(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0087445
(43) 공개일자 2025년06월16일

(51) 국제특허분류(Int. Cl.)
H04N 19/60 (2014.01) *G06N 3/0464* (2023.01)
H04N 19/124 (2014.01) *H04N 19/132* (2014.01)
(52) CPC특허분류
H04N 19/60 (2015.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2024-0154232
(22) 출원일자 2024년11월04일
심사청구일자 없음
(30) 우선권주장
1020230176421 2023년12월07일 대한민국(KR)

(71) 출원인
현대자동차주식회사
서울특별시 서초구 헌릉로 12 (양재동)
기아 주식회사
서울특별시 서초구 헌릉로 12 (양재동)
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
강제원
서울특별시 마포구 독막로 145, 104-904
이지후
서울특별시 성북구 보문사길 111, 102동 1102호
(뒷면에 계속)
(74) 대리인
이철희

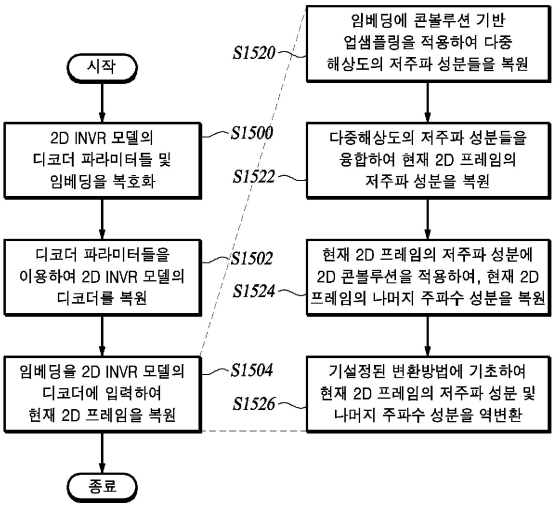
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **암시적 신경망 비디오 표현에 기반하는 비디오 코딩방법 및 장치**

(57) 요 약

본 실시예는 암시적 신경망 비디오 표현에 기반하는 비디오 코딩방법 및 장치를 개시한다. 본 실시예에서, 영상 복호화 장치는 2D 암시적 신경망 비디오 표현(Implicit Neural Video Representation, INVR) 모델의 디코더 파라미터들 및 임베딩을 복호화한다. 여기서, 임베딩은 현재 2D 프레임의 저주파 성분의 압축된 표현이고 2D INVR 모델의 인코더에 의해 생성되고 제공된다. 영상 복호화 장치는 디코더 파라미터들을 이용하여 2D INVR 모델의 디코더를 복원한다. 영상 복호화 장치는 임베딩을 2D INVR 모델의 디코더에 입력하여 현재 2D 프레임을 복원한다.

대 표 도 - 도15



(52) CPC특허분류
H04N 19/124 (2015.01)
H04N 19/132 (2015.01)
(72) 발명자
김진아
서울특별시 송파구 문정로 83, 104-1404
최정아
경기도 화성시 남양읍 현대연구소로 150
허진
경기도 화성시 남양읍 현대연구소로 150
박승욱
경기도 화성시 남양읍 현대연구소로 150

명 세 서

청구범위

청구항 1

영상 복호화 장치가 수행하는, 현재 2D 프레임을 복원하는 방법에 있어서,

비트스트림으로부터 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 디코더 파라미터들 및 임베딩을 복호화하는 단계, 여기서, 상기 임베딩은, 상기 현재 2D 프레임의 저주파 성분의 압축된 표현이고 상기 2D INVR 모델의 인코더에 의해 생성되고 제공됨;

상기 디코더 파라미터들을 이용하여 상기 2D INVR 모델의 디코더를 복원하는 단계; 및

상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계

를 포함하고,

상기 현재 2D 프레임을 복원하는 단계는,

상기 임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원하는 단계; 및

상기 다중해상도의 저주파 성분들을 융합하여 상기 현재 2D 프레임의 저주파 성분을 복원하는 단계

를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 현재 2D 프레임을 복원하는 단계는,

상기 현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 상기 현재 2D 프레임의 나머지 주파수 성분을 복원하는 단계

를 더 포함하는, 방법.

청구항 3

제2항에 있어서,

상기 현재 2D 프레임을 복원하는 단계는,

기설정된 변환방법에 기초하여 상기 현재 2D 프레임의 저주파 성분 및 상기 나머지 주파수 성분을 역변환하는 단계

를 더 포함하는, 방법.

청구항 4

제1항에 있어서,

상기 2D INVR 모델의 인코더와 상기 2D INVR 모델의 디코더는 손실함수에 따라 종단간으로 사전에 트레이닝되고, 상기 손실함수는 상기 2D INVR 모델의 인코더의 입력 프레임과 상기 2D INVR 모델의 디코더의 출력 프레임 간의 차이에 기초하여 정의되는, 방법.

청구항 5

제1항에 있어서,

상기 비트스트림으로부터 부가 정보로서 양자화 파라미터를 복호화하는 단계;

상기 양자화 파라미터를 이용하여 상기 디코더 파라미터들 및 상기 임베딩을 역양자화하는 단계

를 더 포함하는, 방법.

청구항 6

영상 부호화 장치가 수행하는, 현재 2D 프레임을 부호화하는 방법에 있어서,

상기 현재 2D 프레임을 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 인코더에 입력하여 압축된 임베딩을 생성하는 단계;

상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계; 및

상기 임베딩 및 상기 2D INVR 모델의 디코더 파라미터들을 부호화하는 단계

를 포함하고,

상기 임베딩을 생성하는 단계는,

기설정된 변환방법에 기초하여 상기 현재 2D 프레임을 변환하여 상기 현재 2D 프레임의 저주파 성분을 생성하는 단계; 및

상기 저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 상기 임베딩을 생성하는 단계

를 포함하는, 방법.

청구항 7

제6항에 있어서,

상기 현재 2D 프레임을 복원하는 단계는,

상기 임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원하는 단계; 및

상기 다중해상도의 저주파 성분들을 융합하여 상기 현재 2D 프레임의 저주파 성분을 복원하는 단계

를 포함하는, 방법.

청구항 8

제6항에 있어서,

상기 현재 2D 프레임을 복원하는 단계는,

상기 현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 상기 현재 2D 프레임의 나머지 주파수 성분을 복원하는 단계

를 더 포함하는, 방법.

청구항 9

제8항에 있어서,

상기 현재 2D 프레임을 복원하는 단계는,

상기 기설정된 변환방법에 기초하여 상기 현재 2D 프레임의 저주파 성분 및 상기 나머지 주파수 성분을 역변환하는 단계

를 더 포함하는, 방법.

청구항 10

제6항에 있어서,

상기 2D INVR 모델의 인코더와 상기 2D INVR 모델의 디코더를 손실함수에 기초하여 종단간으로 트레이닝하는 단계를 더 포함하고,

상기 손실함수는 상기 2D INVR 모델의 인코더의 입력 프레임과 상기 2D INVR 모델의 디코더의 출력 프레임 간의

차이에 기초하여 정의되는, 방법.

청구항 11

제6항에 있어서,

양자화 파라미터에 기초하여 상기 디코더 파라미터들 및 상기 임베딩을 양자화하는 단계; 및

부가 정보로서 상기 양자화 파라미터를 부호화하는 단계

를 더 포함하는, 방법.

청구항 12

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 2D 프레임을 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 인코더에 입력하여 압축된 임베딩을 생성하는 단계;

상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계; 및

상기 임베딩 및 상기 2D INVR 모델의 디코더 파라미터들을 부호화하는 단계

를 포함하고,

상기 임베딩을 생성하는 단계는,

기설정된 변환방법에 기초하여 상기 현재 2D 프레임을 변환하여 상기 현재 2D 프레임의 저주파 성분을 생성하는 단계; 및

상기 저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 상기 임베딩을 생성하는 단계

를 포함하는 것을 특징으로 하는, 방법.

발명의 설명

기 술 분 야

[0001] 본 개시는 암시적 신경망 비디오 표현에 기반하는 비디오 코딩방법 및 장치에 관한 것이다.

배 경 기 술

[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 비디오 표현 방식으로서, 픽셀 위치별 RGB 픽셀값을 갖는 명시적(explicit) 표현 방식이 사용된다. 명시적 표현 방식을 대체하기 위해 영상의 픽셀 위치인 (x,y) 좌표에서 (r,g,b) 값을 생성하는 함수를 신경망으로 표현하는 암시적 신경망 표현(implicit neural representation) 방식이 도입되고 있다. 명시적 표현 방식에 비하여 암시적 신경망 표현 방식은 이미지의 해상도와 관계 없이 사용되고, 영상 픽셀의 공간 및 시간 축으로의 좌표를 입력하면 (r,g,b) 값이 즉각적으로 복원된다는 점에서 기존의 디코더 대비 실용적 우수성을 갖는다.

[0004] 한편, 암시적 신경망 표현 기반의 비디오 표현 모델(Implicit Neural Video Representation, INVR, 또는 Neural Representation for Videos, NeRV)에 기반하는 비디오 압축 기술은 픽셀 도메인에서 비디오의 암시적

표현을 생성한다. 신경망의 학습과 관련된 NTK(Neural Tangent Kernel) 이론에 따르면, 신경망의 학습은 저주파 성분을 우선 학습하고 고주파 성분을 이어 학습한다. 기존의 INVR도 NTK 이론에 따른 학습을 수행하므로, 고주파 성분을 학습하기 위해 모델의 크기가 커지고 학습 시간이 길어질 수 있다. 따라서, 비디오 부호화 효율을 향상시키고 비디오 화질을 개선하기 위해, 주파수 성분을 고려하여 암시적 신경망 표현 기반의 비디오 표현 모델을 구성하는 방안이 필요하다.

발명의 내용

해결하려는 과제

[0005] 본 개시는, 주파수 도메인 상의 비디오에 암시적 신경망 비디오 표현(Implicit Neural Video Representation, INVR) 모델을 적용하여 압축된 임베딩을 생성하고, 모델 파라미터 및 임베딩을 전송하며, 전송된 모델 파라미터 및 임베딩에 기초하여 비디오를 복원하는 비디오 코딩방법 및 장치를 제공하는 데 목적이 있다.

[0006] 또한, 본 개시는, INVR 모델을 기반으로 비디오의 저주파 성분을 압축하고, 비디오의 고주파 성분을 디코더 측에서 필요에 따라 생성하는 비디오 코딩방법 및 장치를 제공하는 데 목적이 있다.

과제의 해결 수단

[0007] 본 개시의 실시예에 따르면, 영상 복호화 장치가 수행하는, 현재 2D 프레임을 복원하는 방법에 있어서, 비트스트림으로부터 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 디코더 파라미터들 및 임베딩을 복호화하는 단계, 여기서, 상기 임베딩은, 상기 현재 2D 프레임의 저주파 성분의 압축된 표현이고 상기 2D INVR 모델의 인코더에 의해 생성되고 제공됨; 상기 디코더 파라미터들을 이용하여 상기 2D INVR 모델의 디코더를 복원하는 단계; 및 상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계를 포함하고, 상기 현재 2D 프레임을 복원하는 단계는, 상기 임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원하는 단계; 및 상기 다중해상도의 저주파 성분들을 융합하여 상기 현재 2D 프레임의 저주파 성분을 복원하는 단계를 포함하는, 방법을 제공한다.

