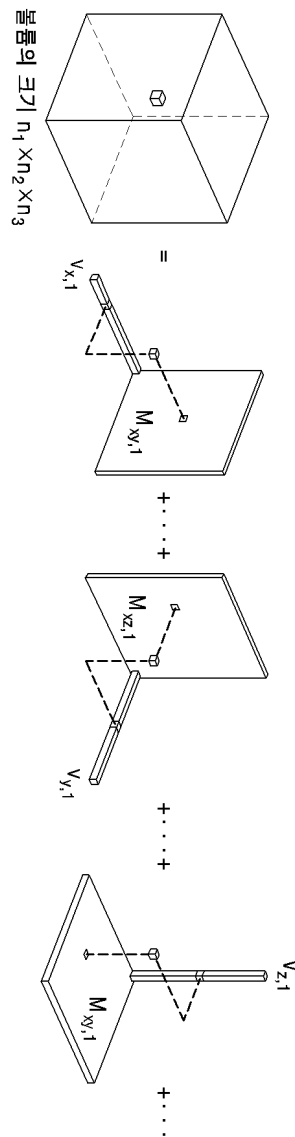
도면7



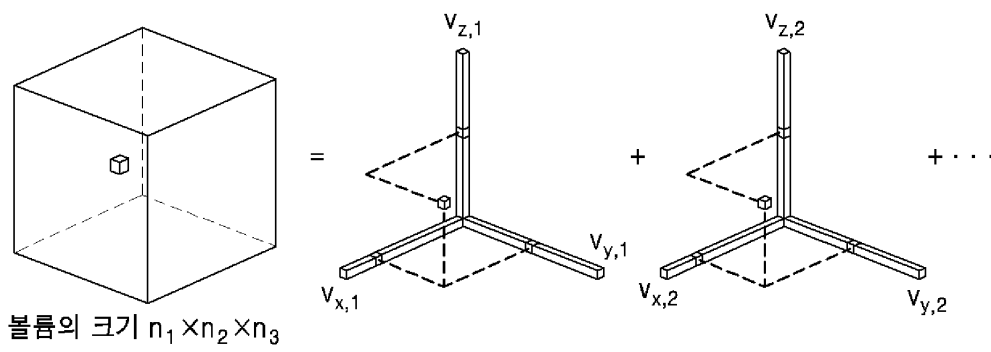
도면8



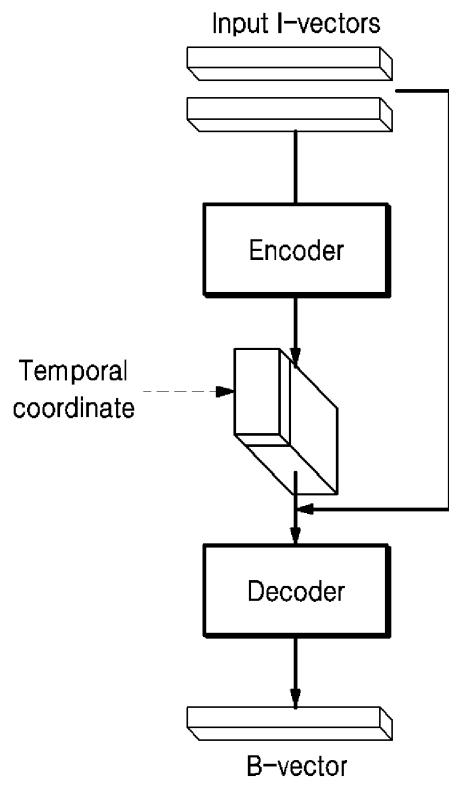
도면9a



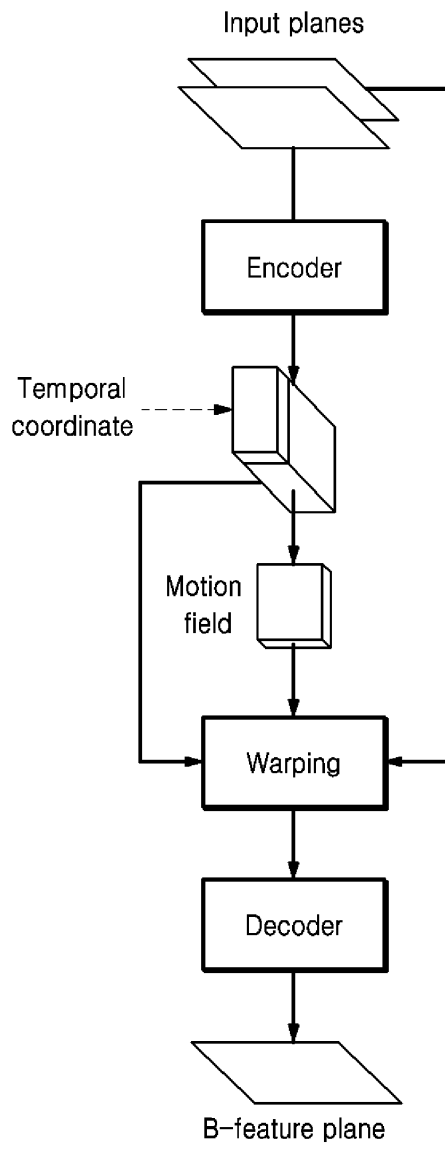
