



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0150298
(43) 공개일자 2022년11월10일

(51) 국제특허분류(Int. Cl.)
H04N 19/463 (2014.01) G06N 3/04 (2006.01)
G06N 3/08 (2006.01) H04N 19/137 (2014.01)
H04N 19/172 (2014.01) H04N 19/42 (2014.01)
H04N 19/503 (2014.01) H04N 19/85 (2014.01)
(52) CPC특허분류
H04N 19/463 (2015.01)
G06N 3/0445 (2013.01)
(21) 출원번호 10-2022-7029923
(22) 출원일자(국제) 2021년01월15일
심사청구일자 없음
(85) 번역문제출일자 2022년08월29일
(86) 국제출원번호 PCT/US2021/013599
(87) 국제공개번호 WO 2021/178050
국제공개일자 2021년09월10일
(30) 우선권주장
62/984,673 2020년03월03일 미국(US)
17/091,570 2020년11월06일 미국(US)

(71) 출원인
켈컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
(72) 발명자
콜린스키 아담 발데마르
미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775
양 양
미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775
(74) 대리인
특허법인코리아나
(뒷면에 계속)

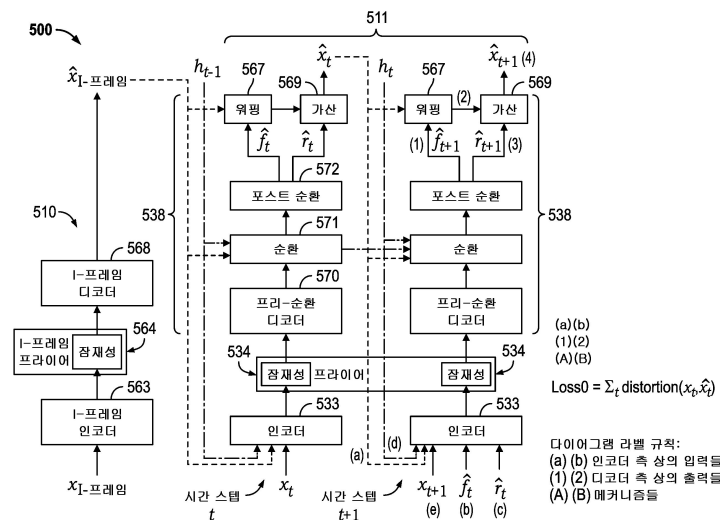
전체 청구항 수 : 총 64 항

(54) 발명의 명칭 순환-기반 머신 러닝 시스템들을 사용한 비디오 압축

(57) 요약

순환-기반 머신 러닝 툴들을 사용하여 비디오 콘텐츠를 코딩하기 위한 기법들이 본 명세서에서 설명된다. 디바이스는 인코더 및 디코더 부분들을 포함하는 뉴럴 네트워크 시스템을 포함할 수 있다. 인코더 부분은 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터에 기초하여 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성할 수 있다. 뉴럴 네트워크 시스템의 디코더 부분은 출력 데이터 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성할 수 있다.

대표도



(52) CPC특허분류

G06N 3/0454 (2013.01)

G06N 3/084 (2013.01)

H04N 19/137 (2015.01)

H04N 19/172 (2015.01)

H04N 19/42 (2015.01)

H04N 19/503 (2015.01)

H04N 19/85 (2015.01)

(72) 발명자

푸레자 레자

미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775

소피에르 기욤 콘라드

미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775

판 로전달 티스 예한

미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775

코헨 타코 세바스티안

미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775

명세서

청구범위

청구항 1

비디오 데이터를 프로세싱하는 방법으로서,

뉴럴 네트워크 시스템의 인코더 부분에 의해, 상기 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 상기 뉴럴 네트워크 시스템의 동작의 상기 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터 상기 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계;

상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 의해, 상기 뉴럴 네트워크 시스템의 동작의 상기 현재 시간 스텝에 대한 출력 데이터를 생성하는 단계로서, 상기 출력 데이터는 상기 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 단계;

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 상기 출력 데이터 및 동작의 상기 이전 시간 스텝으로부터 상기 디코더 부분의 상기 적어도 하나의 순환 계층으로부터의 상기 순환 상태 데이터를 획득하는 단계; 및

동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여, 동작의 상기 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계로서, 상기 재구성된 비디오 프레임은 상기 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하는 단계를 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 2

제 1 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계;

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계; 및

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 3

제 2 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 및

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 4

제 2 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 및

상기 디코더 부분의 포스트-워핑 네트워크에 의해, 상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임을 프로세싱하는 단계를 더 포함하고,

상기 재구성된 비디오 프레임은 상기 포스트-워핑 네트워크에 의한 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임의 프로세싱에 기초하여 동작의 상기 현재 시간 스텝에 대해 생성되는, 비디오 데이터를 프로세싱하는 방법.

청구항 5

제 1 항에 있어서,

상기 뉴럴 네트워크 시스템의 모션 추정 네트워크에 의해, 상기 입력 비디오 프레임과 동작의 상기 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 상기 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계;

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임 또는 상기 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 및

동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 및 상기 워핑된 재구성된 비디오 프레임을 상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 전송하는 단계를 더 포함하고,

동작의 상기 현재 시간 스텝에 대한 생성된 상기 출력 데이터는 동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 및 상기 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 의해 생성되는, 비디오 데이터를 프로세싱하는 방법.

청구항 6

제 5 항에 있어서,

하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 단계로서, 상기 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 1 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 단계;

상기 제 1 손실 함수를 비활성화하는 단계; 및

상기 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 단계로서, 상기 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 2 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 7

제 1 항에 있어서,

상기 순환 상태 데이터는 상기 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 8

제 1 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분은 복수의 순환 계층을 포함하고, 상기 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공하는, 비디오 데이터를 프로세싱하는 방법.

청구항 9

제 1 항에 있어서,

동작의 상기 현재 시간 스텝 또는 동작의 상기 이전 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 상기 이전 시간 스텝으로부터 상기 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 단계를 더 포함하고,

동작의 상기 현재 시간 스텝에 대한 생성된 상기 재구성된 비디오 프레임은 상기 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 비디오 데이터를 프로세싱하는 방법.

청구항 10

제 1 항에 있어서,

상기 디코더 부분에 의해, 재구성된 마스크를 생성하는 단계로서, 상기 재구성된 마스크는 동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 단계; 및

상기 재구성된 마스크, 동작의 상기 현재 시간 스텝에 대한 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 상기 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계로서, 상기 재구성된 마스크는 상기 워핑된 재구성된 비디오 프레임 및 상기 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 11

제 1 항에 있어서,

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계;

동작의 상기 현재 시간 스텝 동안, 상기 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하는 단계;

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분에 의해, 동작의 상기 현재 시간 스텝 동안 상기 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하는 단계;

동작의 상기 현재 시간 스텝 동안, 상기 제 2 재구성된 잔차 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임을 수정하는 단계; 및

수정된 상기 이전에 재구성된 비디오 프레임을 동작의 상기 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 12

제 1 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 의해, 하나 이상의 왜곡 맵을 획득하는 단계; 및

상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 의해, 상기 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 상기 출력 데이터를 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 13

제 12 항에 있어서,

상기 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 동작의 상기 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 14

제 12 항에 있어서,

상기 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함하고, 상기 워핑된 재구성된 비디오 프레임은 동작의 상기 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 생성되는, 비디오 데이터를 프로세싱하는 방법.

청구항 15

제 1 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 인코더 부분의 제 1 계층에 의해, 입력 데이터를 프로세싱하는 단계;

상기 인코더 부분의 상기 제 1 계층에 의해, 상기 인코더 부분의 상기 제 1 계층에 대한 복수의 가중치 값을 결정하는 단계;

상기 뉴럴 네트워크 시스템의 상기 인코더 부분의 제 2 계층의 가중치들을 상기 복수의 가중치 값으로 설정하는 단계; 및

상기 인코더 부분의 상기 제 2 계층에 의해, 상기 복수의 가중치 값으로 설정된 가중치들을 사용하여 상기 입력 데이터를 프로세싱하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 16

제 1 항에 있어서,

상기 출력 데이터를 저장 매체에 저장하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 17

제 1 항에 있어서,

상기 재구성된 비디오 프레임을 저장 매체에 저장하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 18

제 1 항에 있어서,

상기 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 전송하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 19

비디오 데이터를 프로세싱하기 위한 장치로서,

메모리; 및

상기 메모리에 커플링된 프로세서를 포함하고, 상기 프로세서는,

뉴럴 네트워크 시스템의 인코더 부분을 사용하여, 상기 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 상기 뉴럴 네트워크 시스템의 동작의 상기 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터 상기 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고;

상기 인코더 부분을 사용하여, 상기 뉴럴 네트워크 시스템의 동작의 상기 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 상기 출력 데이터는 상기 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하고;

뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 현재 시간 스텝에 대한 상기 출력 데이터 및 동작의 상기 이전 시간 스텝으로부터 상기 디코더 부분의 상기 적어도 하나의 순환 계층으로부터의 상기 순환 상태 데이터를 획득하며; 그리고

동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여, 동작의 상기 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 상기 재구성된 비디오 프레임은 상기 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 20

제 19 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고;

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하며; 그리고

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 21

제 20 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 그리고

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 22

제 21 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 그리고

상기 디코더 부분의 포스트-워핑 네트워크를 사용하여, 상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임을 프로세싱하도록 구성되며,

상기 재구성된 비디오 프레임은 상기 포스트-워핑 네트워크에 의한 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임의 프로세싱에 기초하여 동작의 상기 현재 시간 스텝에 대해 생성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 23

제 19 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 모션 추정 네트워크를 사용하여, 상기 입력 비디오 프레임과 동작의 상기 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 상기 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고;

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임 또는 상기 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하며; 그리고

동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 및 상기 워핑된 재구성된 비디오 프레임을 상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 전송하도록 구성되고,

동작의 상기 현재 시간 스텝에 대한 생성된 상기 출력 데이터는 동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 및 상기 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 상기 뉴럴 네트워크 시스템의 상기 인코더 부분을 사용하여 생성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 24

제 23 항에 있어서,

상기 프로세서는,

하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 것으로서, 상기 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 1 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하고;

상기 제 1 손실 함수를 비활성화하며; 그리고

상기 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 것으로서, 상기 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 2 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 25

제 19 항에 있어서,

상기 순환 상태 데이터는 상기 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 26

제 19 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분은 복수의 순환 계층을 포함하고, 상기 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 27

제 19 항에 있어서,

상기 프로세서는,

동작의 상기 현재 시간 스텝 또는 동작의 상기 이전 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 상기 이전 시간 스텝으로부터 상기 순환 상태 데이터를 수정함으로써 동작의 상기 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하도록 구성되고,

동작의 상기 현재 시간 스텝에 대한 생성된 상기 재구성된 비디오 프레임은 상기 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 28

제 19 항에 있어서,

상기 프로세서는,

상기 디코더 부분을 사용하여, 재구성된 마스크를 생성하는 것으로서, 상기 재구성된 마스크는 동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하고; 그리고

상기 디코더 부분을 사용하여, 상기 재구성된 마스크, 동작의 상기 현재 시간 스텝에 대한 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 상기 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 것으로서, 상기 재구성된 마스크는 상기 워핑된 재구성된 비디오 프레임 및 상기 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 29

제 19 항에 있어서,

상기 프로세서는,

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분을 사용하여, 상기 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하고;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분을 사용하여, 동작의 상기 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하고;

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하고;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분을 사용하여, 동작의 상기 현재 시간 스텝 동안 상기 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하고;

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분을 사용하여, 상기 제 2 재구성된 잔차 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임을 수정하며; 그리고

수정된 상기 이전에 재구성된 비디오 프레임을 동작의 상기 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 30

제 19 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 인코더 부분을 사용하여, 하나 이상의 왜곡 맵을 획득하고; 그리고

상기 뉴럴 네트워크 시스템의 상기 인코더 부분을 사용하여, 상기 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 상기 출력 데이터를 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 31

제 30 항에 있어서,

상기 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 동작의 상기 이전

시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 32

제 30 항에 있어서,

상기 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함하고, 상기 워핑된 재구성된 비디오 프레임은 동작의 상기 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 생성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 33

제 19 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 인코더 부분의 제 1 계층을 사용하여, 입력 데이터를 프로세싱하고;

상기 인코더 부분의 상기 제 1 계층을 사용하여, 상기 인코더 부분의 상기 제 1 계층에 대한 복수의 가중치 값을 결정하고;

상기 뉴럴 네트워크 시스템의 상기 인코더 부분의 제 2 계층의 가중치들을 상기 복수의 가중치 값으로 설정하며; 그리고

상기 인코더 부분의 상기 제 2 계층에 의해, 상기 복수의 가중치 값으로 설정된 가중치들을 사용하여 상기 입력 데이터를 프로세싱하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 34

제 19 항에 있어서,

상기 프로세서는,

상기 출력 데이터가 상기 메모리에 저장되게 하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 35

제 19 항에 있어서,

상기 프로세서는,

상기 재구성된 비디오 프레임이 상기 메모리에 저장되게 하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 36

제 19 항에 있어서,

상기 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 송신하도록 구성된 송신기를 더 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 37

제 19 항에 있어서,

상기 프로세서는 뉴럴 프로세싱 유닛 (NPU) 를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 38

제 19 항에 있어서,

상기 장치는 모바일 디바이스를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 39

제 19 항에 있어서,

상기 장치는 확장 현실 디바이스를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 40

제 19 항에 있어서,

디스플레이를 더 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 41

제 19 항에 있어서,

상기 장치는 텔레비전을 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 42

제 19 항에 있어서,

상기 장치는 하나 이상의 비디오 프레임을 캡처하도록 구성된 카메라를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 43

비디오 데이터를 프로세싱하는 방법으로서,

뉴럴 네트워크 시스템의 디코더 부분에 의해, 상기 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 상기 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계;

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계;

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계;

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 및

상기 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 44

제 43 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 및

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 45

제 44 항에 있어서,

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분에 의해, 동작의 상기 현재 시간 스텝 동안 상기 뉴럴 네트

워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 결정하는 단계;

동작의 상기 현재 시간 스텝 동안, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임을 수정하는 단계; 및

수정된 상기 이전에 재구성된 비디오 프레임을 동작의 상기 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 46

제 43 항에 있어서,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분에 의해, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계;

상기 디코더 부분의 포스트-워핑 네트워크에 의해, 상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임을 프로세싱하는 단계; 및

상기 포스트-워핑 네트워크에 의한 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 47

제 43 항에 있어서,

동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 또는 동작의 상기 이전 시간 스텝에 대한 결정된 모션 추정을 사용하여 동작의 상기 이전 시간 스텝으로부터 상기 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 단계; 및

상기 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 48

제 43 항에 있어서,

상기 디코더 부분에 의해, 재구성된 마스크를 생성하는 단계로서, 상기 재구성된 마스크는 동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 단계; 및

상기 재구성된 마스크, 동작의 상기 현재 시간 스텝에 대한 상기 워핑된 재구성된 비디오 프레임, 및 동작의 상기 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 단계로서, 상기 재구성된 마스크는 상기 워핑된 재구성된 비디오 프레임 및 상기 재구성된 잔차 데이터의 기여도를 가중하는, 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 49

제 43 항에 있어서,

뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 입력 비디오 프레임, 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 상기 뉴럴 네트워크 시스템의 동작의 상기 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터 상기 뉴럴 네트워크 시스템의 상기 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 및

상기 뉴럴 네트워크 시스템의 상기 인코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 출력 데이터를 생성하는 단계로서, 상기 출력 데이터는 상기 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 상기

재구성된 모션 추정 데이터, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 단계를 더 포함하는, 비디오 데이터를 프로세싱하는 방법.