[0008] 본 개시의 다른 실시예에 따르면, 영상 부호화 장치가 수행하는, 현재 2D 프레임을 부호화하는 방법에 있어서, 상기 현재 2D 프레임을 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 인코더에 입력하여 압축된 임베딩을 생성하는 단계; 상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계; 및 상기 임베딩 및 상기 2D INVR 모델의 디코더 파라미터들을 부호화하는 단계를 포함하고, 상기 임베딩을 생성하는 단계는, 기설정된 변환방법에 기초하여 상기 현재 2D 프레임을 변환하여 상기 현재 2D 프레임의 저주파 성분을 생성하는 단계; 및 상기 저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 상기 임베딩을 생성하는 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 2D 프레임을 2D(Dimensional) INVR(Implicit Neural Video Representation) 모델의 인코더에 입력하여 압축된 임베딩을 생성하는 단계; 상기 임베딩을 상기 2D INVR 모델의 디코더에 입력하여 상기 현재 2D 프레임을 복원하는 단계; 및 상기 임베딩 및 상기 2D INVR 모델의 디코더 파라미터들을 부호화하는 단계를 포함하고, 상기 임베딩을 생성하는 단계는, 기설정된 변환방법에 기초하여 상기 현재 2D 프레임을 변환하여 상기 현재 2D 프레임의 저주파 성분을 생성하는 단계; 및 상기 저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 상기 임베딩을 생성하는 단계를 포함하는 것을 특징으로 하는, 방법을 제공한다.

발명의 효과

[0010] 이상에서 설명한 바와 같이 본 실시예에 따르면, 주파수 도메인으로 변환된 비디오에 INVR 모델을 적용하여 압축된 임베딩을 생성하고, 모델 파라미터 및 임베딩을 전송하며, 전송된 모델 파라미터 및 임베딩에 기초하여 비디오를 복원하는 비디오 코딩방법 및 장치를 제공함으로써, 비디오 부호화 효율 및 비디오 화질을 개선하는 것이 가능해지는 효과가 있다.

[0011] 또한 본 실시예에 따르면, INVR 모델을 기반으로 비디오의 저주파 성분을 압축하고, 비디오의 고주파 성분을 디코더 측에서 필요에 따라 생성하는 비디오 코딩방법 및 장치를 제공함으로써, INVR 모델이 고주파 성분을 임베

딩하기 위해 필요한 오버헤드를 감소시키는 것이 가능해지는 효과가 있다.

도면의 간단한 설명

[0012] 도 1은 본 개시의 일 실시예에 따른, INVR(Implicit Neural Video Representation) 기반 영상 부호화 장치를 나타내는 블록도이다.

도 2는 본 개시의 일 실시예에 따른, INVR 기반 영상 복호화 장치를 나타내는 블록도이다.

도 3은 본 개시의 일 실시예에 따른 콘볼루션 레이어의 연산을 나타내는 예시도이다.

도 4는 SISR(Single Image Super Resolution) 네트워크를 나타내는 예시도이다.

도 5는 SISR에 이용되는 잔차 블록을 나타내는 예시도이다.

도 6은 NeRV(Neural Representations for Videos) 모델을 나타내는 예시도이다.

도 7a 및 도 7b는 본 개시의 일 실시예에 따른, NeRV 모델의 개념을 나타내는 예시도이다.

도 8은 본 개시의 일 실시예에 따른, NeRV 모델의 부호화 파이프라인을 나타내는 예시도이다.

도 9는 본 개시의 일 실시예에 따른, NeRV 모델의 복호화 파이프라인을 나타내는 예시도이다.

도 10a 및 도 10b는 본 개시의 일 실시예에 따른, 공간 정보 기반 INVR 모델을 나타내는 예시도이다.

도 11은 본 개시의 일 실시예에 따른, 공간 정보 기반 INVR 모델을 나타내는 예시도이다.

도 12는 본 개시의 일 실시예에 따른, INVR 모델의 부호화 파이프라인을 나타내는 예시도이다.

도 13은 본 개시의 일 실시예에 따른, INVR 모델의 복호화 파이프라인을 나타내는 예시도이다.

도 14는 본 개시의 일 실시예에 따른, 영상 부호화 장치가 2D 비디오를 부호화하는 방법을 나타내는 순서도이다.

도 15는 본 개시의 일 실시예에 따른, 영상 복호화 장치가 2D 비디오를 복원하는 방법을 나타내는 순서도이다.

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 부호를 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 영상(비디오)의 부호화 및 복호화에 관한 것이다. 보다 자세하게는, 주파수 도메인으로 변환된 2D 비디오에 암시적 신경망 비디오 표현(Implicit Neural Video Representation, INVR) 모델을 적용하여 압축된 임베딩을 생성하고, 모델 파라미터 및 임베딩을 전송하며, 전송된 모델 파라미터 및 임베딩에 기초하여 비디오를 복원하는 비디오 코딩방법 및 장치를 제공한다.

[0015] 도 1은 본 개시의 일 실시예에 따른, 2D INVR 기반 영상 부호화 장치를 나타내는 블록도이다.

[0016] 도 1의 예시에서 2D(Dimensional) INVR 기반 영상 부호화 장치(이하, '영상 부호화 장치')는 주파수 도메인 상의 2D 비디오의 임베딩을 생성하는 것을 학습하도록 2D INVR 모델을 생성(즉, 트레이닝)하고, 학습된 2D INVR 모델의 디코더 및 임베딩을 압축하여 비트스트림을 생성하며, 생성된 비트스트림을 전송한다. 영상 부호화 장치는 2D INVR 모델의 인코더 및 디코더에 해당하는 부호화기(110), 복호화기(120) 및 신경망 압축기(Neural Network Compressor, 130)를 포함한다. 여기서, 본 개시에 따른 영상 부호화 장치에 포함되는 구성요소가 반드시 이에 한정되는 것은 아니다. 영상 부호화 장치는 2D INVR 모델을 트레이닝하기 위한 트레이닝부(미도시)를 추가로 구비하거나, 외부의 트레이닝부와 연동되는 형태로 구현될 수 있다.

[0017] 영상 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] 영상 부호화 장치는 부호화된 비디오 데이터의 비트스트림을 비일시적인 기록매체에 저장하거나 통신 네트워크

를 이용하여 2D INVR 기반 영상 복호화 장치로 전송할 수 있다.

[0019] 도 2는 본 개시의 일 실시예에 따른, 2D INVR 기반 영상 복호화 장치를 나타내는 블록도이다.

[0020] 도 2의 예시에서 2D INVR 기반 영상 복호화 장치(이하, '영상 복호화 장치')는 비트스트림을 역압축(decompression)하여 2D INVR 모델 및 임베딩을 복원하고, 복원된 2D INVR 모델을 이용하여 임베딩으로부터 2D 비디오를 생성한다. 영상 복호화 장치는 신경망 역압축기(Neural Network Decompressor, 210), 및 복원된 2D INVR 모델의 디코더를 포함하는 복호화기(220)를 포함한다.

[0021] 도 1에 예시된 영상 부호화 장치와 마찬가지로, 영상 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] 기존의 INVR에서는 픽셀 도메인에서 비디오의 암시적 표현을 생성하나, 본 개시에 따른 영상 부호화 장치는 비디오를 주파수 도메인으로 변환하고, 주파수 도메인 상의 비디오를 2D INVR 모델에 입력하여, 비디오의 압축을 수행한다. 주파수 도메인 상의 비디오에 대해 암시적 신경망을 생성 시, 영상 부호화 장치는 비디오의 저주파 신호와 고주파 신호를 분리한 후, 암시적 신경망을 트레이닝한다. 고주파는 저주파에 비하여 학습이 어려우므로, 영상 부호화 장치는 비디오의 정해진 임계치에 따라 고주파 신호의 부호화 과정을 생략하고, 저주파 신호를 우선적으로 학습하여 암시적 신경망에 임베딩한다. 고주파 신호를 임베딩하기 위해, 신경망의 크기가 증가하고 전송 비트율이 증가될 수 있다. 따라서, 영상 부호화 장치는 필요에 따라 디코더에서 고주파 성분를 생성함으로써, 고주파 성분을 임베딩하기 위해 필요한 오버헤드(overhead)를 감소시킬 수 있다. 또한, 영상 복호화 장치는 고정된 비트스트림의 크기에 기초하여 출력 비디오의 품질을 향상시킬 수 있다.

[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 다수의 뉴런들 간을 연결하는 에지(edge)로 구성된다. 다수의 뉴런들은 레이어들에 포함된다. MLP는 입력 레이어(input layer)와 출력 레이어(output layer) 외에, 하나 이상의 은닉 레이어(hidden layer)를 포함할 수 있다. 은닉 레이어는 하나 이상의 은닉 노드(hidden node)를 포함할 수 있다. 각 에지에는 상이한 가중치(weight)가 할당될 수 있다. 하나의 레이어에서 다음 레이어로 출력이 전파되는 과정에서, 활성화(activation) 함수가 사용될 수 있다. 활성화 함수로는 시그모이드(sigmoid) 함수, 탄젠트 쌍곡선(tangent hyperbolic) 함수. 또는 Relu(Rectified Liner unit) 함수가 사용될 수 있다.

[0026] 대상 행위를 학습하기 위해, MLP를 구성하는 가중치들을 업데이트 과정을 트레이닝이라고 한다. 트레이닝은 일반적으로 역전파(back propagation) 알고리즘에 기초하는 SGD(Stochastic Gradient Descent) 방식을 활용한다. 트레이닝에 따라 생성된 가중치들을 이용하여 MLP의 정방향(feed-forward) 출력을 산정하는 과정을 추론 또는 테스트(test)로 명칭한다.

[0027] 이하, 다층신경망, MLP, 또는 정방향 네트워크는 호환적으로 사용될 수 있다.

[0028] **I-2. CNN(Convolutional Neural Network)**

[0029] CNN은 복수의 콘볼루션 레이어(convolution layer)와 풀링 레이어(pooling layer)로 구성된 신경망을 지칭하고, 영상 처리에 가장 적합한 것으로 알려진 딥러닝 기술이다. 콘볼루션 레이어는 다수의 커널(kernel) 또는 필터(filter)를 이용하여 특징맵(feature map, 또는 '특성'과 호환적으로 이용)을 추출한다. 이때 필터를 구성하는 커널 계수(kernel coefficient)가 학습 과정에서 결정되는 파라미터이다.

[0030] CNN의 콘볼루션 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.

[0031] 도 3은 본 개시의 일 실시예에 따른 콘볼루션 레이어의 연산을 나타내는 예시도이다.

[0032] 콘볼루션 레이어는 콘볼루션 연산을 이용하여 입력 영상으로부터 특징맵을 생성한다. 도 3의 예시에는, 커널 크기가 3×3인 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 3에 예시된 커널은 총 9 개의 커널 파라미터를 갖는다. 커널 파라미터는 초기

에 임의의 값으로 설정되고, 트레이닝(learning)을 기반으로 그 값이 업데이트될 수 있다.

[0033] 콘볼루션 레이어는 입력 영상에서 커널 사이즈만큼의 블록을 이용하여 콘볼루션 연산을 수행한다. 이때, 입력 영상 내의 커널 사이즈만큼의 블록을 윈도우(window)로 지칭한다.

[0034] 입력 영상에 대한 필터링을 래스터 스캔 순서(raster-scan order)로 수행할 때, 윈도우의 이동 크기를 스트라이드(stride)라고 한다. 도 3의 예시에서, 스트라이드는 1이다. 만약 스트라이드를 2로 설정하면, 윈도우를 2 샘플씩 띄어서 콘볼루션 연산을 수행하고, 결과적으로 특징맵의 가로와 세로의 크기는 입력 영상의 가로와 세로의 크기의 반이 된다.

[0035] 전술한 바와 같이, 하나의 콘볼루션 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.

[0036] 패딩(padding)은 콘볼루션 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, 콘볼루션 레이어를 거칠 때마다 출력 데이터의 공간적 크기가 감소하여, 경계의 정보들이 사라지는 문제가 발생할 수 있다. 따라서, 이러한 문제를 방지하기 위해, 패딩이 사용된다. 즉, 콘볼루션 레이어의 출력 데이터와 입력 데이터의 공간적 크기를 동일하게 맞추기 위해 패딩이 사용될 수 있다.