도면9b



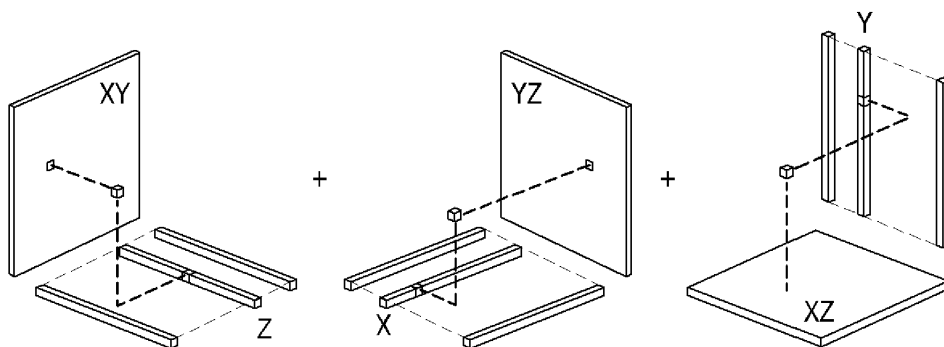
도면10



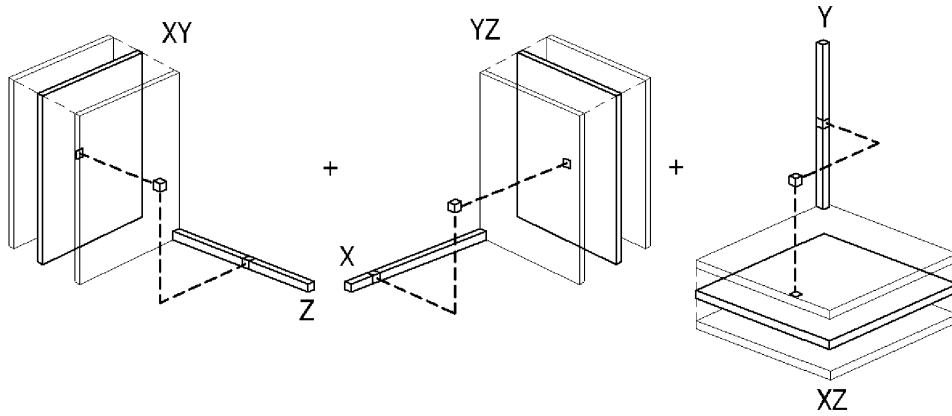
도면11



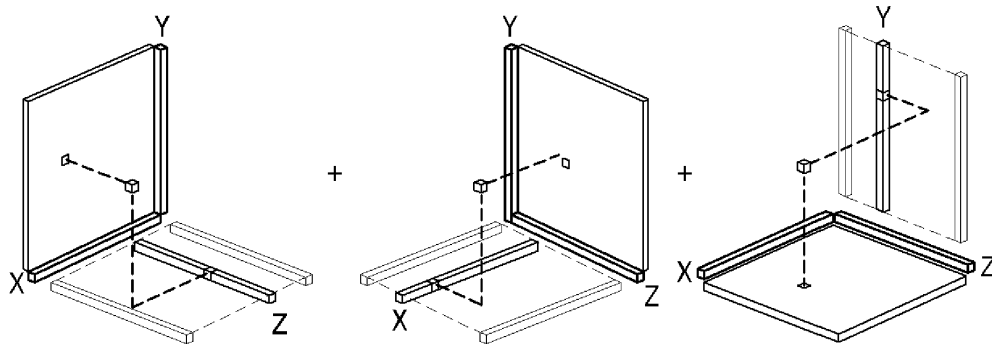
도면12a