청구항 50

비디오 데이터를 프로세싱하기 위한 장치로서,

메모리; 및

상기 메모리에 커플링된 프로세서를 포함하고, 상기 프로세서는,

뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 상기 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 상기 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고;

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고;

동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고;

상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하며; 그리고

상기 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 51

제 50 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 그리고

상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터에 상기 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 52

제 51 항에 있어서,

상기 프로세서는,

동작의 상기 현재 시간 스텝 동안 상기 디코더 부분을 사용하여, 동작의 상기 현재 시간 스텝 동안 상기 뉴럴 네트워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 상기 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하고;

동작의 상기 현재 시간 스텝 동안, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임을 수정하며; 그리고

수정된 상기 이전에 재구성된 비디오 프레임을 동작의 상기 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 53

제 50 항에 있어서,

상기 프로세서는,

상기 뉴럴 네트워크 시스템의 상기 디코더 부분을 사용하여, 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터 및 상기 출력 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고;

상기 디코더 부분의 포스트-워핑 네트워크를 사용하여, 상기 현재 시간 스텝에 대한 결정된 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임을 프로세싱하며; 그리고

상기 포스트-워핑 네트워크에 의한 상기 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 54

제 50 항에 있어서,

상기 프로세서는,

동작의 상기 현재 시간 스텝에 대한 결정된 상기 모션 추정 데이터 또는 동작의 상기 이전 시간 스텝에 대한 결정된 모션 추정을 사용하여 동작의 상기 이전 시간 스텝으로부터 상기 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 상기 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하고; 그리고

상기 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 55

제 50 항에 있어서,

상기 프로세서는,

상기 디코더 부분을 사용하여, 재구성된 마스크를 생성하는 것으로서, 상기 재구성된 마스크는 동작의 상기 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하고; 그리고

상기 재구성된 마스크, 동작의 상기 현재 시간 스텝에 대한 상기 워핑된 재구성된 비디오 프레임, 및 동작의 상기 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 상기 현재 시간 스텝에 대한 상기 재구성된 비디오 프레임을 생성하는 것으로서, 상기 재구성된 마스크는 상기 워핑된 재구성된 비디오 프레임 및 상기 재구성된 잔차 데이터의 기여도를 가중하는, 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 56

제 50 항에 있어서,

상기 프로세서는,

뉴럴 네트워크 시스템의 인코더 부분을 사용하여, 동작의 상기 현재 시간 스텝에 대한 입력 비디오 프레임, 상기 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 상기 뉴럴 네트워크 시스템의 동작의 상기 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터 상기 뉴럴 네트워크 시스템의 상기 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 그리고

상기 뉴럴 네트워크 시스템의 상기 인코더 부분을 사용하여, 동작의 상기 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 상기 출력 데이터는 상기 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 상기 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 57

제 50 항에 있어서,

상기 프로세서는,

상기 재구성된 비디오 프레임이 상기 메모리에 저장되게 하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 58

제 50 항에 있어서,

상기 프로세서는,

상기 재구성된 비디오 프레임이 디스플레이되게 하도록 구성되는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 59

제 50 항에 있어서,

상기 재구성된 비디오 프레임을 디스플레이하도록 구성된 디스플레이를 더 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 60

제 50 항에 있어서,

상기 프로세서는 뉴럴 프로세싱 유닛 (NPU) 를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 61

제 50 항에 있어서,

상기 장치는 모바일 디바이스를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 62

제 50 항에 있어서,

상기 장치는 확장 현실 디바이스를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 63

제 50 항에 있어서,

상기 장치는 텔레비전을 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

청구항 64

제 50 항에 있어서,

상기 장치는 하나 이상의 비디오 프레임을 캡처하도록 구성된 카메라를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치.

발명의 설명

기술 분야

[0001] 본 개시는 일반적으로 비디오의 인코딩 (또는 압축) 및 디코딩 (압축해제) 을 포함하는 비디오 코딩에 관한 것이다. 예를 들어, 본 개시의 양태들은 하나 이상의 순환 (recurrent)-기반 머신 러닝 시스템을 사용하여 비디오 콘텐츠를 코딩 (예를 들어, 인코딩 및/또는 디코딩) 하는 것에 관한 것이다.

배경 기술

[0002] 많은 디바이스들 및 시스템들은 비디오 데이터가 프로세싱되고 소비를 위해 출력되도록 한다. 디지털 비디오 데이터는 소비자 및 비디오 제공자의 요구를 충족시키기 위해 대량의 데이터를 포함한다. 예를 들어, 비디오 데이터의 소비자는 높은 충실도, 해상도, 프레임 레이트 등을 갖는 고 품질의 비디오를 원한다. 결과로서, 이러한 요구를 충족시키기 위해 필요한 대량의 비디오 데이터는 그 비디오 데이터를 프로세싱하고 저장하는 통신 네트워크 및 디바이스에 부담을 지운다.

[0003] 비디오 코딩 기법들이 비디오 데이터를 압축하는데 사용될 수도 있다. 비디오 코딩의 목표는 비디오 품질에 대한 열화를 회피 또는 최소화하면서 더 낮은 비트 레이트를 이용하는 형태로 비디오 데이터를 압축하는 것이다. 끊임없이 진화하는 비디오 서비스가 이용가능하게 됨에 따라, 우수한 코딩 효율을 갖는 인코딩 기법이 필요하다.

발명의 내용

[0004] 하나 이상의 순환-기반 머신 러닝 시스템을 사용하여 비디오 콘텐츠를 코딩 (예를 들어, 인코딩 및/또는 디코딩) 하기 위한 시스템들 및 기법들이 설명된다. 하나의 예시적인 예에서, 비디오 데이터를 프로세싱하는 방법이 제공된다. 방법은, 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 단계로서, 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 단계; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계로서, 재구성된 비디오 프레임은 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하는 단계를 포함한다.

[0005] 다른 예에서, 메모리 및 메모리에 커플링된 프로세서 (예를 들어, 회로부에서 구현됨) 를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치가 제공된다. 일부 예들에서, 하나보다 많은 프로세서가 메모리에 커플링될 수 있고, 동작들 중 하나 이상을 수행하는데 사용될 수 있다. 프로세서는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 인코더 부분을 사용하여, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하고; 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하며; 그리고 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 비디오 프레임은 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하도록 구성된다.

[0006] 다른 예에서, 명령들이 저장된 비일시적 컴퓨터 판독가능 매체가 제공되며, 명령들은 하나 이상의 프로세서에 의해 실행될 때, 하나 이상의 프로세서로 하여금, 뉴럴 네트워크 시스템의 인코더 부분을 사용하여, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하게 하고; 뉴럴 네트워크 시스템의 인코더 부분을 사용하여, 뉴럴 네

트위크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하게 하는 것으로서, 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하게 하고; 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하게 하며; 그리고 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하게 하는 것으로서, 재구성된 비디오 프레임은 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하게 한다.

[0007] 다른 예에서, 비디오 데이터를 프로세싱하기 위한 장치가 제공된다. 장치는 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 수단; 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 수단으로서, 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 수단; 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 수단; 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 수단으로서, 재구성된 비디오 프레임은 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하는 수단을 포함한다.

[0008] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 것; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것; 및 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0009] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것; 및 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0010] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것; 및 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하는 것을 더 포함한다. 이러한 양태들에서, 재구성된 비디오 프레임은 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 기초하여 동작의 현재 시간 스텝에 대해 생성된다.

[0011] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 모션 추정 네트워크에 의해, 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 것; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임 또는 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것; 및 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임을 뉴럴 네트워크 시스템의 인코더 부분에 전송하는 것을 더 포함한다. 이러한 양태들에서, 동작의 현재 시간 스텝에 대한 생성된 출력 데이터는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초

하여 뉴럴 네트워크 시스템의 인코더 부분에 의해 생성된다.

- [0012] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것으로서, 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것; 제 1 손실 함수를 비활성화하는 것; 및 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것으로서, 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것을 더 포함한다.
- [0013] 일부 양태들에서, 순환 상태 데이터는 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함한다.
- [0014] 일부 양태들에서, 뉴럴 네트워크 시스템의 디코더 부분은 복수의 순환 계층을 포함하고, 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공한다.
- [0015] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 동작의 현재 시간 스텝 또는 동작의 이전 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 것을 더 포함한다. 이러한 양태들에서, 동작의 현재 시간 스텝에 대한 생성된 재구성된 비디오 프레임은 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성된다.
- [0016] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 디코더 부분에 의해, 재구성된 마스크를 생성하는 것으로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 것; 및 재구성된 마스크, 동작의 현재 시간 스텝에 대한 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.
- [0017] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것; 동작의 현재 시간 스텝 동안, 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것; 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하는 것; 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것; 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하는 것; 동작의 현재 시간 스텝 동안, 제 2 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 것; 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 것을 더 포함한다.
- [0018] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵을 획득하는 것; 및 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 출력 데이터를 생성하는 것을 더 포함한다. 일부 양태들에서, 하나 이상의 왜곡 맵은 동작의 현재 시간 스텝에 대한 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함한다. 일부 양태들에서, 하나 이상의 왜곡 맵은 동작의 현재 시간 스텝에 대한 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함한다. 이러한 양태들에서, 워핑된 재구성된 비디오 프레임은 동작의 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 생성될 수 있다.
- [0019] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 인코더 부분의 제 1 계층에 의해, 입력 데이터를 프로세싱하는 것; 인코더 부분의 제 1 계층에 의해, 인코더 부분의 제 1 계층에 대한 복수의 가중치 값을 결정하는 것; 뉴럴 네트워크 시스템의 인코더 부분의 제 2 계층의 가중치들을 복수의 가중치 값으로 설정하는 것; 및 인코더 부분의 제 2 계층에 의해, 복수의 가중치 값으로 설정된 가중치들을 사용하여 입력 데이터를 프로세싱하는 것을 더 포함한다.

- [0020] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 저장 매체에 출력 데이터를 저장하는 것을 더 포함한다.
- [0021] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 저장 매체에 재구성된 비디오 프레임을 저장하는 것을 더 포함한다.
- [0022] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 전송하는 것을 더 포함한다.
- [0023] 다른 예시적인 예에서, 비디오 데이터를 프로세싱하는 방법이 제공된다. 방법은, 뉴럴 네트워크 시스템의 디코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 포함한다.
- [0024] 다른 예에서, 메모리 및 메모리에 커플링된 프로세서 (예를 들어, 회로부에서 구현됨)를 포함하는, 비디오 데이터를 프로세싱하기 위한 장치가 제공된다. 일부 예들에서, 하나보다 많은 프로세서가 메모리에 커플링될 수 있고, 동작들 중 하나 이상을 수행하는데 사용될 수 있다. 프로세서는 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하며; 그리고 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성된다.
- [0025] 다른 예에서, 명령들이 저장된 비일시적 컴퓨터 판독가능 매체가 제공되며, 명령들은 하나 이상의 프로세서에 의해 실행될 때, 하나 이상의 프로세서로 하여금, 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하게 하고; 뉴럴 네트워크 시스템의 디코더 부분을 사용하여, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하게 하고; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하게 하며; 그리고 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하게 한다.
- [0026] 다른 예에서, 비디오 데이터를 프로세싱하기 위한 장치가 제공된다. 장치는, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 수단; 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 수단; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 수단; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 수단; 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 수단을 포함한다.
- [0027] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 디코더 부분에 의

해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것; 및 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0028] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 뉴럴 네트워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 결정하는 것; 동작의 현재 시간 스텝 동안, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 것; 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 것을 더 포함한다.

[0029] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것; 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하는 것; 및 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0030] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 또는 동작의 이전 시간 스텝에 대한 결정된 모션 추정을 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 것; 및 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0031] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 디코더 부분에 의해, 재구성된 마스크를 생성하는 것으로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 것; 및 재구성된 마스크, 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하는 것을 더 포함한다.

[0032] 일부 양태들에서, 상술한 방법, 장치들, 및 컴퓨터 판독가능 매체는, 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 것; 및 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 것을 더 포함한다.

[0033] 일부 양태들에서, 장치는 모바일 디바이스, 웨어러블 디바이스, 확장 현실 디바이스 (예를 들어, 가상 현실 (VR) 디바이스, 증강 현실 (AR) 디바이스, 또는 혼합 현실 (MR) 디바이스), 개인용 컴퓨터, 랩탑 컴퓨터, 비디오 서버, 텔레비전, 또는 다른 디바이스를 포함한다. 일부 양태들에서, 장치는 모바일 디바이스 (예를 들어, 모바일 전화 또는 이른바 "스마트폰") 를 포함한다. 일부 양태들에서, 장치는 하나 이상의 이미지 또는 비디오 프레임을 캡처하기 위한 적어도 하나의 카메라를 더 포함한다. 예를 들어, 장치는 비디오 프레임들을 포함하는 하나 이상의 이미지 및/또는 하나 이상의 비디오를 캡처하기 위한 카메라 (예를 들어, RGB 카메라) 또는 다중 카메라들을 포함할 수 있다. 일부 양태들에서, 장치는 하나 이상의 이미지, 비디오, 통지, 또는 다른 디스플레이가능한 데이터를 디스플레이하기 위한 디스플레이를 포함한다. 일부 양태들에서, 장치는 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스로 송신하도록 구성된 송신기를 포함한다. 일부 양태들에서, 프로세서는 뉴럴 프로세싱 유닛 (NPU), 중앙 프로세싱 유닛 (CPU), 그래픽스 프로세싱 유닛 (GPU), 또는 다른 프로세싱 디바이스 또는 컴포넌트를 포함한다.

[0034] 이 개요는 청구된 청구물의 핵심적인 또는 본질적인 특징들을 식별하도록 의도되는 것도 아니고, 청구된 청구물의 범위를 결정하기 위해 별도로 이용되도록 의도되지도 않는다. 청구물은 본 특허의 전체 명세서, 일부 또는 모든 도면들 및 각각의 청구항의 적절한 부분들을 참조하여 이해되어야 한다.