[0037] 디콘볼루션 레이어는 콘볼루션 레이어와 반대되는 연산을 수행한다. 디콘볼루션 레이어는 입력인 특징맵으로부터 원하는 데이터 영상을 출력으로 생성한다.

[0038] 풀링 레이어는 콘볼루션 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.

[0039] 풀링 레이어의 반대 개념은 언풀링(unpooling) 레이어로 정의한다. 언풀링 레이어는 풀링 레이어와 반대로 차원을 확대하는 역할을 하고, 주로 디콘볼루션 레이어 이후에 사용된다.

[0040] 콘볼루션 인코더(convolutional encoder)-디코더(decoder) 구조는 콘볼루션 레이어들와 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 언풀링 레이어로 구성되어, 특징맵 또는 잠재 벡터로부터 출력 영상을 생성한다.

[0041] 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.

[0042] 도 4는 SISR 네트워크를 나타내는 예시도이다.

[0043] CNN이 적용되는 일 예로서 SISR(Single Image Super Resolution)이 있다. SISR 네트워크는 저해상도 입력 영상으로부터 고해상도 이미지를 출력으로 생성한다. SISR 네트워크는 도 4에 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.

[0044] 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 5에 예시된 바와 같이, 입력 특징 x_l에 콘볼루션 연산을 적용하는 경로 외에 스킵 경로를 포함한다. 또한, 잔차 블록은 출력 x_{l+1}을 생성 시, 학습 효율에 기초하여 콘볼루션 연산을 적용하는 경로 또는 스킵 경로를 선택할 수도 있다. 도 5의 예시에서, 잔차 블록은 BN(Batch Normalization) 레이어를 포함한다.

[0045] 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 프레임을 최종 출력에 더해 주는 방식인 잔차신호 기반 학습(residual learning)을 사용한다. VDSR은 네트워크의 맨 마지막에 잔차 신호를 추가하여, 입력 신호에 잔차 신호를 더한다.

[0046] **I-3. 암시적 신경망 표현(implicit neural representation) 모델**

[0047] 최근 이미지를 비롯한 각종 데이터의 표현을 신경망 구조로 나타내는 암시적 신경망 표현 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.

[0048] 암시적 신경망 표현 모델을 이용하는 2차원 비디오 코딩의 일 예로서, 암시적 신경망 표현 기반의 비디오 코딩(Implicit Neural Video Representation for Coding, INVRC)이 존재한다. INVRC 기술은 영상의 픽셀 위치인 (x,y) 좌표에 대해 (r,g,b) 값을 생성한다. INVRC 기술의 일 예로서, NeRV(Neural Representation for Video) 모델이 존재한다. 전술한 바와 같이, 명시적 표현 방식은, 이미지 내 픽셀의 위치 (x,y)와 프레임의 시간 인덱스 t를 이용하여 (x,y,t) 그리드 상에 픽셀값을 표현한다. 암시적 신경망 표현 모델은, (x,y,t)의 위치를 나타내는 입력으로부터 RGB 픽셀값을 출력한다. 하지만, (x,y,t)의 모든 위치에서 신경망을 트레이닝하는 것은 연산 복잡도를 크게 증가시킬 수 있으므로, 도 6의 예시와 같이 NeRV 모델은 시간 인덱스 t만을 입력으로 하는 신경망 구조를 이용하여 시간 인덱스 t에서 전체 프레임의 RGB 이미지를 출력한다.

[0049] 도 7a 및 도 7b는 본 개시의 일 실시예에 따른, NeRV 모델의 개념을 나타내는 예시도이다.

[0050] 시간 인덱스 t 입력에 대해 이미지 출력을 용이하게 제공하기 위해, 기존 암시적 표현 모델의 MLP 네트워크를 이용하는 것보다 콘볼루션 레이어를 이용하는 것이 효과적일 수 있다. NeRV 모델은 다수의 콘볼루션 레이어들로 구성된 NeRV 블록이 적층된 구조를 이용하여 출력을 생성할 수 있다. 도 7a의 예시와 같이, MLP 기반 출력 및 NeRV 블록 기반 출력이 비교될 수 있다. NeRV 블록은 도 7b의 예시와 같이, 콘볼루션을 수행하는 콘볼루션 레이어들, 픽셀 셔플(pixel shuffle) 레이어, 및 활성화(activation) 레이어를 포함한다. 도 7b의 예시에서, C는 채널의 개수를 나타내고, W, H는 NeRV 블록의 입력의 너비, 높이를 나타내며, S는 업스케일링 인자를 나타낸다.

[0051] 또한, NeRV 모델은 수학식 1과 같이 시간 인덱스 t를 고차원 공간에 임베딩(embedding)한 후, 임베딩된 인덱스를 입력으로 사용한다.

수학식 1

[0052] $$\gamma(t) = (\sin(2^0\pi t), \cos(2^0\pi t), \ldots, \sin(2^{L-1}\pi t), \cos(2^{L-1}\pi t))$$

[0053] 수학식 3의 예시와 같이, 시간 인덱스 t는 2L 차원의 벡터 $\gamma(t)$로 매핑될 수 있다. 임베딩된 시간 인덱스 t를 NeRV 모델의 트레이닝에 이용함으로써, NeRV 모델은 고주파 변이를 포함하는 비디오 데이터를 더 잘 예측할 수 있다.

[0054] NeRV 모델의 손실함수로서, L1 손실과 SSIM(Structural Similarity Index) 손실의 조합이 사용될 수 있다. NeRV 모델의 트레이닝 시, 입력에 기초하여 NeRV 모델에 의해 추정된 이미지, 및 GT(Ground-Truth) 이미지(즉, 원본 이미지)의 모든 픽셀 위치들에서, 전술한 바와 같은 손실이 산정될 수 있다. L1 손실은 추정된 이미지와 GT 이미지에서 픽셀들 간 차이의 절대값을 이용하여 산정된다. SSIM 손실은 추정된 이미지와 GT 이미지에서 픽셀들의 평균, 표준편차 및 상관도(correlation)에 기초하여 산정된다. 이후, 산정된 손실을 기반으로 트레이닝 과정에서 NeRV 모델을 구성하는 가중치들(또는 파라미터들)이 업데이트될 수 있다.

[0055] 도 8은 본 개시의 일 실시예에 따른, NeRV 모델의 부호화 파이프라인을 나타내는 예시도이다.

[0056] NeRV 모델을 이용하여 원본 비디오를 근사적으로 표현할 수 있으므로, NeRV 모델을 압축하여 전송함으로써, 신경망을 이용한 종단간(end-to-end) 압축 기술이 구현될 수 있다. NeRV 모델을 부호화하기 위한 부호화 파이프라인은 도 8의 예시와 같이, 비디오 오버피터(video overfitter, 910), 모델 프루너(model pruner, 920), 모델

양자화부(model quantizer, 930), 및 가중치 부호화부(weight encoder, 940)의 전부 또는 일부를 포함할 수 있다.

[0057] 비디오 오버피터(910)는 입력된 비디오 프레임을 NeRV 모델로 표현한다. 이때, 영상 부호화 장치는 전술한 손실 함수를 이용하여 NeRV 모델을 트레이닝함으로써, NeRV 모델로 하여금 비디오 프레임을 표현하도록 한다. 모델 프루너(920)은 MLP 또는 MLP와 콘볼루션 레이어들이 혼합된 형태를 갖는 NeRV 모델 구조를 프루닝을 이용하여 단순화한다. 예컨대, 프루닝은 기설정된 임계치보다 작은 가중치를 영으로 설정한다. 모델 양자화부(930)은 프루닝이 적용된 가중치들(예컨대, 임계치 이상의 가중치들)을 양자화한다. 가중치 부호화부(940)은 양자화된 가중치들에 엔트로피 부호화를 적용하여 가중치들의 비트스트림을 생성한다.

[0058] 한편, 도 8에 예시된 부호화 파이프라인의 역순으로서, NeRV 모델을 복호화하기 위한 복호화 파이프라인은 도 9와 같이 예시될 수 있다. 복호화 파이프라인은 도 9의 예시와 같이, 가중치 복호화부(weight decoder, 1010), 모델 역양자화부(model dequantizer, 1020), 모델 복원부(model reconstructor, 1030), 및 비디오 생성부(video generator, 1040)의 전부 또는 일부를 포함할 수 있다.

[0059] 가중치 복호화부(1010)는 NeRV 모델의 가중치들의 비트스트림에 엔트로피 부호화를 적용하여 양자화된 가중치들을 생성한다. 모델 역양자화부(1020)는 양자화된 가중치들을 역양자화하여 프루닝이 적용된 가중치들을 생성한다. 모델 복원부(1030)는 프루닝이 적용된 가중치들에 기초하여 NeRV 모델을 복원한다. 비디오 생성부(1040)는 복원된 NeRV 모델을 이용하여 시간 인덱스에 따라 복원 비디오 프레임을 생성한다.

[0060] NeRV 모델은 전술한 바와 같이, 프레임의 시간 인덱스 t를 입력으로 이용한다. 따라서, t×h×w 크기의 픽셀들을 갖는 비디오가 입력되는 경우, NeRV 모델은 t×h×w 번이 아니라 t번만 입력 비디오를 샘플링하므로, 인코딩 속도 및 디코딩 속도 측면에서 모두 큰 이득이 기대될 수 있다.

[0061] 한편, 전술한 바와 같은 부호화 파이프라인 및 복호화 파이프라인은 INVR 모델의 압축 및 복원에도 이용될 수 있다.

[0062] 이하, 암시적 신경망 비디오 표현 모델, INVR 모델 및 2D INVR 모델은 호환적으로 사용된다.

[0063] **II. 본 개시에 따른 실시예들**

[0064] 도 1에 예시된 영상 부호화 장치는 2D 비디오를 이용하여 부호화기(110)와 복호화기(120)를 포함하는 2D INVR 모델을 트레이닝함으로써, 2D INVR 모델로 하여금 2D 비디오를 생성할 수 있도록 한다. 이때, 2D INVR 모델의 입력 및 GT(Ground Truth)로는 시간 t에서의 비디오 프레임이 사용된다. 즉, 부호화기(110)와 복호화기(120)는 오토인코더를 구성하고, 종단간으로 트레이닝될 수 있다. 부호화기(110)는 비디오 프레임의 저주파수 성분이 압축된 임베딩을 생성하고, 복호화기(120)는 임베딩에 기초하여 비디오 프레임을 생성하도록 트레이닝될 수 있다. 트레이닝부는 2D INVR 모델의 출력과 GT 간 차이를 이용하여 2D INVR 모델의 파라미터들(가중치들)을 업데이트함으로써, INVR 모델로 하여금 2D 비디오를 생성하는 것을 학습하도록 한다. 학습이 완료된 2D INVR 모델은 모델 내 파라미터들에 기초하여 2D 비디오가 포함하는 2D 평면 정보를 포함할 수 있다.

[0065] 신경망 압축기(130)는 트레이닝된 2D INVR 모델의 파라미터들 및 임베딩을 압축하여 비트스트림을 생성한다. 예컨대, 도 8에 예시된 부호화 파이프라인을 이용하여 2D INVR 모델이 압축될 수 있다. 영상 부호화 장치는 생성된 비트스트림을 영상 복호화 장치로 전송한다.

[0066] 일 예로서, 영상 부호화 장치는 부가 정보로서 INVR 모델의 압축에 관련된 양자화 파라미터를 부호화할 수 있다. 영상 부호화 장치는 부가 정보로서 INVR 모델의 구조에 관한 정보를 부호화할 수 있다.