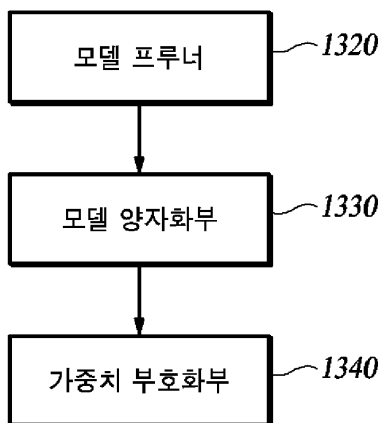
도면12b



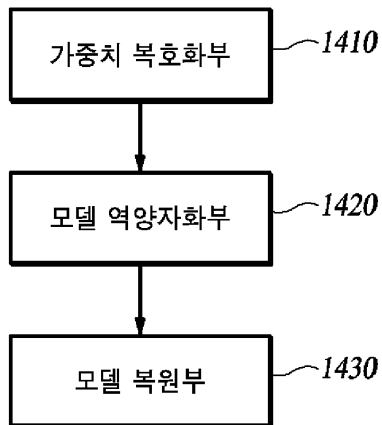
도면12c



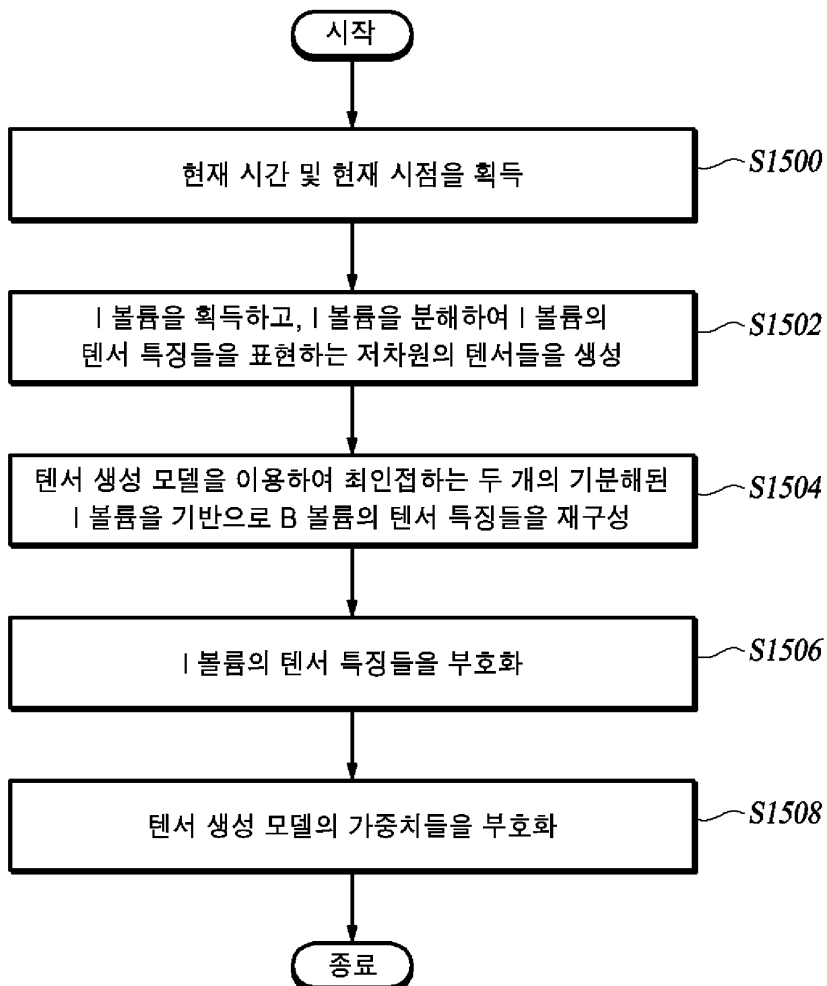
도면13



도면14



도면15



도면16