[0035] 진술한 것은 다른 특징들 및 실시형태들과 함께, 다음의 명세서, 청구항들 및 첨부 도면들을 참조하면 더 명백해질 것이다.

도면의 간단한 설명

[0036] 본 출원의 예시적인 실시형태들은 다음의 도면들을 참조하여 하기에 상세히 설명된다.

도 1 은 시스템-온-칩 (system-on-a-chip; SOC) 의 예시의 구현을 도시한다.

도 2a 는 완전히 연결된 뉴럴 네트워크의 예를 도시한다.

도 2b 는 완전히 연결된 뉴럴 네트워크의 예를 도시한다.

도 2c 는 컨볼루션 뉴럴 네트워크의 예를 도시한다.

도 2d 는 이미지로부터 시각적 특징들을 인식하도록 설계된 딥 컨볼루션 네트워크 (deep convolutional network; DCN) 의 상세한 예를 도시한다.

도 3 은 딥 컨볼루션 네트워크 (DCN) 를 도시하는 블록 다이어그램이다.

도 4 는 일부 예들에 따른, 순환-기반 뉴럴 네트워크들을 사용하여 비디오 코딩을 수행하도록 동작가능한 디바이스를 포함하는 시스템의 예를 도시하는 다이어그램이다.

도 5a 는 일부 예들에 따른, 비디오 코딩을 위해 사용된 뉴럴 네트워크 시스템들의 예를 도시하는 다이어그램이다.

도 5b 는 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크의 상세한 예를 도시하는 다이어그램이다.

도 6 은 일부 예들에 따른, 모션 추정 및 모션 보상 기법들의 예들을 도시하는 다이어그램이다.

도 7 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 8 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 9 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 10 은 일부 예들에 따른, 단일 레벨 순환 및 멀티-레벨 순환의 예들을 도시하는 다이어그램이다.

도 11 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 12 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 13 은 일부 예들에 따른, 비디오 코딩을 위해 사용된 순환-기반 뉴럴 네트워크 시스템의 다른 예를 도시하는 다이어그램이다.

도 14 는 일부 예들에 따른, 순환-기반 뉴럴 네트워크 시스템에 대한 입력으로서 왜곡 맵들을 사용하는 예를 도시하는 다이어그램이다.

도 15 는 일부 예들에 따른, 순환-기반 뉴럴 네트워크 시스템에서 부가 뉴럴 네트워크 계층을 사용하는 예를 도시하는 다이어그램이다.

도 16 은 일부 예들에 따른, 본 명세서에 설명된 기법들을 사용하여 획득된 결과들 대 다른 기법들을 사용하여 획득된 결과들의 예를 도시하는 그래프이다.

도 17 은 일부 예들에 따른, 본 명세서에 설명된 기법들을 사용하여 획득된 결과들 대 다른 기법들을 사용하여 획득된 결과들의 예를 도시하는 그래프이다.

도 18 은 일부 예들에 따른, 비디오 데이터를 프로세싱하기 위한 프로세스의 예를 도시하는 플로우차트이다.

도 19 는 일부 예들에 따른, 비디오 데이터를 프로세싱하기 위한 프로세스의 다른 예를 도시하는 플로우 차트이다.

도 20 은 본 명세서에 설명된 다양한 기법들을 구현할 수 있는 예시의 컴퓨팅 디바이스의 예시의 컴퓨팅 디바이스 아키텍처를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0037] 본 개시의 소정의 양태들 및 실시형태들이 하기에 제공된다. 이들 양태들 및 실시형태들의 일부는 독립적으로 적용될 수도 있고 그 중 일부는 당업자에게 명백한 바와 같이 조합하여 적용될 수도 있다. 다음의 설명에서, 설명의 목적으로, 본 출원의 실시형태들의 철저한 이해를 제공하기 위해 특정 상세들이 제시된다. 그러나, 다양한 실시형태들이 이들 특정 상세들 없이도 실시될 수도 있음이 분명할 것이다. 도면 및 설명은 제한적인 것으로 의도되지 않는다.
- [0038] 다음의 설명은 단지 예시의 실시형태들을 제공하며, 본 개시의 범위, 적용가능성 또는 구성을 제한하도록 의도되지 않는다. 오히려, 예시의 실시형태들의 다음의 설명은 예시의 실시형태들을 구현하기 위해 가능하게 하는 설명을 당업자에게 제공할 것이다. 첨부된 청구항들에 기술된 바와 같이 본 발명의 사상 및 범위를 벗어나지 않으면서 엘리먼트들의 기능 및 배열에서 다양한 변경들이 이루어질 수도 있음을 이해해야 한다.
- [0039] 디지털 비디오 데이터는, 특히 고품질 비디오 데이터에 대한 요구가 계속 증가함에 따라, 많은 양의 데이터를 포함할 수 있다. 예를 들어, 비디오 데이터의 소비자는 통상적으로 높은 충실도, 해상도, 프레임 레이트 등을 갖는 점점 더 고 품질의 비디오를 원한다. 그러나, 이러한 요구를 충족하기 위해 필요한 많은 양의 비디오 데이터는 비디오 데이터를 프로세싱 및 저장하는 디바이스 뿐만 아니라 통신 네트워크에 상당한 부담을 줄 수 있다.
- [0040] 비디오 데이터를 코딩하기 위해 다양한 기법들이 사용될 수 있다. 비디오 코딩은 특정 비디오 코딩 표준에 따라 수행될 수 있다. 예시의 비디오 코딩 표준들은 고-효율 비디오 코딩 (HEVC), 어드밴스드 비디오 코딩 (AVC), MPEG (moving picture experts group) 코딩, 및 다기능 비디오 코딩 (VVC) 를 포함한다. 비디오 코딩은 비디오 이미지 또는 시퀀스에 존재하는 리던던시 (redundancy) 를 이용하는, 인터-예측 또는 인트라-예측과 같은 예측 방법들을 사용한다. 비디오 코딩 기법들의 일반적인 목표는 비디오 품질에 대한 열화를 회피 또는 최소화하면서, 더 낮은 비트 레이트를 사용하는 형태로 비디오 데이터를 압축하는 것이다. 비디오 서비스에 대한 요구가 증가하고 새로운 비디오 서비스가 이용가능하게 됨에 따라, 더 나은 코딩 효율성, 성능 및 레이트 제어를 갖는 코딩 기법들이 필요하다.
- [0041] 하나 이상의 머신 러닝 (ML) 시스템을 사용하여 비디오 코딩을 수행하기 위한 기법들이 본 명세서에서 설명된다. ML 은 인공 지능 (artificial intelligence; AI) 의 서브세트이다. ML 시스템들은 컴퓨터 시스템들이 명시적인 명령들의 사용없이, 패턴들 및 추론에 의존하여 다양한 태스크들을 수행하는데 사용할 수 있는 알고리즘 및 통계 모델들을 포함한다. ML 시스템의 일 예는, 인공 뉴런들 (예를 들어, 뉴런 모델들) 의 상호연결된 그룹으로 구성될 수도 있는, 뉴럴 네트워크 (또한 인공 뉴럴 네트워크로서 지칭됨) 이다. 뉴럴 네트워크들은 특히, 이미지 분석 및/또는 컴퓨터 비전 애플리케이션들, 인터넷 프로토콜 (IP) 카메라들, 사물 인터넷 (IoT) 디바이스들, 자율 차량들, 서비스 로봇들과 같은, 다양한 애플리케이션들 및/또는 디바이스들에 사용될 수도 있다.
- [0042] 뉴럴 네트워크의 개별 노드들은 입력 데이터를 취하고 데이터에 대한 간단한 동작들을 수행하여 생물학적 뉴런들을 에뮬레이트할 수도 있다. 입력 데이터에 대해 수행된 간단한 동작의 결과는 선택적으로 다른 뉴런들로 전달된다. 가중치 값들은 네트워크에서 각각의 벡터 및 노드와 연관되며, 이러한 값들은 입력 데이터가 출력 데이터에 어떻게 관련되는지를 제약한다. 예를 들어, 각각의 노드의 입력 데이터는 대응 가중치 값으로 곱해질 수도 있고, 그 곱들은 합산될 수도 있다. 곱들의 합은 선택적인 바이어스에 의해 조정될 수도 있고, 활성화 함수가 결과에 적용되어, 노드의 출력 신호 또는 "출력 활성화" 를 산출할 수도 있다 (때때로 활성화 맵 또는 특징 맵으로 지칭됨). 가중치 값들은 초기에 네트워크를 통한 훈련 데이터의 반복적인 흐름에 의해 결정

정될 수도 있다 (예를 들어, 가중치 값들은 네트워크가 이들의 통상적인 입력 데이터 특성들에 의해 특정 클래스들을 식별하는 방법을 학습하는 훈련 페이지 동안 확립됨).

[0043] 특히, 순환 뉴럴 네트워크 (recurrent neural network; RNN), 멀티계층 퍼셉트론 (multilayer perceptron; MLP) 뉴럴 네트워크, 컨볼루션 뉴럴 네트워크 (convolutional neural network; CNN) 와 같은, 다양한 타입의 뉴럴 네트워크가 존재한다. RNN 은 계층의 출력을 저장하고 이 출력을 입력에 다시 피딩하여 계층의 결과를 예측하는데 도움이 되는 원리에 따라 작동한다. MLP 뉴럴 네트워크에서, 데이터는 입력 계층에 피드될 수도 있으며, 하나 이상의 은닉된 계층은 데이터에 추상화의 레벨들을 제공한다. 그 후 추상화된 데이터에 기초하여 출력 계층에 대한 예측이 이루어질 수도 있다. MLP들은 입력들이 클래스 또는 라벨에 할당되는 분류 예측 문제들에 특히 적합할 수도 있다. 컨볼루션 뉴럴 네트워크 (CNN) 는 피드-포워드 인공 뉴럴 네트워크의 타입이다. 컨볼루션 뉴럴 네트워크는 각각 수용 필드 (예를 들어, 입력 공간의 공간적으로 로컬화된 영역) 를 갖고 입력 공간을 집합적으로 타일링하는 인공 뉴런의 집합들을 포함할 수도 있다. 컨볼루션 뉴럴 네트워크는 수많은 애플리케이션들을 갖는다. 특히, CNN 은 패턴 인식 및 분류의 영역에서 널리 사용되었다.

[0044] 계층화된 뉴럴 네트워크 아키텍처 (다중의 은닉된 계층들이 존재할 때 딥 뉴럴 네트워크로 지칭됨) 에서, 인공 뉴런들의 제 1 계층의 출력은 인공 뉴런들의 제 2 계층에 대한 입력이 되고, 인공 뉴런들의 제 2 계층의 출력은 인공 뉴런들의 제 3 계층에 대한 입력이 되는 등이다. 컨볼루션 뉴럴 네트워크는 특징들의 계위를 인식하도록 훈련될 수도 있다. 컨볼루션 뉴럴 네트워크 아키텍처에서의 계산은 하나 이상의 계산 체인으로 구성될 수도 있는, 프로세싱 노드들의 집단에 걸쳐 분산될 수도 있다. 이러한 멀티-계층화된 아키텍처는 한 번에 하나의 계층씩 훈련될 수도 있고 역 전파를 사용하여 미세-튜닝될 수도 있다.

[0045] 도 1 은 본 명세서에 설명된 기능들 중 하나 이상을 수행하도록 구성된, 중앙 프로세싱 유닛 (CPU)(102) 또는 멀티-코어 CPU 를 포함할 수도 있는, 시스템-온-칩 (SOC)(100) 의 예시의 구현을 도시한다. 특히, 파라미터들 또는 변수들 (예를 들어, 뉴럴 신호들 및 시냅스(synaptic) 가중치들), 계산 디바이스와 연관된 시스템 파라미터들 (예를 들어, 가중치들을 갖는 뉴럴 네트워크), 지연들, 주파수 빈 정보, 및 태스크 정보는 뉴럴 프로세싱 유닛 (NPU)(108) 과 연관된 메모리 블록, CPU (102) 와 연관된 메모리 블록, 그래픽스 프로세싱 유닛 (GPU)(104) 과 연관된 메모리 블록, 디지털 신호 프로세서 (DSP)(106) 와 연관된 메모리 블록, 메모리 블록 (118) 에 저장될 수도 있고, 및/또는 다중 블록들에 걸쳐 분산될 수도 있다. CPU (102) 에서 실행되는 명령들은 CPU (102) 와 연관된 프로그램 메모리로부터 로딩될 수도 있거나 메모리 블록 (118) 으로부터 로딩될 수도 있다.

[0046] SOC (100) 는 또한, GPU (104), DSP (106), 5세대 (5G) 접속성, 4세대 롱 텀 에볼루션 (4G LTE) 접속성, Wi-Fi 접속성, USB 접속성, 블루투스 접속성 등을 포함할 수도 있는 접속성 블록 (110), 및 예를 들어, 제스처들을 검출 및 인식할 수도 있는 멀티미디어 프로세서 (112) 와 같은, 특정 기능들에 맞게 조정된 부가 프로세싱 블록들을 포함할 수도 있다. 일 구현에서, NPU 는 CPU (102), DSP (106), 및/또는 GPU (104) 에서 구현된다. SOC (100) 는 또한 센서 프로세서 (114), 이미지 신호 프로세서들 (ISP들)(116), 및/또는 내비게이션 모듈 (120) 을 포함할 수 있으며, 이는 글로벌 포지셔닝 시스템을 포함할 수도 있다.

[0047] SOC (100) 는 ARM 명령 세트에 기초할 수도 있다. 본 개시의 양태에서는, CPU (102) 에 로딩된 명령들이 입력 값과 필터 가중치의 승산 곱에 대응하는 룩업 테이블 (LUT) 에서 저장된 승산 결과를 탐색하기 위한 코드를 포함할 수도 있다. CPU (102) 에 로딩된 명령들은 또한 승산 곱의 룩업 테이블 히트가 검출될 때 승산 곱의 승산 연산 동안 승산기를 디스에이블하기 위한 코드를 포함할 수도 있다. CPU (102) 에 로딩된 명령들은 또한 승산 곱의 룩업 테이블 미스가 검출될 때 필터 가중치 및 입력 값의 계산된 승산 곱을 저장하기 위한 코드를 포함할 수도 있다.