[0067] 도 2에 예시된 영상 복호화 장치는 신경망 역압축기(210)를 이용하여 비트스트림으로부터 2D INVR 모델의 파라미터들 및 임베딩을 복호화한다. 예컨대, 도 9에 예시된 복호화 파이프라인을 이용하여 2D INVR 모델의 복호화기가 복원될 수 있다.

[0068] 일 예로서, 영상 복호화 장치는 부가 정보로서 INVR 모델의 압축에 관련된 양자화 파라미터를 복호화할 수 있다. 영상 복호화 장치는 부가 정보로서 INVR 모델의 구조에 관한 정보를 복호화할 수 있다.

[0069] 복호화기(220)는 복원된 2D INVR 모델의 디코더를 이용하여 임베딩으로부터 2D 비디오를 생성한다. 영상 복호화 장치는 임베딩을 복호화한 후, 복호화된 임베딩을 2D INVR 모델에 입력하여 복원 2D 비디오를 생성한다.

[0070] 영상 부호화 장치 내 복호화기(120)와 영상 복호화 장치 내 복호화기(220)는 동일한 구조를 갖는다. 다만, 트레이닝부에 의해 영상 부호화 장치 내 복호화기(120)는 부호화기(110)와 함께 종단간으로 트레이닝되고, 영상 복

호화 장치 내 복호화기(220)는 트레이닝된 파라미터들에 기초하여 동작한다. 이하, 영상 부호화 장치 내 부호화기(110)와 복호화기(120)를 중심으로 기술한다. 영상 복호화 장치 내 복호화기(220)의 구조 및 동작은 영상 부호화 장치 내 복호화기(120)에 대한 설명으로 대체될 수 있다.

[0071] 일 예로서, 공간 정보 기반 INVR 모델을 이용하는 영상 부호화 장치를 기술한다.

[0072] 도 10a 및 도 10b는 본 개시의 일 실시예에 따른, 공간 정보 기반 INVR 모델을 나타내는 예시도이다.

[0073] INVR 모델은, 스펙트럼 편향의 완화에 따라 비디오의 암시적 표현의 학습을 향상시키기 위해, 2D 이산 웨이블렛 변환(Discrete Wavelet Transform, DWT)을 활용하여 입력 비디오를 저주파 및 고주파 성분으로 분할한다. INVR 모델의 부호화기(110)는 프레임 전반에서 변화가 작은 저주파 성분을 사용하여 간결(compact)한 임베딩을 형성한다. 임베딩은 입력된 비디오 프레임에 대한 공간 정보를 포함한다. INVR 모델의 복호기화(120)는, 부호화기(110)에 의해 전달된 임베딩에 기초하여 저주파 성분을 복원하고, 복원된 저주파 성분을 활용하여 고주파 성분을 복원한다.

[0074] 비디오 $\mathbf{V}=[V_1, V_2, \ldots, V_T]$에서 임의의 시간 t에서의 비디오 프레임인 V_t가 부호화기(110)의 입력으로 제공된다. 이때, $1 \leq t \leq T$이고, $V_t \in \mathbf{V}$이다. 부호화기(110)는 비디오 V_t를 변환하여 주파수 도메인 상의 신호 w_t를 생성하고, INVR 기반 압축을 수행할 수 있다. 도 10a의 예시와 같이, 부호화기(110)는 2D DWT 모듈, 및 CNN 기반 다운샘플링 블록(Downsampling Block, DB)를 포함할 수 있다.

[0075] 변환 후에도 공간 정보를 활용하기 위해, 부호화기(110)는 비디오 프레임에 2D DWT을 적용하여 다수의 하위 대역 웨이블렛 성분으로 변환할 수 있다. 예를 들어, Haar Wavelet 변환을 적용하여, 부호화기(110)는 비디오 프레임 V_t를 $w_t=[w_t(LL), w_t(LH), w_t(HL), w_t(HH)]$로 변환한다. 여기서, $w_t(LL)$, $w_t(LH)$, $w_t(HL)$, $w_t(HH)$는 Haar Wavelet 변환을 수직 방향 및 수평 방향으로 저주파, 고주파 기저함수를 이용하여 생성된 주파수 도메인 신호이다. 예컨대, $w_t(LL)$는 수직 및 수평 방향 저주파 기저함수의 적용에 따라 생성된 저주파 신호이다.

[0076] INVR 모델의 화질 정보를 향상시키기 위해, 부호화기(110)는 w_t의 저주파 성분만을 임베딩할 수 있다. 부호화기(110)는 w_t 중 $w_t(LL)$만을 DB에 입력하고, 다운샘플링에 기초하는 공간 적응적 임베딩을 생성할 수 있다. 부호화기(110)는 생성된 임베딩을 복호화기(120)로 전달한다.

[0077] 복호화기(120)는 임베딩에 기초하여 주파수 도메인 신호 w_t를 복원하고 복원된 w_{t_hat}를 역변환하여 복원 프레임 V_{t_hat}을 생성한다. 도 10a의 예시와 같이, 복호화기(120)는 CNN 기반 업샘플링 블록(Upsampling Block, UB), 다중해상도 융합기(Multi-resolution Fusion Unit, MFU), 고주파 복원기(High frequency restorer (HFR), 및 2D 역이산 웨이블렛 변환(Inverse DWT, IDWT) 모듈의 전부 또는 일부를 포함할 수 있다.

[0078] 복호화기(120)는 다수의 UB을 이용하여 임베딩으로부터 저주파 성분을 복원한다. 다수의 UB는, 2D 콘볼루션과 픽셀셔플에 기초하여 임베딩의 공간적 크기를 원본 크기로 순차적으로 업샘플링함으로써, 저주파 성분을 원본 차원으로 복원한다.

[0079] 업샘플링만으로는 복원 품질을 향상시키는데 한계가 있으므로, 복호화기(120)는 coarse-to-fine 구조의 MFU를 이용하여 복원된 저주파 성분의 품질을 향상시킨다. UB에 의해 생성되는 다양한 해상도의 저주파 성분(예를 들어, 도 10a에서 S1, S2, S3)을 활용하여, MFU는 복원된 저주파 성분의 품질을 향상시킬 수 있다. 품질이 향상된 복원 저주파 성분을 $w_{t_hat}(LL)$로 표현한다.

[0080] 도 10b의 예시와 같이, MFU는 적어도 하나 이상의 다중해상도 융합블록(Multi-resolution Fusion Block, MFB)을 포함하고, 각 MFB는 전치 콘볼루션(Transposed Convolution, CT) 블록과 잔차블록(Residual Block, RB)들을 포함한다. MFB는, 제1 해상도의 저주파 성분(예컨대, 도 10b에서 S1)에 전치 콘볼루션을 적용하여 제2 해상도의 제1 저주파 성분을 생성한다. 제1 저주파 성분의 제2 해상도는 제2 저주파 성분(예컨대, 도 10b에서 S2)의 해상도와 동일할 수 있다. MFB는, 전치 콘볼루션이 적용된 제1 저주파 성분과 제2 저주파 성분을 결합(concatenation)하고, 결합된 저주파 성분을 잔차블록에 전달한다. RB는 도 10b의 예시와 같이, 두 개의 2D 콘볼루션(Conv2D) 블록 및 LReLU(Leaky Rectified Linear Unit) 활성화(activation) 블록을 포함한다. 또한, RB는 입출력 간에 스킵 경로를 포함한다. RB는 결합된 저주파 성분의 잔차를 생성하고, 생성된 잔차와 결합된 저주파 성분을 가산한다. MFU는 MFB를 반복적으로 적용함으로써, 다양한 해상도의 저주파 성분의 융합에 기초하여 복원된 저주파 성분의 품질을 향상시킬 수 있다.

[0081] HFR은 복원 저주파 성분 $w_{t_hat}(LL)$를 이용하여 나머지 대역 성분인 $w_t(LH)$, $w_t(HL)$, $w_t(HH)$을 복원함으로써, w_{t_hat}를 구성할 수 있다. HFR은 도 10b의 예시와 같이, 두 개의 Conv2D 블록 및 LReLU 활성화 블록을 포함한다. HFR은 저주파 성분에 2D 콘볼루션을 적용하여 나머지 주파수 성분을 복원할 수 있다.

[0082] 복호화기(120)는 복원 주파수 성분들을 2D IDWT 모듈에 적용하여 역이산 웨이블렛 변환함으로써, 복원 비디오 프레임 V_{t_hat}를 출력한다. 고주파 성분의 학습에 기초하여 고주파 성분을 복원하는 대신, 복호화기(120)가 HFR에 기초하여 고주파 성분을 합성함으로써, 트레이닝 과정에 사용되는 파라미터의 개수가 감소될 수 있다.

[0083] 전술한 바와 같이, 트레이닝부는 부호화기(110) 및 복호화기(120)에 포함된 딥러닝 기반 구성요소들(DB, UB, MFU, HFR 등)을 종단간으로 트레이닝시킬 수 있다. 예컨대, 손실함수는 복호화기(120)의 복원 프레임과 GT 간 차이에 기반하여 정의될 수 있다. 트레이닝부는 손실함수를 감소시키는 방향으로, INVR 모델에 포함된 부호화기(110) 및 복호화기(120)의 파라미터들을 업데이트할 수 있다.

[0084] 전술한 바와 같이, 영상 복호화 장치 내 복호화기(220)는 영상 부호화 장치 내 복호화기(120)와 동일한 구조를 갖고, 트레이닝된 파라미터들에 기초하여 영상 부호화 장치 내 복호화기(120)와 동일하게 동작한다. 따라서, 영상 복호화 장치 내 복호화기(220)의 구조 및 동작에 대한 기술을 생략한다.

[0085] INVR 기반 압축에 따라 생성되는 전체 비트들은, 임베딩 벡터와 복호화기(120)의 네트워크 정보의 크기이다. 임베딩 벡터의 크기는, $w_t=[w_t(LL), w_t(LH), w_t(HL), w_t(HH)]$에서 전송할 저주파 대역을 결정하는 것에 의존한다. 또한, 복호화기(120)의 네트워크 정보의 크기는 네트워크 파라미터 및 네트워크 구조 정보를 포함하는데, 네트워크의 복원력과 비례 관계일 수 있다. 따라서, 원본 비디오 V_t와 복원 비디오 V_{t_hat} 간의 왜곡인 D_t 및 전체 비트 R_t의 결합에 기초하는 율왜곡 최적화 측면에서, 전송할 저주파 대역이 선택될 수 있다. 예를 들어, 수학식 2에 나타낸 J_t를 최소로 하는 저주파 대역이 선택될 수 있다.

수학식 2

[0086] $$J_t = D_t + \lambda R_t$$

[0087] 수학식 2에서, λ는 라그랑지 승수(Lagrangian multiplier)이다.

[0088] 전술한 바와 같이, 공간 정보 기반 INVR 모델은 각 비디오 프레임을 별도로 학습한다. 시간 정보도 반영하기 위한 다른 예로서, 시공간 정보 기반 INVR 모델을 이용하는 영상 부호화 장치를 기술한다.

[0089] 도 11은 본 개시의 일 실시예에 따른, 시공간 정보 기반 INVR 모델을 나타내는 예시도이다.

[0090] 시공간 정보 기반 INVR 모델은, 현재 프레임의 정보에 대한 임베딩을 활용할 뿐만 아니라, 인접한 프레임들 간의 차이에 대한 정보를 추가적으로 입력으로 활용한다. 시공간 정보 기반 INVR 모델은 인접한 프레임의 차이에 대한 정보를 활용하기 위해, 공간 정보 기반 INVR 모델과 동일한 구성요소들 외에 2D DWT 모듈들 및 1D DWT 모듈을 추가로 포함한다.