[0048] SOC (100) 및/또는 그의 컴포넌트들은 본 명세서에서 논의된 본 개시의 양태들에 따른 머신 러닝 기법들을 사용하여 비디오 압축 및/또는 압축해제 (비디오 인코딩 및/또는 디코딩으로 또한 지칭되고, 총괄적으로 비디오 코딩으로 지칭됨) 를 수행하도록 구성될 수도 있다. 비디오 압축 및/또는 압축해제를 수행하기 위해 딥 러닝 아키텍처를 사용함으로써, 본 개시의 양태들은 디바이스 상에서 비디오 압축 및/또는 압축해제의 효율성을 증가시킬 수 있다. 예를 들어, 설명된 비디오 코딩 기법들을 사용하는 디바이스는 머신 러닝 기반 기법들을 사용하여 비디오를 보다 효율적으로 압축할 수 있고, 압축된 비디오를 다른 디바이스로 송신할 수 있으며, 다른 디바이스는 본 명세서에 설명된 머신 러닝 기반 기법들을 사용하여 압축된 비디오를 보다 효율적으로 압축해제할 수 있다.

- [0049] 위에 언급한 바와 같이, 뉴럴 네트워크는 머신 러닝 시스템의 예이며, 입력 계층, 하나 이상의 은닉된 계층, 및 출력 계층을 포함할 수 있다. 입력 계층의 입력 노드들로부터 데이터가 제공되고, 하나 이상의 은닉된 계층의 은닉 노드에 의해 프로세싱이 수행되며, 출력 계층의 출력 노드들을 통해 출력이 생성된다. 딥 러닝 네트워크들은 통상적으로 다중의 은닉된 계층들을 포함한다. 뉴럴 네트워크의 각각의 계층은 인공 뉴런들 (또는 노드들) 을 포함할 수 있는 특징 맵들 또는 활성화 맵들을 포함할 수 있다. 특징 맵은 필터, 커널 등을 포함할 수 있다. 노드들은 계층들 중 하나 이상에 대한 노드들의 중요도를 표시하는데 사용된 하나 이상의 가중치를 포함할 수 있다. 일부 경우들에서, 딥 러닝 네트워크는 일련의 많은 은닉 계층들을 가질 수 있고, 초기의 계층들은 입력의 단순하고 낮은 레벨 특성들을 결정하며, 나중의 계층들은 더 복잡하고 추상적인 특성들의 계위를 구축한다.
- [0050] 딥 러닝 아키텍처는 특징들의 계위를 학습할 수도 있다. 예를 들어, 시각적 데이터로 제시되면, 제 1 계층은 입력 스트림에서, 에지들과 같은 비교적 간단한 특징들을 인식하는 것을 학습할 수도 있다. 다른 예에서, 청각적 데이터로 제시되면, 제 1 계층은 특정 주파수들에서의 스펙트럼 전력을 인식하는 것을 학습할 수도 있다. 제 1 계층의 출력을 입력으로서 취하는 제 2 계층은, 시각 데이터에 대한 간단한 형상들 또는 청각 데이터에 대한 사운드들의 조합들과 같은 특징들의 조합들을 인식하는 것을 학습할 수도 있다. 예를 들어, 상위 계층들은 시각적 데이터에서의 복잡한 형상들 또는 청각적 데이터에서의 단어들을 나타내는 것을 학습할 수도 있다. 여전히 상위 계층들은 공통 시각적 오브젝트들 또는 구어체들을 인식하는 것을 학습할 수도 있다.
- [0051] 딥 러닝 아키텍처들은 자연스러운 계위 구조를 갖는 문제들에 적용될 때 특히 잘 수행할 수도 있다. 예를 들어, 모터구동 차량들 (motorized vehicle) 의 분류는 휠들, 윈드쉴드들 및 다른 특징들을 인식하는 것을 먼저 학습하는 것으로 이익을 얻을 수 있다. 이러한 특징들은 자동차, 트럭, 및 비행기를 인식하기 위해 상이한 방식으로 상위 계층에서 조합될 수도 있다.
- [0052] 뉴럴 네트워크들은 다양한 접속성 패턴들로 설계될 수도 있다. 피드-포워드 네트워크들에서, 정보는 하위 계층에서 상위 계층으로 전달되고, 주어진 계층에서의 각각의 뉴런은 상위 계층들에서의 뉴런들에 통신한다. 계위적 표현은 상술한 바와 같이, 피드-포워드 네트워크의 연속적인 계층들에 구축될 수도 있다. 뉴럴 네트워크들은 또한 순환 또는 피드백 (또한 하향식이라 함) 연결들을 가질 수도 있다. 순환 연결에서, 주어진 계층의 뉴런으로부터의 출력은 동일한 계층의 다른 뉴런으로 통신될 수도 있다. 순환 아키텍처는 시퀀스로 뉴럴 네트워크에 전달되는 입력 데이터 청크들 중 하나보다 많은 청크들에 걸쳐 있는 패턴들을 인식하는데 도움이 될 수도 있다. 주어진 계층의 뉴런에서 하위 계층의 뉴런으로의 연결은 피드백 (또는 하향식) 연결이라고 한다. 많은 피드백 연결들을 갖는 네트워크는 하이-레벨 개념의 인식이 입력의 특정 로우-레벨 특징들을 식별하는 것을 보조할 수도 있을 때 도움이 될 수도 있다.
- [0053] 뉴럴 네트워크의 계층들 사이의 연결들은 완전히 연결되거나 로컬로 연결될 수도 있다. 도 2a 는 완전히 연결된 뉴럴 네트워크 (202) 의 예를 도시한다. 완전히 연결된 뉴럴 네트워크 (202) 에서, 제 1 계층에서의 뉴런은 제 2 계층에서의 모든 뉴런에 그의 출력을 통신할 수도 있으므로, 제 2 계층에서의 각각의 뉴런이 제 1 계층에서의 모든 뉴런으로부터 입력을 수신할 것이다. 도 2b 는 로컬로 연결된 뉴럴 네트워크 (204) 의 예를 도시한다. 로컬로 연결된 뉴럴 네트워크 (204) 에서, 제 1 층에서의 뉴런은 제 2 계층에서의 제한된 수의 뉴런들에 연결될 수도 있다. 보다 일반적으로, 로컬로 연결된 뉴럴 네트워크 (204) 의 로컬로 연결된 계층은 계층에서의 각각의 뉴런이 동일하거나 유사한 접속성 패턴을 가질 것이지만, 상이한 값들 (예를 들어, 210, 212, 214, 및 216) 을 가질 수도 있는 연결 강도들을 갖도록 구성될 수도 있다. 로컬로 연결된 접속성 패턴은 상위 계층에서 공간적으로 별개의 수용 필드들을 발생할 수도 있는데, 이는 주어진 영역에서 상위 계층 뉴런들이 네트워크에 대한 총 입력의 제한된 부분의 특성들에 대한 훈련을 통해 튜닝되는 입력들을 수신할 수도 있기 때문이다.
- [0054] 로컬로 연결된 뉴럴 네트워크의 일 예는 컨볼루션 뉴럴 네트워크이다. 도 2c 는 컨볼루션 뉴럴 네트워크 (206) 의 예를 도시한다. 컨볼루션 뉴럴 네트워크 (206) 는 제 2 계층에서의 각각의 뉴런에 대한 입력들과 연관된 연결 강도들이 공유되도록 (예를 들어, 208) 구성될 수도 있다. 컨볼루션 뉴럴 네트워크들은 입력들의 공간적 위치가 의미있는 문제들에 매우 적합할 수도 있다. 컨볼루션 뉴럴 네트워크 (206) 는 본 개시의 양태들에 따른, 비디오 압축 및/또는 압축해제의 하나 이상의 양태를 수행하기 위해 사용될 수도 있다.
- [0055] 컨볼루션 뉴럴 네트워크의 하나의 타입은 딥 컨볼루션 네트워크 (DCN) 이다. 도 2d 는 자동차-탑재형 (car-mounted) 카메라와 같은 이미지 캡처 디바이스 (230) 로부터 입력된 이미지 (226) 로부터 시각적 특징들을 인식

하도록 설계된 DCN (200) 의 상세한 예를 도시한다. 본 예의 DCN (200) 은 교통 표지판 및 교통 표지판 상에 제공된 번호를 식별하도록 훈련될 수도 있다. 물론, DCN (200) 은 차선 마킹들을 식별하거나 신호등들을 식별하는 것과 같은 다른 태스크들을 위해 훈련될 수도 있다.

[0056] DCN (200) 은 지도 학습으로 훈련될 수도 있다. 훈련 동안, DCN (200) 은 속도 제한 표지판의 이미지 (226) 와 같은 이미지로 제시될 수도 있고, 그 후 순방향 패스가 출력 (222) 을 생성하기 위해 계산될 수도 있다. DCN (200) 은 특징 추출 섹션 및 분류 섹션을 포함할 수도 있다. 이미지 (226) 를 수신하면, 컨볼루션 계층 (232) 은 이미지 (226) 에 컨볼루션 커널들 (미도시) 을 적용하여 특징 맵들 (218) 의 제 1 세트를 생성할 수도 있다. 예로서, 컨볼루션 계층 (232) 에 대한 컨볼루션 커널은 28x28 특징 맵들을 생성하는 5x5 커널일 수도 있다. 본 예에서, 4개의 상이한 특징 맵이 특징 맵들의 제 1 세트 (218) 에서 생성되기 때문에, 4개의 상이한 컨볼루션 커널이 컨볼루션 계층 (232) 에서 이미지 (226) 에 적용되었다. 컨볼루션 커널들은 또한 필터들 또는 컨볼루션 필터들로 지칭될 수도 있다.

[0057] 특징 맵들의 제 1 세트 (218) 는 특징 맵들의 제 2 세트 (220) 를 생성하기 위해 최대 풀링 계층 (미도시) 에 의해 서브샘플링될 수도 있다. 최대 풀링 계층은 특징 맵들 (218) 의 제 1 세트의 사이즈를 감소시킨다. 즉, 14x14 와 같은 특징 맵들의 제 2 세트 (220) 의 사이즈는 28x28 과 같은 특징 맵들의 제 1 세트 (218) 의 사이즈보다 작다. 감소된 사이즈는 메모리 소비를 감소시키면서 후속 계층에 유사한 정보를 제공한다. 특징 맵들의 제 2 세트 (220) 는 추가로, 특징 맵들의 하나 이상의 후속 세트 (미도시) 를 생성하기 위해 하나 이상의 후속 컨볼루션 계층 (미도시) 을 통해 컨볼루션될 수도 있다.

[0058] 도 2d 의 예에서, 특징 맵들의 제 2 세트 (220) 는 제 1 특징 벡터 (224) 를 생성하도록 컨볼루션된다. 또한, 제 1 특징 벡터 (224) 는 제 2 특징 벡터 (228) 를 생성하도록 추가로 컨볼루션된다. 제 2 특징 벡터 (228) 의 각각의 특징은 "표지판", "60" 및 "100" 과 같은 이미지 (226) 의 가능한 특징에 대응하는 수를 포함할 수도 있다. 소프트맥스 함수 (softmax function)(미도시) 는 제 2 특징 벡터 (228) 에서의 수들을 확률로 변환할 수도 있다. 이와 같이, DCN (200) 의 출력 (222) 은 하나 이상의 특징을 포함하는 이미지 (226) 의 확률이다.

[0059] 본 예에서, "표지판" 및 "60" 에 대한 출력 (222) 에서의 확률들은 "30", "40", "50", "70", "80", "90" 및 "100" 과 같은 출력 (222) 의 다른 것들의 확률들보다 높다. 훈련 전에, DCN (200) 에 의해 생성된 출력 (222) 은 부정확할 가능성이 있다. 따라서, 출력 (222) 과 타겟 출력 사이에 에러가 계산될 수도 있다. 타겟 출력은 이미지 (226) 의 실측 자료(ground truth)(예를 들어, "표지판" 및 "60") 이다. DCN (200) 의 가중치들은 그 후 DCN (200) 의 출력 (222) 이 타겟 출력과 더 밀접하게 정렬되도록 조정될 수도 있다.

[0060] 가중치들을 조정하기 위해, 러닝 알고리즘은 가중치들에 대한 그래디언트 벡터를 계산할 수도 있다. 그래디언트는 가중치가 조정되었으면 에러가 증가 또는 감소할 양을 표시할 수도 있다. 최상위 계층에서, 그래디언트는 끝에서 두번째 계층에서의 활성화된 뉴런 및 출력 계층에서의 뉴런을 연결하는 가중치의 값에 직접 대응할 수도 있다. 하위 계층들에서, 그래디언트는 가중치들의 값 및 상위 계층들의 계산된 에러 그래디언트들에 의존할 수도 있다. 가중치들은 그 후 에러를 감소시키기 위해 조정될 수도 있다. 가중치를 조정하는 이러한 방식은 뉴럴 네트워크를 통한 "역방향 패스" 를 수반하기 때문에 "역 전파" 로 지칭될 수도 있다.

[0061] 실제로, 가중치들의 에러 그래디언트는 작은 수의 예들에 걸쳐 계산될 수도 있어서, 계산된 그래디언트는 실제 에러 그래디언트에 근사한다. 이러한 근사화 방법은 확률적 그래디언트 하강법 (stochastic gradient descent) 으로 지칭될 수도 있다. 확률적 그래디언트 하강법은 전체 시스템의 달성가능한 에러율이 감소하는 것을 멈출 때까지 또는 에러율이 타겟 레벨에 도달할 때까지 반복될 수도 있다. 학습 후, DCN 이 새로운 이미지들로 제시될 수도 있고 네트워크를 통한 순방향 패스가 DCN 의 추론 또는 예측으로 간주될 수도 있는 출력 (222) 을 산출할 수도 있다.

[0062] DBN (deep belief network) 은 은닉된 노드들의 다중 계층들을 포함하는 확률 모델이다. DBN 은 훈련 데이터 세트의 계위적 표현을 추출하는데 사용될 수도 있다. DBN 은 제한된 볼츠만 머신 (Restricted Boltzmann Machines)(RBM) 의 계층들을 적층하여 획득될 수도 있다. RBM 은 입력들의 세트에 걸친 확률 분포를 학습할 수 있는 인공 뉴럴 네트워크의 타입이다. RBM들은 각각의 입력이 카테고리화되어야 하는 클래스에 관한 정보의 부재 시 확률 분포를 학습할 수 있기 때문에, RBM들은 종종 비지도 학습에 사용된다. 하이브리드 비지도 및 지도 패러다임을 사용하여, DBN 의 최하위 RBM들은 비지도 방식으로 훈련될 수도 있고 특징 추출기들로서 작용할 수도 있으며, 최상위 RBM 은 (이전 계층 및 타겟 클래스들로부터의 입력들의 공동 분포에 대해) 지도 방

식으로 훈련될 수도 있고 분류기로서 작용할 수도 있다.

- [0063] 딥 컨볼루션 네트워크 (DCN) 는 부가 풀링 및 정규화 계층들로 구성된, 컨볼루션 네트워크의 네트워크이다. DCN들은 많은 태스크들에 대해 최첨단 성능을 달성하였다. DCN들은 입력 및 출력 타겟들 양자 모두가 많은 예시들에 대해 알려져 있고 그라디언트 하강 방법들의 사용에 의해 네트워크의 가중치들을 수정하는데 사용되는 지도 학습을 사용하여 훈련될 수 있다.
- [0064] DCN 은 피드-포워드 네트워크일 수도 있다. 또한, 상술한 바와 같이, DCN 의 제 1 계층에서의 뉴런으로부터 다음 상위 계층에서의 뉴런들의 그룹으로의 연결들은 제 1 계층에서의 뉴런들에 걸쳐 공유된다. DCN들의 피드-포워드 및 공유 연결들은 빠른 프로세싱을 위해 이용될 수도 있다. DCN 의 계산 부담은 예를 들어, 순환 또는 피드백 연결들을 포함하는 유사하게 사이징된 뉴럴 네트워크의 것보다 훨씬 적을 수도 있다.
- [0065] 컨볼루션 네트워크의 각각의 계층의 프로세싱은 공간적으로 불변 템플릿 또는 베이스 투영으로 간주될 수도 있다. 입력이 컬러 이미지의 적색, 녹색 및 청색 채널들과 같은 다중 채널들로 먼저 분해되면, 그 입력에 대해 훈련된 컨볼루션 네트워크는 이미지의 축들을 따라 2개의 공간 차원 및 컬러 정보를 캡처하는 제 3 차원을 갖는, 3차원으로 간주될 수도 있다. 컨볼루션 연결들의 출력들은 후속 계층에서 특징 맵을 형성하는 것으로 간주될 수도 있고, 특징 맵의 각각의 엘리먼트 (예를 들어, 220) 는 이전 계층에서의 뉴런들의 범위 (예를 들어, 특징 맵들 (218)) 로부터 그리고 다중 채널들 각각으로부터 입력을 수신한다. 특징 맵에서의 값들은 추가로, 정류, $\max(0, x)$ 와 같은 비선형성으로 프로세싱될 수도 있다. 인접한 뉴런들로부터의 값들은 추가로 풀링될 수도 있으며, 이는 다운 샘플링에 대응하고, 부가적인 로컬 불변 및 차원성 감소를 제공할 수도 있다.
- [0066] 도 3 은 딥 컨볼루션 네트워크 (350) 의 예를 도시하는 블록 다이어그램이다. 딥 컨볼루션 네트워크 (350) 는 접속성 및 가중치 공유에 기초하여 다중의 상이한 타입들의 계층들을 포함할 수도 있다. 도 3 에 나타난 바와 같이, 딥 컨볼루션 네트워크 (350) 는 컨볼루션 블록들 (354A, 354B) 을 포함한다. 컨볼루션 블록들 (354A, 354B) 의 각각은 컨볼루션 계층 (CONV)(356), 정규화 계층 (LNorm)(358), 및 최대 풀링 계층 (MAX POOL)(360) 으로 구성될 수도 있다.
- [0067] 컨볼루션 계층들 (356) 은 특징 맵을 생성하기 위해 입력 데이터 (352) 에 적용될 수도 있는 하나 이상의 컨볼루션 필터를 포함할 수도 있다. 2개의 컨볼루션 블록 (354A, 354B) 만이 나타나 있지만, 본 개시는 그렇게 제한되지 않으며, 대신에, 설계 선호도에 따라 임의의 수의 컨볼루션 블록 (예를 들어, 블록들 (354A, 354B)) 이 딥 컨볼루션 네트워크 (350) 에 포함될 수도 있다. 정규화 계층 (358) 은 컨볼루션 필터들의 출력을 정규화할 수도 있다. 예를 들어, 정규화 계층 (358) 은 화이트닝 또는 측면 억제제를 제공할 수도 있다. 최대 풀링 계층 (360) 은 로컬 불변 및 차원성 감소를 위해 공간에 걸쳐 다운 샘플링 집성을 제공할 수도 있다.
- [0068] 예를 들어, 딥 컨볼루션 네트워크의 병렬 필터 뱅크들은 고성능 및 저전력 소비를 달성하기 위해 SOC (100) 의 CPU (102) 또는 GPU (104) 상에 로딩될 수도 있다. 대안의 실시형태들에서, 병렬 필터 뱅크들은 SOC (100) 의 DSP (106) 또는 ISP (116) 상에 로딩될 수도 있다. 또한, 딥 컨볼루션 네트워크 (350) 는 센서들 및 내비게이션에 각각 전용된, 센서 프로세서 (114) 및 내비게이션 모듈 (120) 과 같은 SOC (100) 상에 존재할 수도 있는 다른 프로세싱 블록들에 액세스할 수도 있다.
- [0069] 딥 컨볼루션 네트워크 (350) 는 또한 하나 이상의 완전히 연결된 계층, 예컨대 계층 (362A)("FC1" 으로 라벨링 됨) 및 계층 (362B)("FC2" 로 라벨링 됨) 을 포함할 수도 있다. 딥 컨볼루션 네트워크 (350) 는 로지스틱 회귀 (LR) 계층 (364) 을 더 포함할 수도 있다. 딥 컨볼루션 네트워크 (350) 의 각각의 계층 (356, 358, 360, 362, 364) 사이에는 업데이트될 가중치들 (미도시) 이 있다. 계층들 (예를 들어, 356, 358, 360, 362, 364) 각각의 출력은 컨볼루션 블록들 (354A) 중 첫번째에 공급된 입력 데이터 (352)(예를 들어, 이미지들, 오디오, 비디오, 센서 데이터 및/또는 다른 입력 데이터) 로부터 계위적 특징 표현들을 학습하기 위해 딥 컨볼루션 네트워크 (350) 에서 계층들 (예를 들어, 356, 358, 360, 362, 364) 중 후속하는 하나의 입력으로서 작용할 수도 있다. 딥 컨볼루션 네트워크 (350) 의 출력은 입력 데이터 (352) 에 대한 분류 스코어 (366) 이다. 분류 스코어 (366) 는 확률들의 세트일 수도 있고, 여기서 각각의 확률은 특징들의 세트로부터의 특징을 포함하는, 입력 데이터의 확률이다.
- [0070] 위에 언급된 바와 같이, 디지털 비디오 데이터는 많은 양의 데이터를 포함할 수 있으며, 이는 비디오 데이터를 프로세싱 및 저장하는 디바이스들 뿐만 아니라 통신 네트워크들에 상당한 부담을 줄 수 있다. 예를 들어,

압축되지 않은 비디오 콘텐츠를 기록하면 일반적으로 기록된 비디오 콘텐츠의 해상도가 증가함에 따라 크게 증가하는 파일 사이즈를 야기한다. 하나의 예시적인 예에서, 1080p/24 로 기록된 채널당 압축되지 않은 16-비트 비디오 (예를 들어, 폭이 1920 픽셀이고 높이가 1080 픽셀인 해상도, 초당 24 프레임 캡처) 는 프레임당 12.4 메가바이트 또는 초당 297.6 메가바이트를 점유할 수도 있다. 초당 24 프레임의 4K 해상도로 기록된 채널당 압축되지 않은 16-비트 비디오는 프레임당 49.8 메가바이트 또는 초당 1195.2 메가바이트를 점유할 수도 있다.

[0071] 네트워크 대역폭은 큰 비디오 파일들이 문제가 될 수 있는 다른 제약이다. 예를 들어, 비디오 콘텐츠는 종종 무선 네트워크들을 통해 전달되고 (예를 들어, LTE, LTE-어드밴스드, 뉴 라디오 (New Radio; NR), WiFi[™], 블루투스[™], 또는 다른 무선 네트워크들), 소비자 인터넷 트래픽의 많은 부분을 구성할 수 있다. 무선 네트워크들에서 이용가능한 대역폭의 양의 발전에도 불구하고, 이러한 네트워크들에서 비디오 콘텐츠를 전달하는데 사용된 대역폭의 양을 감소시키는 것이 여전히 바람직할 수도 있다.

[0072] 압축되지 않은 비디오 콘텐츠는 물리적 스토리지를 위한 상당한 크기의 메모리와 송신을 위한 상당한 대역폭을 수반할 수도 있는 큰 파일들을 야기할 수 있기 때문에, 비디오 코딩 기법들이 활용되어 이러한 비디오 콘텐츠를 압축한 후 압축해제할 수 있다.

[0073] 비디오 콘텐츠의 사이즈 - 및 이에 따른 비디오 콘텐츠를 저장하는데 수반된 스토리지의 양 - 및 비디오 콘텐츠를 전달하는데 수반된 대역폭의 양을 감소시키기 위해, 다양한 비디오 코딩 기법들은 특히, HEVC, AVC, MPEG, VVC 와 같은 특정 비디오 코딩 표준에 따라 수행될 수 있다. 비디오 코딩은 비디오 이미지 또는 시퀀스에 존재하는 리던던시를 이용하는, 인터-예측 또는 인트라-예측과 같은 예측 방법들을 사용한다. 비디오 코딩 기법들의 일반적인 목표는 비디오 품질에 대한 열화를 회피 또는 최소화하면서, 더 낮은 비트 레이트를 사용하는 형태로 비디오 데이터를 압축하는 것이다. 비디오 서비스에 대한 수요가 증가하고 새로운 비디오 서비스가 이용가능하게 됨에 따라, 더 나은 코딩 효율성, 성능 및 레이트 제어를 갖는 코딩 기법들이 필요하다.

[0074] 일반적으로 인코딩 디바이스는 비디오 코딩 표준에 따라 비디오 데이터를 인코딩하여 인코딩된 비디오 비트스트림을 생성한다. 일부 예들에서, 인코딩된 비디오 비트스트림 (또는 "비디오 비트스트림" 또는 "비트스트림") 은 일련의 하나 이상의 코딩된 비디오 시퀀스들이다. 인코딩 디바이스는 각각의 픽처를 다중 슬라이스들로 파티셔닝함으로써 픽처들의 코딩된 표현들을 생성할 수 있다. 슬라이스는 그 슬라이스 내의 정보가 동일한 픽처 내의 다른 슬라이스들로부터의 데이터에 종속되지 않고 코딩되도록 다른 슬라이스들에 대해 독립적이다. 슬라이스는 독립 슬라이스 세그먼트를 포함하는 하나 이상의 슬라이스 세그먼트들, 및 존재하는 경우, 이전 슬라이스 세그먼트들에 의존하는 하나 이상의 종속 슬라이스 세그먼트들을 포함한다. HEVC 에서, 슬라이스들은 루마 샘플들 및 크로마 샘플들의 코딩 트리 블록들 (CTB들) 로 파티셔닝된다. 샘플들에 대한 선택스와 함께, 루마 샘플들의 CTB 와 크로마 샘플들의 하나 이상의 CTB 는 코딩 트리 유닛 (CTU) 으로서 지칭된다. CTU 는 또한 "트리 블록" 또는 "최대 코딩 유닛" (LCU) 으로 지칭될 수도 있다. CTU 는 HEVC 인코딩을 위한 기본 프로세싱 단위이다. CTU 는 다양한 사이즈의 다중 코딩 유닛들 (CU들) 로 분할될 수 있다. CU 는 코딩 블록들 (CB들) 으로서 지칭되는 루마 및 크로마 샘플 배열들을 포함한다.

[0075] 루마 및 크로마 CB 들은 예측 블록들 (PB) 로 더 분할될 수 있다. PB 는 (이용 가능하거나 사용을 위해 인에이블될 때) 인터 예측 또는 인트라 블록 복사 (IBC) 예측에 대해 동일한 모션 파라미터들을 사용하는 루마 성분 또는 크로마 성분의 샘플들의 블록이다. 루마 PB 및 하나 이상의 크로마 PB 들은, 연관된 선택스와 함께, 예측 유닛 (PU) 을 형성한다. 인터 예측을 위해, 모션 파라미터들의 세트 (예를 들어, 하나 이상의 모션 벡터들, 참조 인덱스들 등) 가 각각의 PU 에 대해 비트스트림으로 시그널링되고, 루마 PB 및 하나 이상의 크로마 PB 들의 인터 예측을 위해 사용된다. 모션 파라미터들은 모션 정보로서도 지칭될 수 있다. CB 는 또한 하나 이상의 변환 블록들 (TB) 로 파티셔닝될 수 있다. TB 는 예측 잔차 신호를 코딩하기 위해 잔차 변환 (예를 들어, 일부 경우들에서, 동일한 2차원 변환) 이 적용되는 컬러 컴포넌트의 샘플들의 정사각형 블록을 나타낸다. 변환 유닛 (TU) 은 루마 및 크로마 샘플들의 TB 들 및 대응하는 선택스 엘리먼트들을 나타낸다. 변환 코딩은 하기에서 더 상세히 설명된다.

[0076] HEVC 표준에 따르면, 변환들은 TU들을 사용하여 수행될 수도 있다. TU 들은 주어진 CU 내의 PU 들의 사이즈에 기초하여 사이즈가 정해질 수도 있다. TU들은 PU들과 동일한 사이즈이거나 그 보다 더 작을 수도 있다. 일부 예들에서, CU 에 대응하는 잔차 샘플들은 잔차 쿼드트리 (RQT) 로서 알려진 쿼드트리 구조를 사용하여 더 작은 유닛들로 세분화될 수도 있다. RQT 의 리프 노드들은 TU들에 대응할 수도 있다. TU들과 연관된 픽셀 차이 값들은 변환 계수들을 생성하도록 변환될 수도 있다. 변환 계수들은 그 후 인코딩 디바이스에 의

해 양자화될 수도 있다.

[0077] 비디오 데이터의 픽처들이 CU 들로 파티셔닝되면, 인코딩 디바이스는 예측 모드를 사용하여 각각의 PU 를 예측한다. 그 후, 예측 유닛 또는 예측 블록은 잔차들 (하기에 기재됨) 을 얻기 위해 원래의 비디오 데이터로부터 감산된다. 각각의 CU 에 대해, 예측 모드는 선택스 데이터틀 사용하여 비트스트림 내부에서 시그널링될 수도 있다. 예측 모드는 인트라 예측 (또는 인트라-픽처 예측) 또는 인터 예측 (또는 인터-픽처 예측) 을 포함할 수도 있다. 인트라 예측은 픽처 내에서 공간적으로 이웃하는 샘플들 간의 상관관계를 활용한다. 예를 들어, 인트라-예측을 사용하여, 각각의 PU 는, 예를 들어, PU 에 대한 평균값을 발견하기 위한 DC 예측, PU 에 대해 평면 표면을 피팅하기 위한 평면 예측, 이웃하는 데이터로부터 외삽하기 위한 방향 예측, 또는 임의의 다른 적절한 타입의 예측을 사용하여 동일한 픽처 내의 이웃하는 이미지 데이터로부터 예측된다. 인터 예측은 이미지 샘플들의 블록에 대한 모션 보상된 예측을 도출하기 위해 픽처들 간의 시간 상관관계를 사용한다. 예를 들어, 인터-예측을 사용하여, 각각의 PU 는 (출력 순서로 현재 픽처의 전 또는 후의) 하나 이상의 기준 픽처들에서의 이미지 데이터로부터의 모션 보상 예측을 사용하여 예측된다. 인터 픽처 또는 인트라 픽처 예측을 사용하여 픽처 영역을 코딩할지 여부의 결정은 예를 들어 CU 레벨에서 행해질 수도 있다.