[0091] 부호화기(110)는 t 번째 프레임 및 인접한 t-1 번째, t+1 번째 프레임들을 활용하여 다음과 같이 3D 이산 웨이블렛 변환을 수행한다. 부호화기(110)는 2D 이산 웨이블렛 변환을 이용하여 공간 축에 대해 저주파 및 고주파 성분을 분해하고, 2D 변환 결과에 대한 추가적인 1D 이산 웨이블렛 변환을 이용하여 시간 축에 대해 저주파 및 고주파 성분을 분해한다. t 번째 프레임에 대한 임베딩(도 11에서, embedding (t)) 외에, 부호화기(110)는 생성된 시공간에 대한 저주파 성분을 DB를 이용하여 다운샘플링함으로써, 추가적인 임베딩 두 개(도 11에서, embedding (t-1), embedding (t+1))를 생성할 수 있다. 부호화기(110)는 임베딩들을 복호화기(120)로 전달한다. 복호화기(120)는 임베딩들을 주파수 도메인 신호를 복원하고, 전술한 바와 같이, 복원된 w_{t_hat}를 역변환하여 시간 t에서의 복원 프레임 V_{t_hat}을 생성한다.

[0092] 2D INVR 모델은 복원하고자 하는 영상을 표현할 수 있으므로, 해당 모델을 전송하는 것이 영상을 전송하는 것과 동일한 효과를 가진다. 즉, 2D INVR 모델을 압축하여 전송함으로써, 신경망을 이용한 종단간(end-to-end) 압축 기술이 구현될 수 있다. 이하, 영상을 생성하는 복호화기(120)의 파라미터 및 출력 임베딩을 부호화함으로써,

영상을 압축하는 방법을 기술한다.

[0093] 도 12는 본 개시의 일 실시예에 따른, 2D INVR 모델의 부호화 파이프라인을 나타내는 예시도이다.

[0094] 2D INVR 모델의 부호화 파이프라인은 도 12의 예시와 같이, 비디오 오버피터(910), 모델 프루너(920), 모델 양자화부(930), 임베딩 양자화부(embedding quantizer, 1210) 및 파라미터 부호화부(parameter encoder, 1220)의 전부 또는 일부를 포함할 수 있다.

[0095] 비디오 오버피터(910)는 2D INVR 모델의 파라미터들에 기초하여 복원 영상을 표현한다. 영상 부호화 장치는 전술한 손실함수를 이용하여 INVR 모델을 트레이닝함으로써, INVR 모델로 하여금 비디오 프레임을 표현하도록 한다. 손실함수로서, 복원 프레임과 GT 간 차이에 기반하여 L1 손실과 SSIM 손실의 조합이 사용될 수 있다. 복원 프레임과 GT 간 차이에 기반하는 손실 외에, 복원된 2D 웨이블렛 계수들과 관련된 손실을 추가적으로 산정함으로써, 고주파 성분이 더 잘 표현될 수 있다.

[0096] 모델 프루너(920)는 프루닝을 이용하여 중요도가 낮은 파라미터들을 제거함으로써, INVR 모델의 모델을 경량화한다. 일 예로서, 모델의 전역 범위에 대하여 가장 작은 연결을 제거하는 global unstructured pruning 기법이 사용될 수 있다. 다른 예로서, 프루닝은 기설정된 임계치보다 작은 파라미터를 영으로 설정할 수 있다.

[0097] 모델 양자화부(930)은 프루닝이 적용된 파라미터들(예컨대, 임계치 이상의 파라미터들)을 양자화한다. 임베딩 양자화부(1210)는 부호화기(110)에 의해 생성되는 임베딩을 양자화한다. 경량화된 INVR 모델의 최종 파라미터들 및 임베딩은 각각 m 비트와 n 비트(m, n은 임의의 양의 정수)로 양자화될 수 있다.

[0098] 파라미터 부호화부(1220)는 양자화된 파라미터들 및 임베딩의 손실 없는 압축을 위해, 예컨대, 허프만 코딩(Huffman coding) 또는 엔트로피 코딩을 활용하여 비트스트림의 크기를 감소시킬 수 있다. 더 많은 파라미터를 프루닝함에 따라, 파라미터 부호화를 이용하여 더 적은 비트를 사용함으로써, 모델의 파라미터들이 효과적으로 압축될 수 있다.

[0099] 전술한 바와 같이, 영상 부호화 장치는 임베딩을 양자화하고, DWT에 기초하여 입력 공간적 크기를 감소시킴으로써, 영상 복호화 장치에서 요구되는 파라미터의 크기를 감소시킬 수 있다.

[0100] 일 예로서, 영상 부호화 장치는 부가 정보로서 INVR 모델의 압축에 관련된 양자화 파라미터를 부호화할 수 있다. 영상 부호화 장치는 부가 정보로서 INVR 모델의 구조에 관한 정보를 부호화할 수 있다. 일 예로서, 부가 정보를 전달하기 위해 IPS(INVR model Parameter Set)이 활용될 수 있다. 영상 부호화 장치는 IPS를 비트스트림 상에 포함시킬 수 있다.

[0101] 한편, 도 12에 예시된 부호화 파이프라인의 역순으로서, INVR 모델을 복호화하기 위한 복호화 파이프라인은 도 13과 같이 예시될 수 있다. 복호화 파이프라인은 도 10의 예시와 같이, 파라미터 복호화부(parameter decoder, 1310), 임베딩 역양자화부(1320, embedding dequantizer), 모델 역양자화부(1020), 모델 복원부(1030), 및 비디오 생성부(1040)의 전부 또는 일부를 포함할 수 있다.

[0102] 파라미터 복호화부(1310)는, 예컨대, 허프만 코딩 또는 엔트로피 코딩을 이용하여, 비트스트림으로부터 양자화된 가중치들 및 양자화된 임베딩을 생성한다. 임베딩 역양자화부(1320)는 양자화된 임베딩을 역양자화하여 복원 임베딩을 생성한다. 모델 역양자화부(1020)는 양자화된 가중치들을 역양자화하여 프루닝이 적용된 파라미터들을 생성한다. 모델 복원부(1030)는 프루닝이 적용된 파라미터들에 기초하여 INVR 모델을 복원한다. 비디오 생성부(1040)는 복원된 INVR 모델을 이용하여 복원 임베딩에 따라 복원 비디오 프레임을 생성한다.

[0103] 일 예로서, 영상 복호화 장치는 부가 정보로서 INVR 모델의 압축에 관련된 양자화 파라미터를 복호화할 수 있다. 영상 복호화 장치는 부가 정보로서 INVR 모델의 구조에 관한 정보를 복호화할 수 있다.

[0104] 일 예로서, 2D DWT의 사용 외에, 다른 1D/2D 변환(예를 들어, DCT(discrete cosine transform), DST(discrete sine transform, 비분리 변환(non-separable transform) 등)도 활용될 수 있다. 예컨대, 영상 부호화 장치는 임의의 변환에 기초하여 변환계수들을 생성한 후, 변환된 블록/픽처 내 특정 좌표값 (x_{th}, y_{th})를 사용하여 변환계수들을 저주파와 고주파 성분으로 구분할 수 있다. 변환계수의 좌표값이 $0 \le x \le x_{th}$, $0 \le y \le y_{th}$인 경우, 영상 부호화 장치는 해당되는 변환계수들을 저주파 성분으로 구분하고, 나머지 변환계수들을 고주파 성분으로 구분할 수 있다. 여기서, 특정 좌표값 x_{th}, y_{th}는 0 이상의 양의 정수이고, 영상 부호화 장치와 영상 복호화 장치 간 약속에 따라 미리 설정될 수 있다. 또는, 특정 좌표값은 율왜곡 최적화 측면에서 영상 부호화 장치에서 결정되고, 영상 복호화 장치에게 시그널링될 수 있다.

[0105] 다른 예로서, 생성된 변환계수들에 스캔(scan)이 수행되는 경우, 영상 부호화 장치는 기설정된 임계치에 기초하여 저주파와 고주파 성분을 구분할 수 있다. 고주파 성분부터 스캔되는 경우, 스캔에 따른 인덱스가 기설정된 임계치 이상인 경우, 영상 부호화 장치는 해당되는 변환계수들을 저주파 성분으로 구분하고, 나머지 변환계수들을 고주파 성분으로 구분할 수 있다. 저주파 성분부터 스캔되는 경우, 스캔에 따른 인덱스가 기설정된 임계치 이하인 경우, 영상 부호화 장치는 해당되는 변환계수들을 저주파 성분으로 구분하고, 나머지 변환계수들을 고주파 성분으로 구분할 수 있다. 여기서, 기설정된 임계치는 0 이상의 양의 정수이고, 영상 부호화 장치와 영상 복호화 장치 간 약속에 따라 미리 설정될 수 있다. 또는, 기설정된 임계치는 율왜곡 최적화 측면에서 영상 부호화 장치에서 결정되고, 영상 복호화 장치에게 시그널링될 수 있다.

[0106] 이하, 도 14 및 도 15의 도시를 이용하여, 2D INVR 모델에 기반하여 2D 비디오를 부호화하거나 복원하는 방법을 기술한다.

[0107] 도 14는 본 개시의 일 실시예에 따른, 영상 부호화 장치가 2D 비디오를 부호화하는 방법을 나타내는 순서도이다.

[0108] 영상 부호화 장치는 현재 2D 프레임을 2D INVR 모델의 인코더에 입력하여 압축된 임베딩을 생성한다(S1400).

[0109] 영상 부호화 장치는 다음과 같이 임베딩을 생성할 수 있다.

[0110] 영상 부호화 장치는 기설정된 변환방법에 기초하여 현재 2D 프레임을 변환하여 현재 2D 프레임의 저주파 성분 및 나머지 주파수 성분을 생성한다(S1410). 여기서, 기설정된 변환방법은 2D 이산 웨이블렛 변환, DCT, DST, 또는 비분리 변환을 포함할 수 있다.

[0111] 영상 부호화 장치는 저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 임베딩을 생성한다(S1412).

[0112] 영상 부호화 장치는 임베딩을 2D INVR 모델의 디코더에 입력하여 현재 2D 프레임을 복원한다(S1402).

[0113] 영상 부호화 장치는 다음과 같이 현재 2D 프레임을 복원할 수 있다.

[0114] 영상 부호화 장치는 임베딩을 업샘플링하여 다중 해상도의 저주파 성분들을 복원한다(S1420).

[0115] 영상 부호화 장치는 다중해상도의 저주파 성분들을 융합하여 현재 2D 프레임의 저주파 성분을 복원한다(S1422).

[0116] 영상 부호화 장치는 현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 현재 2D 프레임의 나머지 주파수 성분을 복원한다(S1424).

[0117] 영상 부호화 장치는 기설정된 변환방법에 기초하여 현재 2D 프레임의 저주파 성분 및 나머지 주파수 성분을 역변환한다(S1426).

[0118] 영상 부호화 장치는 2D INVR 모델의 인코더와 2D INVR 모델의 디코더를 손실함수에 기초하여 종단간으로 트레이닝할 수 있다. 여기서, 손실함수는 2D INVR 모델의 인코더의 입력 프레임과 2D INVR 모델의 디코더의 출력 프레임 간의 차이에 기초하여 정의될 수 있다.

[0119] 영상 부호화 장치는 임베딩 및 2D INVR 모델의 디코더 파라미터들을 부호화한다(S1404).

[0120] 추가적으로, 영상 부호화 장치는 양자화 파라미터에 기초하여 디코더 파라미터들 및 임베딩을 양자화할 수 있다. 영상 부호화 장치는 부가 정보로서 양자화 파라미터를 부호화할 수 있다.

[0121] 도 15는 본 개시의 일 실시예에 따른, 영상 복호화 장치가 2D 비디오를 복원하는 방법을 나타내는 순서도이다.