[0078] 일부 예들에서, 픽처의 하나 이상의 슬라이스는 슬라이스 타입이 할당된다. 슬라이스 타입은 I 슬라이스, P 슬라이스 및 B 슬라이스를 포함한다. I 슬라이스 (인트라-프레임, 독립적으로 디코딩 가능) 는 인트라-예측에 의해서만 코딩되는 픽처의 슬라이스이며, 따라서 I 슬라이스는 슬라이스의 임의의 예측 유닛 또는 예측 블록을 예측하기 위해 프레임 내의 데이터만을 필요로 하기 때문에 독립적으로 디코딩 가능하다. P 슬라이스 (단방향 예측 프레임들) 는 인트라-예측 및 단방향 인터-예측으로 코딩될 수도 있는 픽처의 슬라이스이다. P 슬라이스 내의 각각의 예측 유닛 또는 예측 블록은 인트라 예측 또는 인터-예측으로 코딩된다. 인터-예측이 적용될 때, 예측 유닛 또는 예측 블록은 하나의 참조 픽처에 의해서만 예측되며 이에 따라, 참조 샘플들은 하나의 프레임의 하나의 참조 영역으로부터만이다. B 슬라이스 (양방향 예측성 프레임들) 는 인트라-예측 및 인터-예측 (예를 들어, 이중 예측 또는 단일 예측) 으로 코딩될 수도 있는 픽처의 슬라이스이다. B 슬라이스의 예측 유닛 또는 예측 블록은 2개의 참조 픽처들로부터 양방향으로 예측될 수도 있고, 여기서, 각각의 픽처가 하나의 참조 영역에 기여하고 2개의 참조 영역들의 샘플 세트들이 (예를 들어, 동일한 가중치 또는 상이한 가중치를 이용하여) 가중되어 양방향 예측 블록의 예측 신호를 생성한다. 위에 설명된 바와 같이, 하나의 픽처의 슬라이스들은 독립적으로 코딩된다. 일부 경우들에서, 픽처는 단지 하나의 슬라이스로서 코딩될 수 있다.

[0079] 인트라 및/또는 인터 예측을 사용하여 예측을 수행한 후, 인코딩 디바이스는 변환 및 양자화를 수행할 수 있다. 예를 들어, 예측에 후속하여, 인코딩 디바이스는 PU 에 대응하는 잔차 값들을 계산할 수도 있다. 잔차 값들은 코딩되는 픽셀들의 현재 블록 (PU) 과 현재 블록을 예측하기 위해 이용되는 예측 블록 (예를 들어, 현재 블록의 예측된 버전) 사이의 픽셀 차이 값들을 포함할 수도 있다. 예를 들어, 예측 블록을 생성한 후 (예를 들어, 인터-예측 또는 인트라-예측을 발행한 후), 인코딩 디바이스는 현재 블록으로부터 예측 유닛에 의해 생성된 예측 블록을 감산하는 것에 의해 잔차 블록을 생성할 수 있다. 잔차 블록은 현재 블록의 픽셀 값들과 예측 블록의 픽셀 값들 사이의 차이들을 정량화하는 픽셀 차이 값들의 세트를 포함한다. 일부 예들에서, 잔차 블록은 2 차원 블록 포맷 (예를 들어, 2 차원 매트릭스 또는 픽셀 값들의 어레이) 으로 표현될 수도 있다. 이러한 예들에서, 잔차 블록은 픽셀 값들의 2 차원 표현이다.

[0080] 예측이 수행된 후에 남을 수도 있는 임의의 잔차 데이터는 이산 코사인 변환, 이산 사인 변환, 정수 변환, 웨이브렛 변환, 다른 적절한 변환 함수 또는 이들의 임의의 조합에 기초할 수도 있는 블록 변환을 사용하여 변환된다. 일부 경우들에서, 하나 이상의 블록 변환 (예를 들어, 사이즈들 32 x 32, 16 x 16, 8 x 8, 4 x 4, 또는 다른 적절한 사이즈) 이 각각의 CU 에서 잔차 데이터에 적용될 수도 있다. 일부 실시형태들에서, TU 는 인코딩 디바이스에 의해 구현되는 변환 및 양자화 프로세스들을 위해 사용될 수도 있다. 하나 이상의 PU를 갖는 주어진 CU 는 또한 하나 이상의 TU를 포함할 수도 있다. 하기에 더 상세히 설명되는 바와 같이, 잔차 값들은 블록 변환들을 사용하여 변환 계수들로 변환될 수도 있고, 그 후 TU들을 사용하여 양자화되고 스캔되어 엔트로피 코딩을 위한 직렬화된 변환 계수들을 생성할 수도 있다.

[0081] 인코딩 디바이스는 변환 계수들의 양자화를 수행할 수도 있다. 양자화는 변환 계수들을 양자화하여 추가의 압축을 제공하여 계수들을 표현하는데 사용된 데이터의 양을 감소시킨다. 예를 들어, 양자화는 그 계수들의 일부 또는 전부와 연관되는 비트 깊이를 감소시킬 수도 있다. 일 예에서, n-비트 값을 갖는 계수는 양자화 동안 m-비트 값으로 라운드 다운될 수도 있으며, 여기서, n 은 m 보다 더 크다.

- [0082] 양자화가 수행되면, 코딩된 비디오 비트스트림은 양자화된 변환 계수들, 예측 정보 (예를 들어, 예측 모드들, 모션 벡터들, 블록 벡터들 등), 파티셔닝 정보, 및 다른 신택스 데이터와 같은 임의의 다른 적합한 데이터를 포함한다. 코딩된 비디오 비트스트림의 상이한 엘리먼트들은 그 후 인코딩 디바이스에 의해 엔트로피 인코딩될 수도 있다. 일부 예들에서, 인코딩 디바이스는 양자화된 변환 계수들을 스캔하기 위한 미리정의된 스캔 순서를 활용하여 엔트로피 인코딩될 수 있는 직렬화된 벡터를 생성할 수도 있다. 일부 예들에서, 인코딩 디바이스는 적응적 스캔을 수행할 수도 있다. 벡터 (예를 들어, 1차원 벡터) 를 형성하기 위해 양자화된 변환 계수들을 스캐닝한 후, 인코딩 디바이스는 그 벡터를 엔트로피 인코딩할 수도 있다. 예를 들어, 인코딩 디바이스는 컨텍스트 적응적 가변 길이 코딩, 컨텍스트 적응적 이진 산술 코딩, 신택스 기반 컨텍스트 적응적 이진 산술 코딩, 확률 간격 파티셔닝 엔트로피 코딩, 또는 다른 적절한 엔트로피 인코딩 기법을 사용할 수도 있다.
- [0083] 인코딩 디바이스는 인코딩된 비디오 비트스트림을 저장할 수 있고 및/또는 인코딩된 비디오 비트스트림 데이터를 통신 링크를 통해 디코딩 디바이스를 포함할 수 있는 수신 디바이스로 전송할 수 있다. 디코딩 디바이스는 인코딩된 비디오 데이터를 구성하는 하나 이상의 코딩된 비디오 시퀀스의 엘리먼트들을 (예를 들어, 엔트로피 디코더를 이용하여) 엔트로피 디코딩하고 추출함으로써 인코딩된 비디오 비트스트림 데이터를 디코딩할 수도 있다. 그 후, 디코딩 디바이스는 인코딩된 비디오 비트스트림 데이터에 대해 리스케일링 및 역변환을 수행할 수도 있다. 잔차 데이터는 그 후 디코딩 디바이스의 예측 스테이지로 전달된다. 그 후 디코딩 디바이스는 인트라 예측, 인터 예측, IBC, 및/또는 다른 타입의 예측을 사용하여 픽셀들의 블록 (예를 들어, PU) 을 예측한다. 일부 예들에서, 예측은 역 변환의 출력 (잔차 데이터) 에 추가된다. 디코딩 디바이스는 디코딩된 비디오를 비디오 목적지 디바이스에 출력할 수도 있고, 비디오 목적지 디바이스는 디코딩된 비디오 데이터를 콘텐츠의 소비자에게 디스플레이하기 위한 디스플레이 또는 다른 출력 디바이스를 포함할 수도 있다.
- [0084] 다양한 비디오 코딩 표준들에 의해 정의된 비디오 코딩 시스템들 및 기법들 (예를 들어, 상술한 HEVC 비디오 코딩 기법들) 은 원시 비디오 콘텐츠에 많은 정보를 보유할 수도 있고 신호 프로세싱 및 정보 이론 개념들에 기초하여 선형적으로 정의될 수도 있다. 그러나, 미리정의된 압축 알고리즘이 일반적으로 (예를 들어, 임의의 타입의 비디오 콘텐츠에) 적용될 수도 있지만, 이러한 미리정의된 비디오 코딩 기법들은 네트워크 기반 환경 또는 다른 타입의 환경에서 유익한 소정의 특성들이 부족할 수도 있다. 예를 들어, 비디오 코딩 시스템들은 저 레이턴시 동작을 구현할 수 있어야 한다. 그러나, 기존 비디오 코딩 표준들에 따라 동작하는 비디오 코딩 시스템들 및 관련 기법들은 비디오를 여러 프레임들의 배치(batch)(때때로 픽처들의 그룹 (Group-of-Pictures) 또는 GoP 로 지칭됨) 로 인코딩 및 송신하며, 따라서 송신될 배치 (예를 들어, GoP) 의 프레임들의 누적을 대기하는데 필요하다. 이러한 배치-기반 비디오 코딩은 프레임별 코딩 및 송신을 사용하여 달성할 수 있는 것보다 더 높은 레이턴시를 초래한다. 또한, 이러한 비디오 코딩 시스템들 및 기법들은 비디오의 특성들에 의존하여 런타임 시 GoP 사이즈 (즉, 잔차 정보이기 보다는, 전체 프레임을 송신하는 빈도) 를 조정하는 능력이 부족하여, 레이트-왜곡 공간에서 적은 최적의 동작 포인트들을 유도한다.
- [0085] 인코더 스테이지 및/또는 디코더 스테이지에 피드백을 제공하는 순환 뉴럴 네트워크 아키텍처를 사용하여 비디오 코딩 (비디오 압축/인코딩 및/또는 비디오 압축해제/디코딩) 을 수행하는 방법들 (또한 프로세스들로 지칭됨), 시스템들, 장치들 및 컴퓨터 판독가능 매체들 (총괄적으로 "시스템들 및 기법들" 로 지칭됨) 이 본 명세서에 설명된다. 일부 예들에서, 시스템들 및 기법들은 다중의 뉴럴 네트워크 시스템들을 활용하여 비디오 프레임들 또는 이미지들을 압축 및 압축해제할 수 있다. 예를 들어, 제 1 뉴럴 네트워크 시스템은 인트라 예측된 프레임 (또한 I-프레임으로 지칭됨) 과 유사한, 하나 이상의 이전 프레임에 기초하여 모션 보상을 사용하지 않으면서 전체 프레임을 압축하기 위해 이미지 인코더 (예를 들어, 이미지 오토인코더) 로서 작용할 수 있다. 제 2 뉴럴 네트워크 시스템 (예를 들어, 순환 오토인코더) 은 하나 이상의 이전 프레임에 기초한 모션 보상을 사용하고 잔차 정보를 사용하여 프레임들을 인코딩 및/또는 디코딩하는데 사용될 수 있다 (단일 참조 프레임 (P-프레임이라고도 함) 을 사용하는 단일-예측된 프레임 또는 2개의 참조 프레임 (또한 B-프레임으로 지칭됨) 을 사용한 이중-예측된 프레임과 같은, 인터-예측된 프레임과 유사함).
- [0086] 제 2 뉴럴 네트워크 시스템에 의해 수행된 모션 보상은 이전에 재구성된 프레임으로부터의 픽셀들이 잔차 정보와 함께 현재 프레임에서의 새로운 포지션들로 어떻게 이동하는지를 설명함으로써 현재 프레임의 데이터를 결정하는데 사용될 수 있다. 제 2 뉴럴 네트워크 시스템은 향후 시간 스텝들에서 뉴럴 네트워크의 소정 계층들에 피드백될 수 있는 상태 정보를 저장하는 (예를 들어, 제 2 뉴럴 네트워크의 디코더 부분에서의) 하나 이상의 순환 계층을 포함할 수 있다. 예를 들어, 제 2 뉴럴 네트워크 시스템은 현재 시간 스텝 t 에서 프레임 x_t

를 프로세싱할 수 있고, 하나 이상의 순환 계층으로부터의 프레임 x_t 을 프로세싱하는 것과 연관된 상태 정보를 다음 시간 스텝 $t + 1$ 에서의 후속 프레임 x_{t+1} 을 프로세싱하기 위한 제 2 뉴럴 네트워크 시스템에 피드백할 수 있다. 하나 이상의 순환 계층은 하나 이상의 순환 뉴럴 네트워크(RNN), 하나 이상의 게이팅된 순환 유닛(GRU), 하나 이상의 장단기 메모리(LSTM) 네트워크, 하나 이상의 ConvRNN, 하나 이상의 ConvGRU, 하나 이상의 ConvLSTM, 이들의 임의의 조합, 및/또는 다른 타입의 뉴럴 네트워크 아키텍처의 일부일 수 있다.

[0087] 이러한 뉴럴 네트워크 시스템과 관련된 다양한 양태들이 본 명세서에 설명된다. 예를 들어, 일부 구현들에서, 제 2 뉴럴 네트워크는(예를 들어, 전형적인 표준 기반 비디오 코덱에서 사용된 GoP 구조와 같은) 시간-역방향 프레임 종속성 없이 프레임 단위로(이전 프레임만 참조) 동작할 수 있는 완전 자기회귀 조건부 구조를 가질 수 있다. 일부 경우들에서, 제 2 뉴럴 네트워크의 출력은 모션 보상에 의해 워핑된 이전의 재구성된 프레임, 잔차, 재구성된 프레임, 모션 추정(예를 들어, 광학 흐름, 블록 모션 벡터 또는 적응적 컨볼루션 커널), 이들의 임의의 조합, 및/또는 다른 출력을 포함할 수 있다. 제 1 시간 스텝(예를 들어, 시간 스텝 t) 으로부터의 피드백 정보는 디코더로부터 인코더로 및/또는 나중 시간 스텝에서(예를 들어, 시간 스텝 $t+1$ 에서) 디코더로 제공될 수 있다. 피드백은 디코딩 스테이지에서 도입된 에러들을 나타내는데 사용될 수 있으며, 이는 인코더에 의해 사용되어 그러한 에러들을 보상할 수 있다. 이전 시간 스텝의 디코더로부터 다음 시간 스텝의 인코더로 제공되는 정보는, 하나 이상의 이전에 재구성된 프레임(들), 하나 이상의 이전에 재구성된 모션 추정(들), 하나 이상의 이전에 재구성된 잔차(들), 디코더의(예를 들어, 디코더의 하나 이상의 순환 계층으로부터) 하나 이상의 이전 순환 상태(들), 이들의 임의의 조합, 및/또는 다른 출력을 포함할 수 있다. 순환 계층의 순환 상태는 또한 본 명세서에서 은닉된 상태로 지칭될 수 있다.