[0122] 영상 복호화 장치는 비트스트림으로부터 2D INVR 모델의 디코더 파라미터들 및 임베딩을 복호화한다(S1500). 여기서, 임베딩은 현재 2D 프레임의 저주파 성분의 압축된 표현이고, 2D INVR 모델의 인코더의 의해 생성되고 제공된다.

[0123] 추가적으로, 영상 복호화 장치는 비트스트림으로부터 부가 정보로서 양자화 파라미터를 복호화한다. 영상 복호화 장치는 양자화 파라미터를 이용하여 디코더 파라미터들 및 임베딩을 역양자화할 수 있다.

[0124] 영상 복호화 장치는 디코더 파라미터들을 이용하여 2D INVR 모델의 디코더를 복원한다(S1502).

[0125] 영상 복호화 장치는 임베딩을 2D INVR 모델의 디코더에 입력하여 현재 2D 프레임을 복원한다(S1504).

[0126] 영상 복호화 장치는 다음과 같이 현재 2D 프레임을 복원할 수 있다.

[0127] 영상 복호화 장치는 임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원한다(S1514).

[0128] 영상 복호화 장치는 다중해상도의 저주파 성분들을 융합하여 현재 2D 프레임의 저주파 성분을 복원한다(S1512).

[0129] 영상 복호화 장치는 현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 현재 2D 프레임의 나머지 주파수 성분을 복원한다(S1516).

[0130] 영상 복호화 장치는 기설정된 변환방법에 기초하여 현재 2D 프레임의 저주파 성분 및 나머지 주파수 성분을 역변환한다(S1518). 여기서, 기설정된 변환방법은 2D 이산 웨이블렛 변환, DCT, DST, 또는 비분리 변환을 포함할 수 있다.

[0131] 2D INVR 모델의 인코더와 2D INVR 모델의 디코더는 손실함수에 따라 종단간으로 사전에 트레이닝될 수 있다. 손실함수는 2D INVR 모델의 인코더의 입력 프레임과 2D INVR 모델의 디코더의 출력 프레임 간의 차이에 기초하여 정의될 수 있다.

[0132] 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.

[0133] 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.

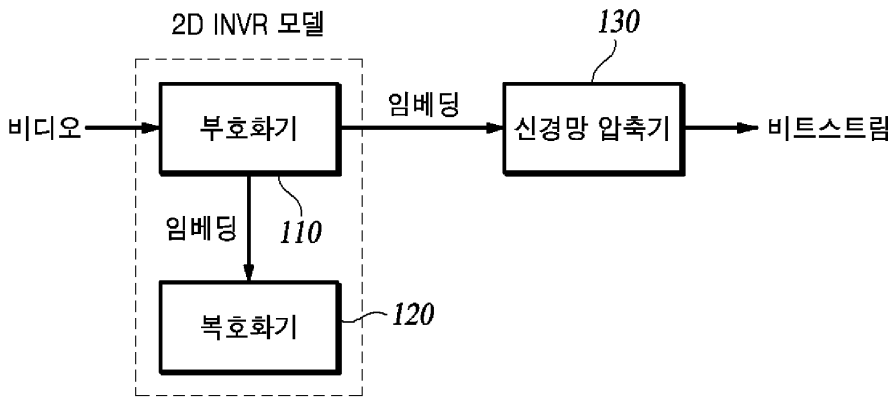
[0134] 한편, 본 실시예에서 설명된 다양한 기능들 혹은 방법들은 하나 이상의 프로세서에 의해 판독되고 실행될 수 있는 비일시적 기록매체에 저장된 명령어들로 구현될 수도 있다. 비일시적 기록매체는, 예를 들어, 컴퓨터 시스템에 의하여 판독가능한 형태로 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 예를 들어, 비일시적 기록매체는 EPROM(erasable programmable read only memory), 플래시 드라이브, 광학 드라이브, 자기 하드 드라이브, 솔리드 스테이트 드라이브(SSD)와 같은 저장매체를 포함한다.

[0135] 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.

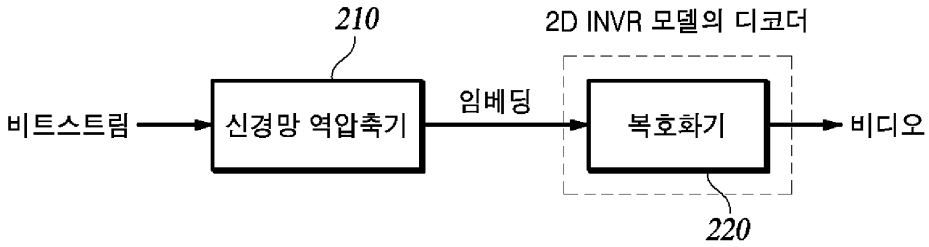
[0136]