[0088] 본 명세서에 설명된 하나 이상의 순환-기반뉴럴 네트워크 시스템들 및 기법들을 사용하여 비디오 콘텐츠를 코딩(예를 들어, 코딩/압축 및 디코딩/압축해제) 함으로써, 본 개시의 양태들은 종래 선형적으로 정의된 비디오 압축 알고리즘(예를 들어, 특히 HEVC, VVC, MPEG-4) 에 접근 또는 초과하는 비디오 비트레이트 및/또는 품질을 갖는 압축된 비디오 콘텐츠를 생성하는 비디오 인코딩 및 디코딩 메커니즘을 제공할 수 있다. 일부 경우들에서, 본 명세서에 설명된 순환-기반뉴럴 네트워크 시스템들 및 기법들을 사용하는 비디오 인코딩 및 디코딩 메커니즘은 뉴럴 네트워크(들)을 재훈련함으로써 다양한 사용 경우들에 적응가능할 수도 있다. 또한, 본 개시의 양태들은 비트스트림으로 송신될 많은 양의 선택스를 요구하지 않으면서 비디오 인코더들 및 디코더들의 자율 생성을 제공할 수도 있다.

[0089] 도 4 는 비디오 인코딩 및 디코딩을 수행하도록 구성된 디바이스(402) 를 포함하는 시스템(400) 을 도시한다. 디바이스(402) 는 카메라(407) 및 저장 매체(414)(예를 들어, 데이터 저장 디바이스) 에 커플링된다. 일부 구현들에서, 카메라(407) 는 제 1 뉴럴 네트워크 시스템(410) 및/또는 제 2 뉴럴 네트워크 시스템(411) 에 의한 인코딩을 위해 프로세서(404) 에 이미지 데이터(408)(예를 들어, 비디오 데이터 스트림) 를 제공하도록 구성된다. 일부 구현들에서, 디바이스(402) 는 다중의 카메라들(예를 들어, 듀얼 카메라 시스템, 3개의 카메라, 또는 다른 수의 카메라) 에 커플링될 수 있고 및/또는 이를 포함할 수 있다. 일부 경우들에서, 디바이스(402) 는 마이크로폰 및/또는 다른 입력 디바이스(예를 들어, 키보드, 마우스, 터치 입력 디바이스, 예컨대 터치스크린 및/또는 터치패드, 및/또는 다른 입력 디바이스) 에 커플링될 수 있다. 일부 예들에서, 카메라(407), 저장 매체(414), 마이크로폰, 및/또는 다른 입력 디바이스는 디바이스(402) 의 일부일 수 있다.

[0090] 디바이스(402) 는 또한 하나 이상의 무선 네트워크, 하나 이상의 유선 네트워크, 또는 이들의 조합과 같은 송신 매체(418) 를 통해 제 2 디바이스(490) 에 커플링된다. 예를 들어, 송신 매체(418) 는 무선 네트워크, 유선 네트워크, 또는 유선 및 무선 네트워크의 조합에 의해 제공되는 채널을 포함할 수 있다. 송신 매체(418) 는 패킷 기반 네트워크, 예컨대 로컬 영역 네트워크, 광역 네트워크, 글로벌 네트워크, 예컨대 인터넷의 부분을 형성할 수도 있다. 송신 매체(418) 는 라우터들, 스위치들, 기지국들, 또는 소스 디바이스로부터 목적지 디바이스로의 통신을 용이하게 하는데 유용할 수도 있는 임의의 다른 장비를 포함할 수도 있다. 무선 네트워크가 임의의 무선 인터페이스 또는 무선 인터페이스들의 조합을 포함할 수도 있고, 임의의 적합한 무선 네트워크(예를 들어, 인터넷 또는 다른 광역 네트워크, 패킷 기반 네트워크, WiFi™, 무선 주파수(RF), UWB, WiFi-Direct, 셀룰러, LTE(Long-Term Evolution), WiMax™ 등)를 포함할 수도 있다. 유선 네트워크가 임의의 유선 인터페이스(예를 들어, 섬유, 이더넷, 전력선 이더넷, 동축 케이블을 통한 이더넷, 디지털 신호 라인(DSL) 등) 를 포함할 수도 있다. 유선 및/또는 무선 네트워크들은 여러 장비들, 이를 테면, 기지국, 라우터, 액세스 포인트, 브리지, 게이트웨이, 스위치 등을 이용하여 구현될 수도 있다. 인코딩된

비디오 데이터는 무선 통신 프로토콜과 같은 통신 표준에 따라 변조되고, 수신 디바이스로 송신될 수도 있다.

- [0091] 디바이스 (402) 는 메모리 (406), 제 1 인터페이스 ("I/F 1")(412), 및 제 2 인터페이스 ("I/F 2")(416) 에 커플링된 하나 이상의 프로세서 (본 명세서에서는 "프로세서" 로 지칭됨)(404) 를 포함한다. 프로세서 (404) 는 카메라 (407), 메모리 (406) 및/또는 저장 매체 (414) 로부터 이미지 데이터 (408) 를 수신하도록 구성된다. 프로세서 (404) 는 제 1 인터페이스 (412) 를 통해 (예를 들어, 메모리 버스를 통해) 저장 매체 (414) 에 커플링되고, 제 2 인터페이스 (416)(예를 들어, 네트워크 인터페이스 디바이스, 무선 트랜시버 및 안테나, 하나 이상의 다른 네트워크 인터페이스 디바이스, 또는 이들의 조합) 를 통해 송신 매체 (418) 에 커플링된다.
- [0092] 프로세서 (404) 는 인코더 부분 (462) 및 디코더 부분 (466) 을 포함하는 제 1 뉴럴 네트워크 시스템 (410), 및 인코더 부분 (432) 및 디코더 부분 (436) 을 포함하는 제 2 뉴럴 네트워크 시스템 (411) 을 포함한다. 일부 구현들에서, 뉴럴 네트워크 시스템 (410) 및/또는 뉴럴 네트워크 시스템 (411) 은 하나 이상의 오토인코더를 포함할 수 있다. 인코더 부분 (462) 은 입력 데이터 (470) 를 수신하고 입력 데이터 (470) 를 프로세싱하여 입력 데이터 (470) 에 적어도 부분적으로 기초하여 출력 데이터 (474) 를 생성하도록 구성된다. 유사하게, 인코더 부분 (432) 은 입력 데이터 (440) 를 수신하고 입력 데이터 (440) 를 프로세싱하여 입력 데이터 (440) 에 적어도 부분적으로 기초하여 출력 데이터 (444) 를 생성하도록 구성된다. 일부 경우들에서, 뉴럴 네트워크 시스템 (411) 의 인코더 부분 (432) 은 하기에서 더 설명되는 바와 같이, 입력 데이터 (440) 에 기초하여 그리고 또한 제 1 상태 데이터 (450) 에 기초하여 출력 데이터 (444) 를 생성하도록 구성된다. 일부 경우들에서, 인코더 부분 (432) 은 입력 데이터 (440) 및 제 1 상태 데이터 (450) 에 기초하여 그리고 또한 하기에서 더 설명되는 바와 같이, (예를 들어, 이전 시간 스텝 $t-1$ 에서 입력 데이터 (470) 를 프로세싱할 때) 이전 시간 스텝 $t-1$ 에서 뉴럴 네트워크 시스템 (410) 의 디코더 부분 (466) 에 의해 생성된 표현 (476) 에 기초하여 현재 시간 스텝 t 에서 출력 데이터 (444) 를 생성하도록 구성된다. 일부 경우들에서, 인코더 부분 (432) 은 입력 데이터 (440) 및 제 1 상태 데이터 (450) 에 기초하여 그리고 또한 하기에서 더 설명되는 바와 같이, (예를 들어, 이전 시간 스텝 $t-1$ 에서 입력 데이터 (440) 를 프로세싱할 때) 이전 시간 스텝 $t-1$ 에서 뉴럴 네트워크 시스템 (411) 의 디코더 부분 (436) 에 의해 생성된 표현 (446) 에 기초하여 현재 시간 스텝 t 에서 출력 데이터 (444) 를 생성하도록 구성된다.
- [0093] 일부 구현들에서, 뉴럴 네트워크 시스템 (410) 의 인코더 부분 (462) 은 출력 데이터 (474) 를 생성하기 위해 입력 데이터 (470) 의 손실 압축을 수행하도록 구성되어, 출력 데이터 (474) 가 입력 데이터 (470) 보다 적은 비트를 갖도록 한다. 인코더 부분 (462) 은 임의의 이전 표현들 (예를 들어, 하나 이상의 이전에 재구성된 프레임) 에 기초하여 모션 보상을 사용하지 않으면서 입력 데이터 (470)(예를 들어, 비디오 프레임들) 를 압축하도록 훈련될 수 있다. 예를 들어, 인코더 부분 (462) 은 비디오 프레임으로부터의 비디오 데이터만을 사용하고 이전에 재구성된 프레임들의 임의의 데이터를 사용하지 않으면서 비디오 프레임을 압축할 수 있다. 인코더 부분 (462) 에 의해 프로세싱된 비디오 프레임들은 본 명세서에서 인트라 예측된 프레임 (I-프레임) 으로 지칭될 수 있다. 일부 예들에서, I-프레임들은 (예를 들어, HEVC, VVC, MPEG-4, 또는 다른 비디오 코딩 표준에 따른) 전형적인 비디오 코딩 기법들을 사용하여 생성될 수 있다. 이러한 예들에서, 프로세서 (404) 는 HEVC 표준과 관련하여 상술한 것과 같은, 블록 기반 인트라 예측을 수행하도록 구성된 비디오 코딩 디바이스 (예를 들어, 인코딩 디바이스) 를 포함하거나 이와 커플링될 수도 있다. 이러한 예들에서, 뉴럴 네트워크 시스템 (410) 은 프로세서 (404) 로부터 제외될 수도 있다.
- [0094] 일부 구현들에서, 뉴럴 네트워크 시스템 (411) 의 인코더 부분 (432) 은 출력 데이터 (444) 를 생성하기 위해 입력 데이터 (440) 의 손실 압축을 수행하도록 구성되어, 입력 데이터 (440) 보다 적은 비트를 갖는 출력 데이터 (444) 를 야기한다. 인코더 부분 (432) 은 이전 표현들 (예를 들어, 하나 이상의 이전에 재구성된 프레임) 에 기초하여 모션 보상을 사용하여 입력 데이터 (440)(예를 들어, 비디오 프레임들) 를 압축하도록 훈련될 수 있다. 예를 들어, 인코더 부분 (432) 은 그 비디오 프레임으로부터의 비디오 데이터를 사용하고 이전에 재구성된 프레임들의 데이터를 사용하여 비디오 프레임을 압축할 수 있다. 인코더 부분 (462) 에 의해 프로세싱된 비디오 프레임들은 본 명세서에서 인트라 예측된 프레임 (P-프레임들) 으로 지칭될 수 있다. 모션 보상은 이전에 재구성된 프레임으로부터의 픽셀들이 잔차 정보와 함께 현재 프레임에서의 새로운 포지션들로 어떻게 이동하는지를 설명함으로써 현재 프레임의 데이터를 결정하는데 사용될 수 있다. 모션 보상 기법들의 예들은 도 6 과 관련하여 하기에 설명된다.
- [0095] 특정 구현에서, 뉴럴 네트워크 시스템 (410) 의 인코더 부분 (462) 은 뉴럴 네트워크 (463) 및 양자화기 (464) 를 포함한다. 뉴럴 네트워크 (463) 는 하나 이상의 컨볼루션 뉴럴 네트워크 (CNN), 하나 이상의 완전 연결

된 뉴럴 네트워크, 하나 이상의 게이팅된 순환 유닛 (GRU), 이들의 임의의 조합, 및/또는 중간 데이터 (472) 를 생성하는 다른 타입의 뉴럴 네트워크 아키텍처들을 포함할 수 있다. 중간 데이터 (472) 는 양자화기 (464) 에 입력된다. 뉴럴 네트워크 시스템 (411) 의 인코더 부분 (432) 은 뉴럴 네트워크 (433) 및 양자화기 (434) 를 포함한다. 뉴럴 네트워크 (433) 는 하나 이상의 CNN, 하나 이상의 완전 연결된 뉴럴 네트워크, 하나 이상의 GRU, 이들의 임의의 조합, 및/또는 그 후 양자화기 (434) 에 입력되는 중간 데이터 (442) 를 생성하는 다른 타입의 뉴럴 네트워크 아키텍처들을 포함한다. 인코더 부분 (462) 및 인코더 부분 (432) 에 포함될 수도 있는 컴포넌트들의 예들은 도 5a 내지 도 15 에 도시되어 있다.

[0096]

양자화기 (464) 는 출력 데이터 (474) 를 생성하기 위해 중간 데이터 (472) 의 양자화 및 일부 경우들에서 엔트로피 코딩을 수행하도록 구성된다. 출력 데이터 (474) 는 양자화된 (그리고 일부 경우들에서는 엔트로피 코딩된) 데이터를 포함할 수 있다. 유사하게, 양자화기 (434) 는 출력 데이터 (444) 를 생성하기 위해 중간 데이터 (442) 의 양자화 및 일부 경우들에서 엔트로피 코딩을 수행하도록 구성된다. 출력 데이터 (444) 는 양자화된 (그리고 일부 경우들에서는 엔트로피 코딩된) 데이터를 포함한다. 양자화기 (464) 및 양자화기 (434) 에 의해 수행되는 양자화 동작들은 동일한 동작들을 수행할 수 있으며, 이는 중간 데이터 (472) 및 중간 데이터 (442) 로부터 양자화된 코드들 (또는 뉴럴 네트워크 시스템 (410) 및/또는 뉴럴 네트워크 시스템 (411) 에 의해 생성된 양자화된 코드들을 나타내는 데이터) 의 생성을 야기할 수 있다. 양자화 코드들 (또는 양자화된 코드들을 나타내는 데이터) 은 또한 잠재성 코드들 또는 잠재성으로 지칭될 수 있다. 잠재성에 적용되는 엔트로피 모델은 본 명세서에서 "이전" 으로 지칭될 수 있다. 일부 예들에서, 양자화 및 엔트로피 코딩 동작들은 기존 비디오 코딩 표준들에 따라 비디오 데이터를 인코딩 및/또는 디코딩할 때 수행되는 기존 양자화 및 엔트로피 코딩 동작들을 사용하여 수행될 수 있다. 일부 예들에서는, 뉴럴 네트워크 시스템 (410) 및/또는 뉴럴 네트워크 시스템 (411) 에 의해 수행될 수 있다. 예를 들어, 뉴럴 네트워크 시스템 (410) 및/또는 뉴럴 네트워크 시스템 (411) 은 지도 훈련을 사용하여 훈련될 수 있고, 간차 데이터는 입력으로 사용되고 양자화된 코드들 및 엔트로피 코드들은 훈련 동안 알려진 출력 (라벨들) 으로 사용된다.