도면

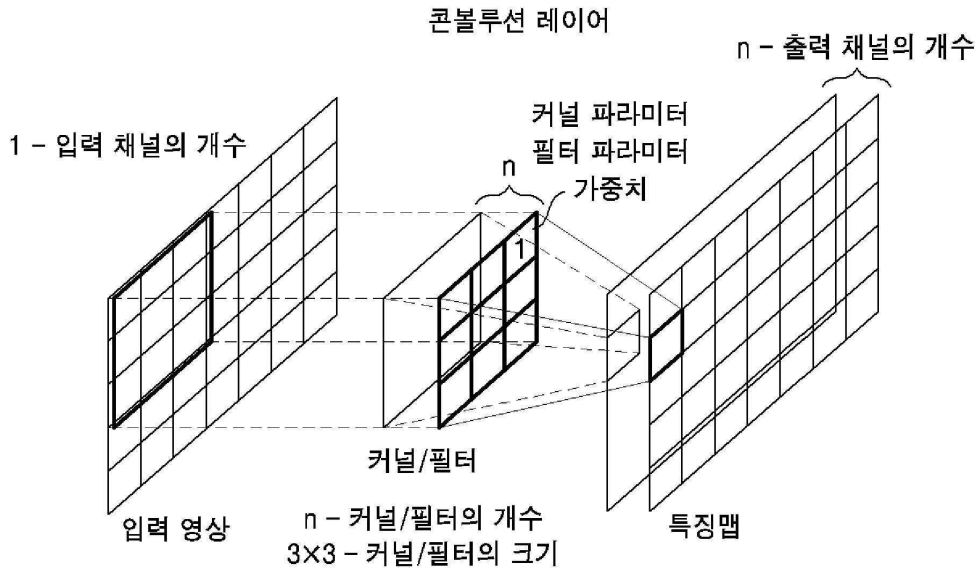
도면1



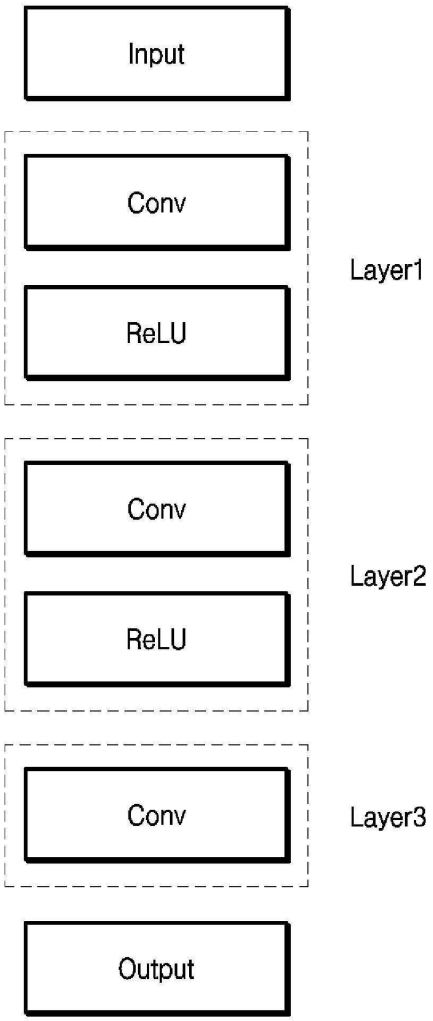
도면2



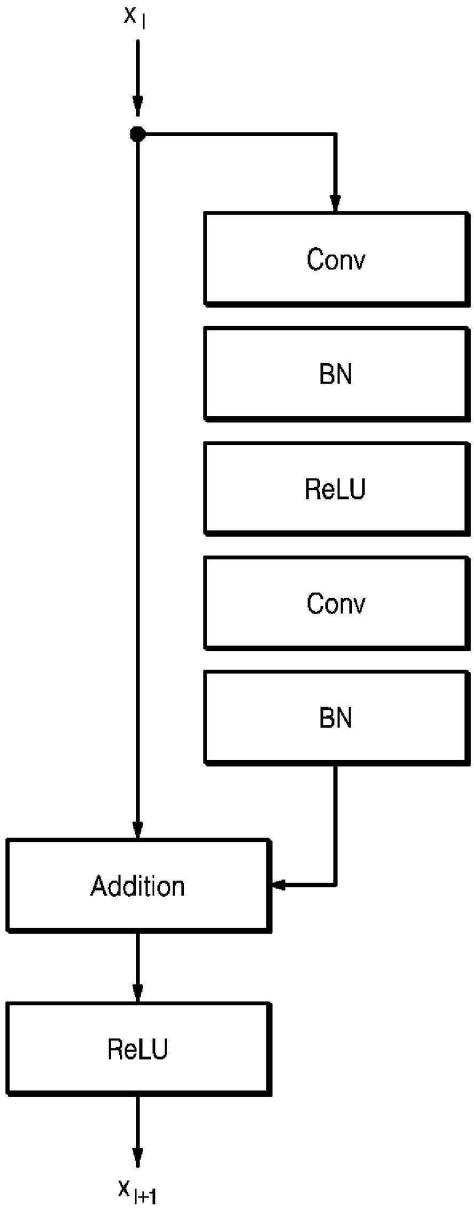
도면3



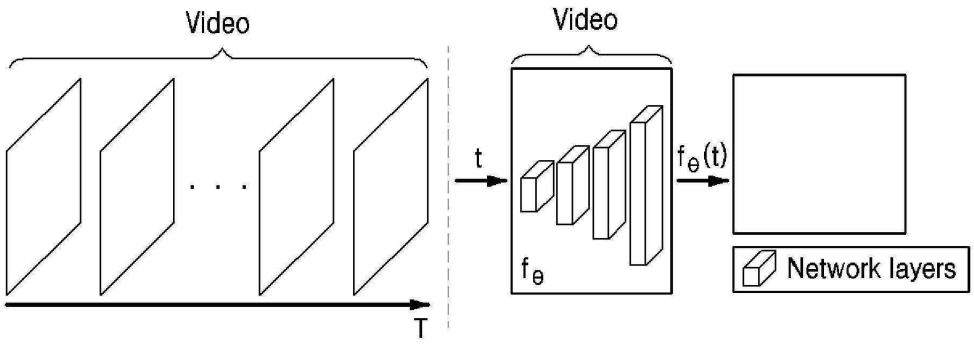
도면4



도면5

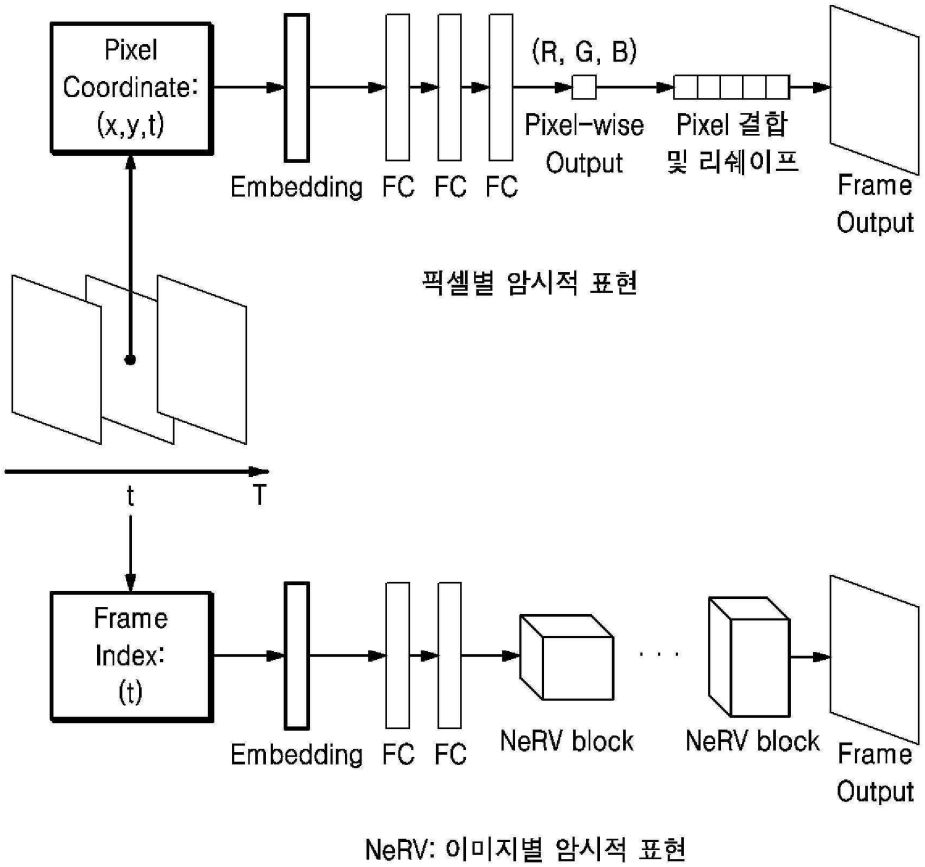


도면6

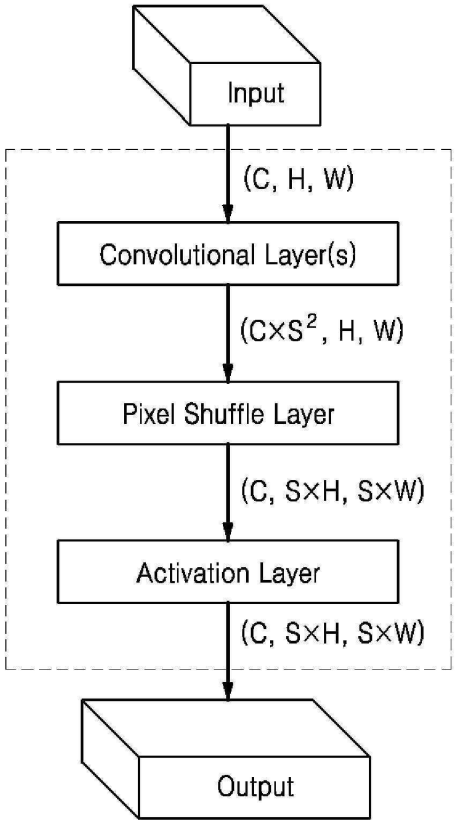


명시적 비디오 표현 암시적 신경망 표현(e.g., NeRV)

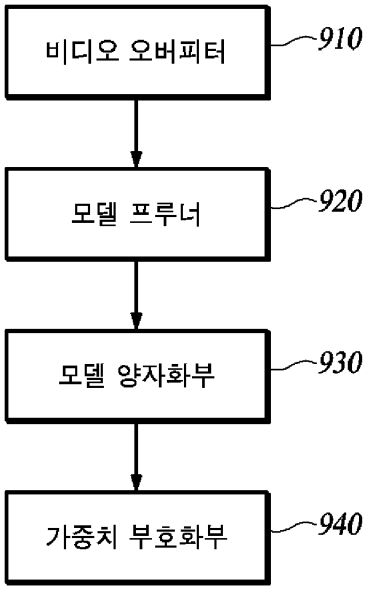
도면7a


Pixel
Coordinate:
(x,y,t)
Embedding
FC
FC
FC
(R, G, B)
Pixel-wise
Output
Pixel 결합
및 리쉐이프
Frame
Output
픽셀별 암시적 표현
t
T
Frame
Index:
(t)
Embedding
FC
FC
NeRV block
NeRV block
Frame
Output
NeRV: 이미지별 암시적 표현

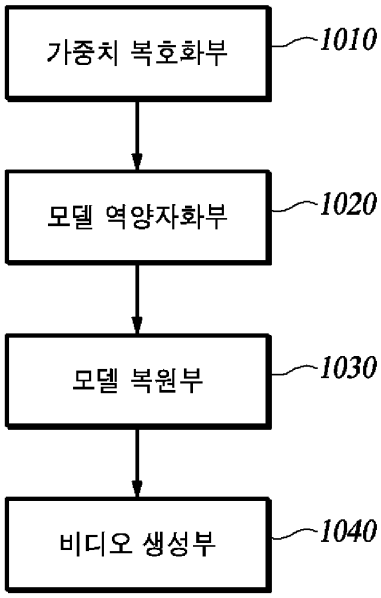
도면7b


Input
(C, H, W)
Convolutional Layer(s)
(C×S², H, W)
Pixel Shuffle Layer
(C, S×H, S×W)
Activation Layer
(C, S×H, S×W)
Output

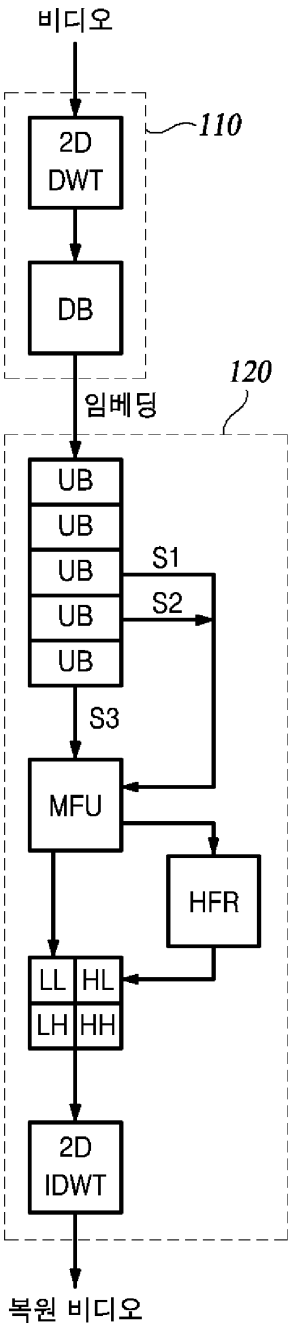
도면8


비디오 오버피터
910
모델 프루너
920
모델 양자화부
930
가중치 부호화부
940

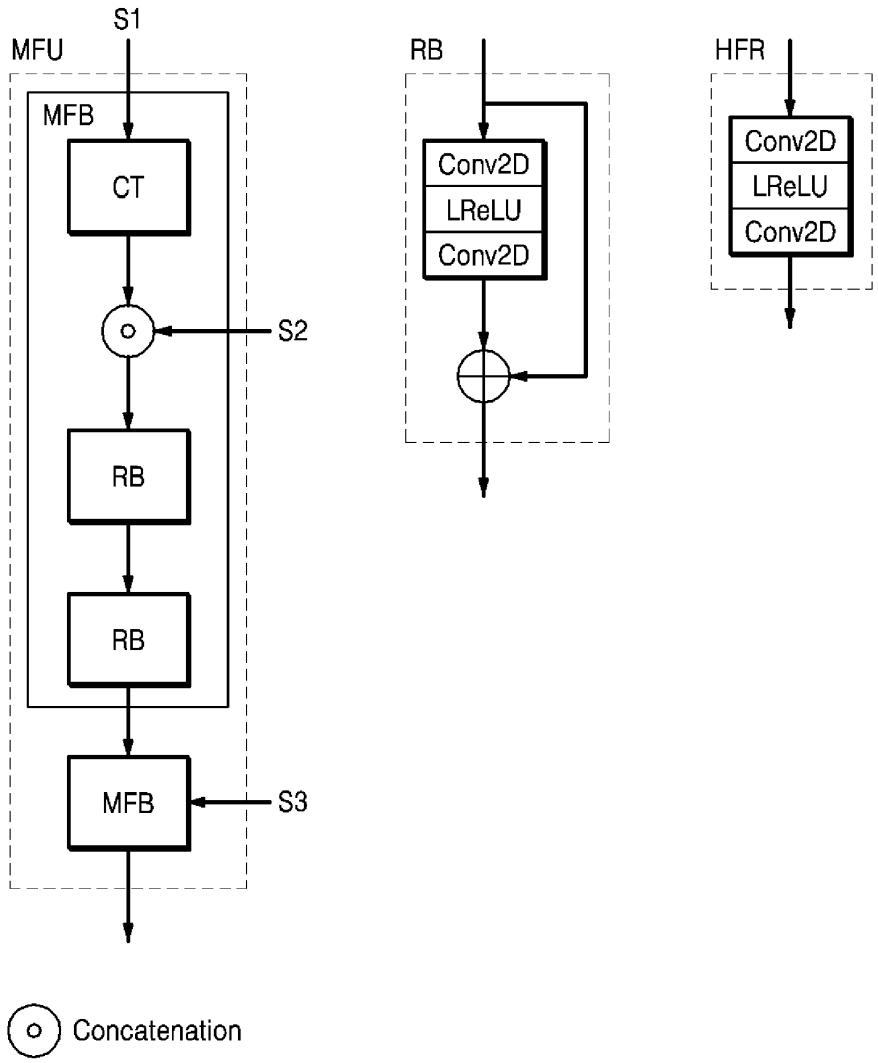
도면9



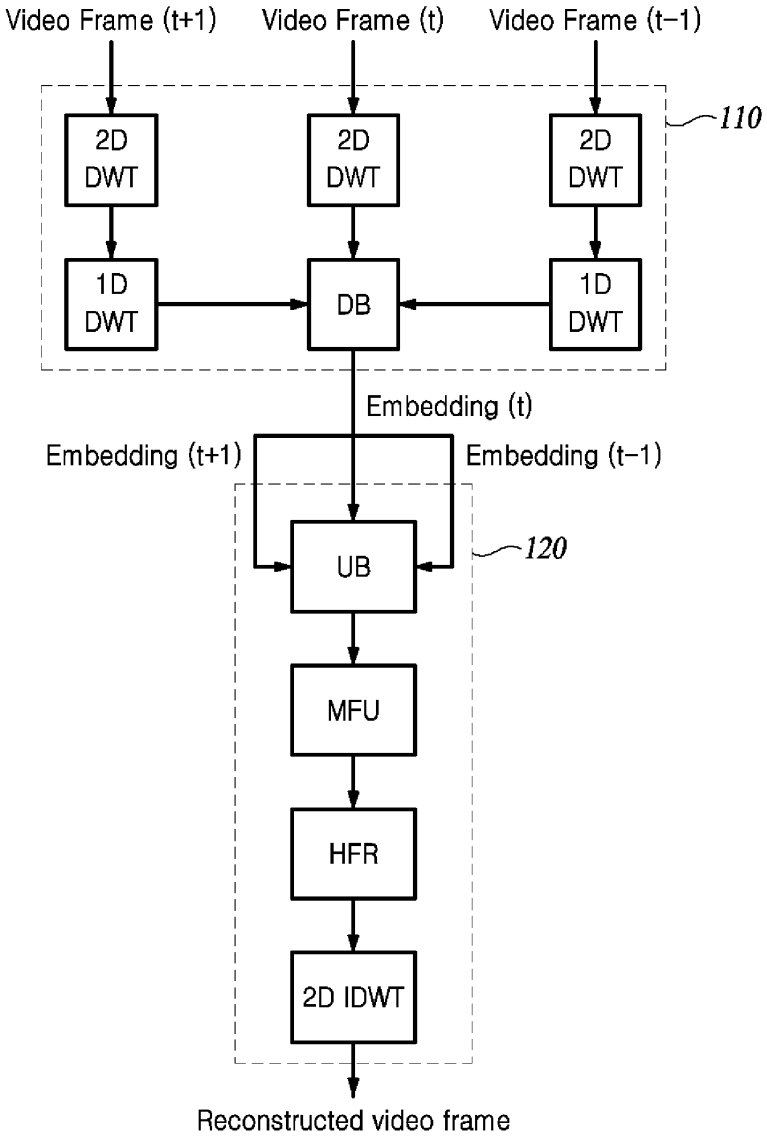
도면10a


비디오
110
2D
DWT
DB
120
임베딩
UB
UB
UB
UB
UB
S1
S2
S3
MFU
HFR
LL
HL
LH
HH
2D
IDWT
복원 비디오

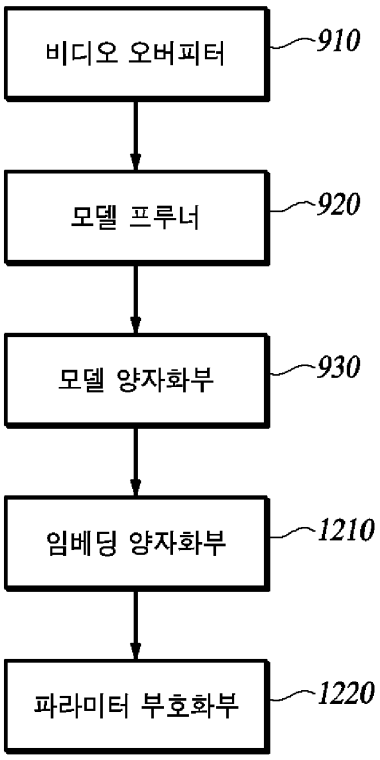
도면10b


S1
MFU
MFB
CT
S2
RB
RB
MFB
S3
RB
Conv2D
LReLU
Conv2D
HFR
Conv2D
LReLU
Conv2D
Concatenation

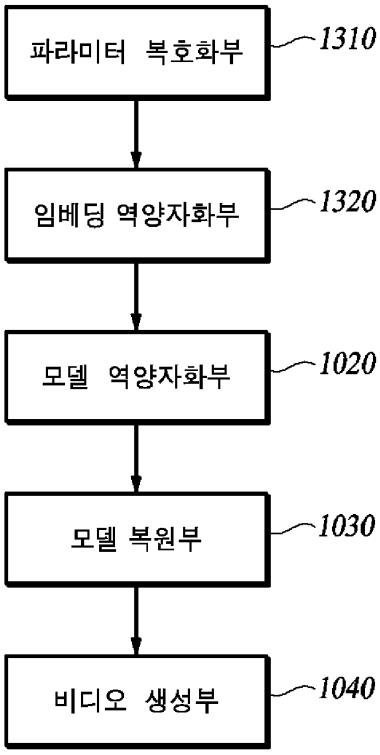
도면11


Video Frame (t+1)
Video Frame (t)
Video Frame (t−1)
2D DWT
2D DWT
2D DWT
110
1D DWT
DB
1D DWT
Embedding (t)
Embedding (t+1)
Embedding (t−1)
UB
120
MFU
HFR
2D IDWT
Reconstructed video frame

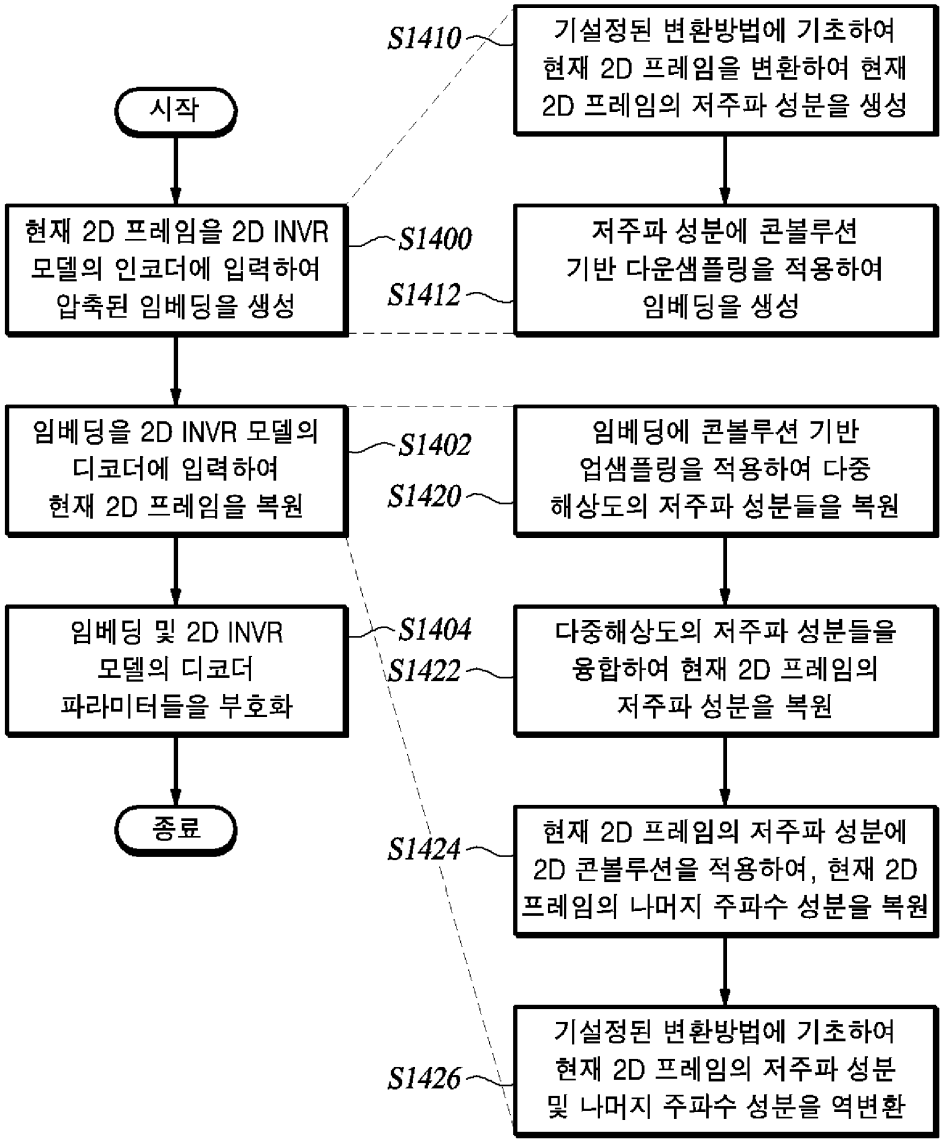
도면12


비디오 오버피터
910
모델 프루너
920
모델 양자화부
930
임베딩 양자화부
1210
파라미터 부호화부
1220

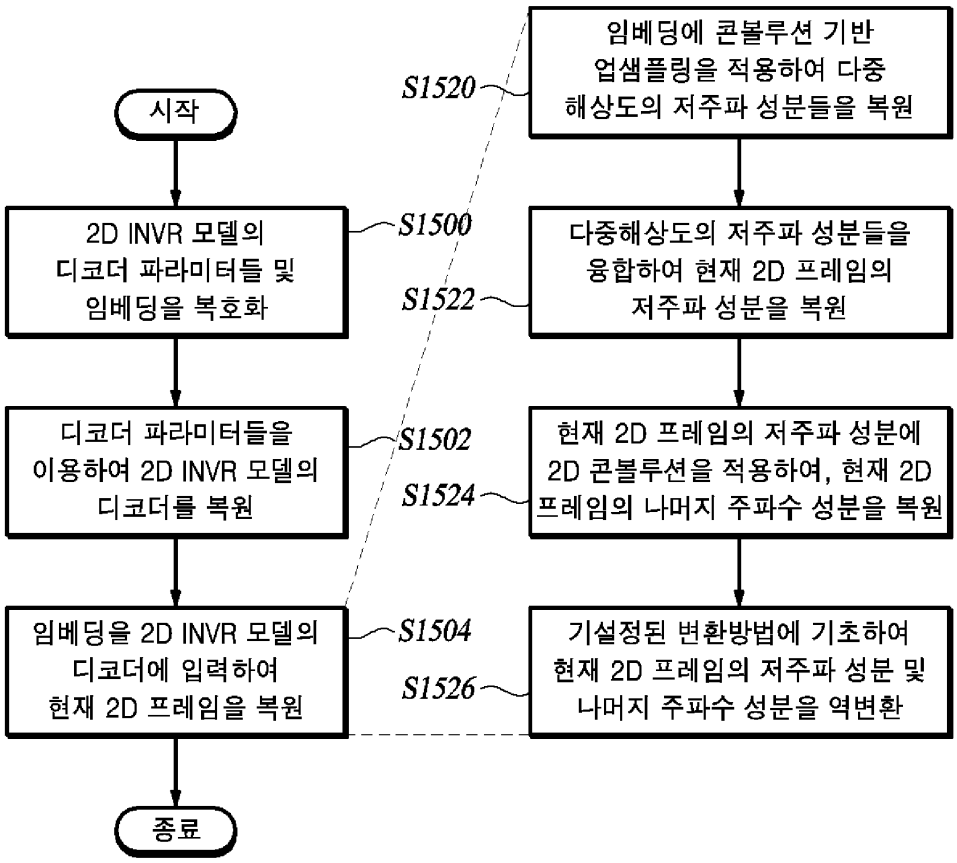
도면13


파라미터 복호화부
1310
임베딩 역양자화부
1320
모델 역양자화부
1020
모델 복원부
1030
비디오 생성부
1040

도면14


시작
현재 2D 프레임을 2D INVR 모델의 인코더에 입력하여 압축된 임베딩을 생성
S1400
S1410
기설정된 변환방법에 기초하여 현재 2D 프레임을 변환하여 현재 2D 프레임의 저주파 성분을 생성
S1412
저주파 성분에 콘볼루션 기반 다운샘플링을 적용하여 임베딩을 생성
임베딩을 2D INVR 모델의 디코더에 입력하여 현재 2D 프레임을 복원
S1402
S1420
임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원
임베딩 및 2D INVR 모델의 디코더 파라미터들을 부호화
S1404
S1422
다중해상도의 저주파 성분들을 융합하여 현재 2D 프레임의 저주파 성분을 복원
종료
S1424
현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 현재 2D 프레임의 나머지 주파수 성분을 복원
S1426
기설정된 변환방법에 기초하여 현재 2D 프레임의 저주파 성분 및 나머지 주파수 성분을 역변환

도면15


시작
2D INVR 모델의 디코더 파라미터들 및 임베딩을 복호화
S1500
디코더 파라미터들을 이용하여 2D INVR 모델의 디코더를 복원
S1502
임베딩을 2D INVR 모델의 디코더에 입력하여 현재 2D 프레임을 복원
S1504
종료
S1520
임베딩에 콘볼루션 기반 업샘플링을 적용하여 다중 해상도의 저주파 성분들을 복원
S1522
다중해상도의 저주파 성분들을 융합하여 현재 2D 프레임의 저주파 성분을 복원
S1524
현재 2D 프레임의 저주파 성분에 2D 콘볼루션을 적용하여, 현재 2D 프레임의 나머지 주파수 성분을 복원
S1526
기설정된 변환방법에 기초하여 현재 2D 프레임의 저주파 성분 및 나머지 주파수 성분을 역변환