[0097]

뉴럴 네트워크 시스템 (410) 의 디코더 부분 (466) 은 (예를 들어, 양자화기 (464) 및/또는 저장 매체 (414) 로부터 직접) 출력 데이터 (474) 를 수신하고 출력 데이터 (474) 를 프로세싱하여 출력 데이터 (474) 에 적어도 부분적으로 기초하여 입력 데이터 (470) 의 표현 (476) 을 생성하도록 구성된다. 유사하게, 뉴럴 네트워크 시스템 (411) 의 디코더 부분 (436) 은 (예를 들어, 양자화기 (434) 및/또는 저장 매체 (414) 로부터 직접) 출력 데이터 (444) 를 수신하고 출력 데이터 (444) 를 프로세싱하여 출력 데이터 (444) 에 적어도 부분적으로 기초하여 입력 데이터 (440) 의 표현 (446) 을 생성하도록 구성된다. 일부 경우들에서, 디코더 부분 (436) 은 하기에서 더 설명되는 바와 같이, 출력 데이터 (444) 및 제 2 상태 데이터 (452) 에 기초하여 표현 (446) 을 생성하도록 구성된다. 일부 경우들에서, 디코더 부분 (436) 은 출력 데이터 (444) 및 제 2 상태 데이터 (452) 에 기초하여 그리고 또한 하기에서 더 설명되는 바와 같이, (예를 들어, 이전 시간 스텝 $t-1$ 에서 입력 데이터 (470) 를 프로세싱할 때) 이전 시간 스텝 $t-1$ 에서 뉴럴 네트워크 시스템 (410) 의 디코더 부분 (466) 에 의해 생성된 표현 (476) 에 기초하여 현재 시간 스텝 t 에서 표현 (446) 을 생성하도록 구성된다. 일부 경우들에서, 디코더 부분 (436) 은 출력 데이터 (444) 및 제 2 상태 데이터 (452) 에 기초하여 그리고 또한 하기에서 더 설명되는 바와 같이, (예를 들어, 이전 시간 스텝 $t-1$ 에서 입력 데이터 (440) 를 프로세싱할 때) 이전 시간 스텝 $t-1$ 에서 디코더 부분 (466) 에 의해 생성된 표현 (446) 에 기초하여 현재 시간 스텝 t 에서 표현 (446) 을 생성하도록 구성된다.

[0098]

특정 구현에서, 뉴럴 네트워크 시스템 (410) 의 디코더 부분 (466) 은 하나 이상의 CNN, 하나 이상의 완전 연결된 뉴럴 네트워크, 하나 이상의 GRU, 이들의 임의의 조합, 및/또는 다른 타입의 뉴럴 네트워크 아키텍처들을 포함할 수도 있는 뉴럴 네트워크 (468) 를 포함한다. 뉴럴 네트워크 시스템 (411) 의 디코더 부분 (436) 은 하나 이상의 CNN, 하나 이상의 완전 연결된 뉴럴 네트워크, 하나 이상의 GRU, 하나 이상의 장단기 메모리 (LSTM) 네트워크, 하나 이상의 ConvRNN, 하나 이상의 ConvGRU, 하나 이상의 ConvLSTM, 이들의 임의의 조합, 및/또는 다른 타입의 뉴럴 네트워크 아키텍처를 포함할 수도 있는 뉴럴 네트워크 (438) 를 포함한다. 디코더 부분 (436) 에 포함될 수도 있는 컴포넌트들의 예들은 도 5a 내지 도 15 에 도시되어 있다.

[0099]

디코더 부분 (436) 의 뉴럴 네트워크 (438) 는 하나 이상의 RNN, GRU, LSTM, ConvRNN, ConvGRU, ConvLSTM, 이들의 임의의 조합, 및/또는 하나 이상의 순환 계층을 갖는 다른 뉴럴 네트워크 아키텍처를 포함할 수 있는 순환 컴포넌트를 갖는다. 순환 컴포넌트 (예를 들어, 하나 이상의 순환 계층) 는 제 1 상태 데이터 (450) 및 제 2 상태 데이터 (452) 를 제공할 수 있다. 제 1 상태 데이터 (450) 및 제 2 상태 데이터 (452) 는 일 세트의 입력 데이터 (440) 에 대한 표현 (446) 의 생성으로부터 야기되는 뉴럴 네트워크 (438) 의 상태에 대응한다.

제 1 상태 데이터 (450) 및 제 2 상태 데이터 (452)(예를 들어, 현재 비디오 프레임의 프로세싱 동안과 같은, 시간 스텝 t 와 연관됨)는, 입력 데이터 (440)의 다음 순차 세트(예를 들어, 현재 비디오 프레임 이후 프로세싱되는 다음 비디오 프레임의 프로세싱 동안과 같은, 다음 시간 스텝 $t+1$ 에서)에 대해, 각각 인코더 부분 (432) 및 디코더 부분 (436)에 대한 입력으로서 제공된다. 예시적인 예에서, 상태 데이터 (450, 452)에 표현된 디코더 부분 (436)의 상태는 뉴럴 네트워크 (438) 내의 노드들의 하나 이상의 값을 포함한다. 결과로서, 뉴럴 네트워크 시스템 (411)은 입력 데이터 (440)와 입력 데이터 (440)의 표현 (446)사이의 차이를 훈련 세트(예를 들어, 입력 이미지들과 출력 이미지들의 훈련 세트)에 비해 감소시키도록 훈련될 수 있는 순환-기반뉴럴 네트워크 시스템으로서 동작한다. 일부 구현들에서, 제 1 상태 데이터 (450)는 제 2 상태 데이터 (452)와 매칭한다(예를 들어, 제 1 상태 데이터 (450)는 제 2 상태 데이터 (452)와 동일하다). 다른 구현들에서, 제 1 상태 데이터 (450)는 제 2 상태 데이터 (452)와 상이할 수 있다.

[0100]

도 4에 나타난 바와 같이, 피드백 정보는 뉴럴 네트워크 시스템 (411)의 디코더 부분 (436)으로부터 뉴럴 네트워크 시스템 (411)의 인코더 부분 (432)으로 제공된다. 피드백은 디코딩 스테이지에서 도입된 에러들을 나타내는데 사용될 수 있으며, 이는 인코더 부분 (432)에 의해 사용되어 그러한 에러들을 보상할 수 있다. 다음 시간 스텝 t 에서 디코더 부분 (436)으로부터 인코더로 제공될 수 있는 이전 시간 스텝 $t-1$ 으로부터의

정보는 디코더 부분 (436)의 뉴럴 네트워크 (438)로부터의 하나 이상의 이전에 재구성된 프레임(들)($\hat{x}_t$ 로 표기됨), 이전에 재구성된 모션 추정(들), 이전에 재구성된 잔차(들), 및/또는 이전 순환 상태(들)을 포함할 수 있다. 예를 들어, 시간 상관을 나타낼 수도 있는 비디오 프레임들의 시퀀스에 대해, 비디오의 프레임과 연관된 디코더 부분 (436)의 이전 순환 상태는 비디오의 다음 프레임을 프로세싱할 때 인코더 부분 (432) 및 디코더 부분 (436)양자 모두에 대한 입력으로서 제공될 수 있다. 이러한 순환 상태 뿐만 아니라 다른 피드백 정보를 사용하면, 압축 비율(예를 들어, 출력 데이터 (444)에서의 비트 수와 비교하여 입력 데이터 (440)의 비트 수)을 증가시키고, 따라서 인코더 부분 (432)이 디코더 부분 (436)에 의해 직면된 에러들을 설명하는 것을 가능하게 함으로써 코딩 효율을 증가시킬 수 있다. 예시적인 예로서, 상태 데이터 (450, 452)는 인코더 부분 (432) 및 디코더 부분 (436)이 일련의 프레임들 중 다음 프레임에 대한 예측을 생성하는 것을 가능하게 하는 입력으로서 기능할 수 있고 인코더 부분 (432)은 디코더 부분 (436)의 이력을 고려하지 않으면서 출력 데이터 (444)를 인코딩하는 것과 비교하여 감소된 비트 레이트에서 출력 데이터 (444)로서 잔차(예를 들어, 다음 프레임과 다음 프레임의 예측 사이의 차이)를 인코딩할 수 있다. 결과로서, 인코딩된 데이터는 감소된 양의 저장 공간을 사용하여 저장되고, 감소된 양의 네트워크 대역폭을 사용하여 또는 더 높은 송신 레이트로, 또는 이들의 임의의 조합으로 송신될 수 있다.

[0101]

프로세서 (404)는 출력 데이터 (444)를 송신 매체 (418) 또는 저장 매체 (414) 중 적어도 하나로 전송하도록 구성된다. 예를 들어, 출력 데이터 (444)는 출력 데이터 (444)를 생성하기 위해 인코딩된/압축된 비디오 데이터의 재생과 같은, 재구성된 데이터로서 입력 데이터 (440)의 표현 (446)을 생성하기 위해 디코더 부분 (436)에 의한 나중 취출 및 압축해제를 위해 저장 매체 (414)에 저장될 수도 있다. 일부 구현들에서, 출력 데이터 (444)는 디코더 부분 (436)과 매칭하는 다른 디코더 디바이스에서(예를 들어, 디바이스 (402)에서 또는 다른 디바이스에서) 디코딩되어 재구성된 데이터로서 입력 데이터 (440)의 표현 (446)을 생성할 수도 있다. 다른 예로서, 제 2 디바이스 (490)는 디코더 부분 (436)과 매칭(또는 실질적으로 매칭)하는 디코더를 포함할 수도 있고, 출력 데이터 (444)는 제 2 디바이스 (490)에서 재구성된 데이터로서 입력 데이터 (440)의 표현 (446)을 생성하기 위해 송신 매체 (418)를 통해 송신될 수도 있다.

[0102]

시스템 (400)의 컴포넌트들은 하나 이상의 프로그램가능한 전자 회로(예를 들어, 마이크로프로세서, 그래픽스 프로세싱 유닛 (GPU), 디지털 신호 프로세서 (DSP), 중앙 프로세싱 유닛 (CPU), 및/또는 다른 적절한 전자 회로)를 포함할 수 있고 및/또는 이를 사용하여 구현될 수 있으며, 및/또는 본 명세서에 설명된 다양한 동작들을 수행하기 위해 컴퓨터 소프트웨어, 펌웨어, 또는 이들의 임의의 조합을 포함하고 및/또는 이를 사용하여 구현될 수 있다.

[0103]

시스템 (400)이 소정의 컴포넌트들을 포함하는 것으로 나타나 있지만, 당업자는 시스템 (400)이 도 4에 나타난 것들보다 더 많거나 더 적은 컴포넌트들을 포함할 수 있음을 알 것이다. 예를 들어, 시스템 (400)은 또한 입력 디바이스 및 출력 디바이스(나타내지 않음)를 포함하는 컴퓨팅 디바이스를 포함하거나 그 일부일 수 있다. 일부 구현들에서, 시스템 (400)은 또한 하나 이상의 메모리 디바이스(예를 들어, 하나 이상의 랜덤 액세스 메모리 (RAM) 컴포넌트, 판독 전용 메모리 (ROM) 컴포넌트, 캐시 메모리 컴포넌트, 버퍼 컴포넌트, 데이터베이스 컴포넌트 및/또는 다른 메모리 디바이스), 하나 이상의 메모리 디바이스와 통신하고 및/또는 이에 전

기적으로 연결되는 하나 이상의 프로세싱 디바이스 (예를 들어, 하나 이상의 CPU, GPU 및/또는 다른 프로세싱 디바이스), 무선 통신을 수행하기 위한 하나 이상의 무선 인터페이스 (예를 들어, 각각의 무선 인터페이스에 대한 하나 이상의 트랜시버 및 베이스대역 프로세서를 포함함), 하나 이상의 하드와이어링 연결들을 통한 통신들을 수행하기 위한 하나 이상의 유선 인터페이스 (예를 들어, 직렬 인터페이스, 예컨대 유니버설 직렬 버스 (USB) 입력, 라이트닝 커넥터, 및/또는 다른 유선 인터페이스), 및/또는 도 4 에 나타내지 않은 다른 컴포넌트들을 포함하는 컴퓨팅 디바이스를 포함할 수도 있거나 그 부분일 수 있다.

[0104] 일부 구현들에서, 시스템 (300) 은 컴퓨팅 디바이스에 의해 로컬로 구현되고 및/또는 컴퓨팅 디바이스에 포함될 수 있다. 예를 들어, 컴퓨팅 디바이스는 모바일 디바이스, 개인용 컴퓨터, 태블릿 컴퓨터, 가상 현실 (VR) 디바이스 (예를 들어, 헤드-장착형 디스플레이 (HMD) 또는 다른 VR 디바이스), 증강 현실 (AR) 디바이스 (예를 들어, HMD, AR 안경 또는 다른 AR 디바이스), 웨어러블 디바이스, 서버 (예를 들어, SaaS (Software as a Service) 시스템 또는 다른 서버 기반 시스템), 텔레비전 및/또는 본 명세서에 설명된 기법들을 수행하기 위한 리소스 능력들을 갖는 다른 컴퓨팅 디바이스를 포함할 수 있다.

[0105] 일 예에서, 뉴럴 네트워크 시스템 (411)(및 일부 경우들에서는 뉴럴 네트워크 시스템 (410)) 은 프로세서 (404) 에 커플링되고 프로세서 (404) 에 의해 실행가능한 명령들을 저장하도록 구성된 메모리 (406), 및 안테나와 프로세서 (404) 에 커플링되고 출력 데이터 (444) 를 원격 디바이스에 송신하도록 동작가능한 무선 트랜시버를 포함하는 휴대용 전자 디바이스에 통합될 수 있다.

[0106] 도 5a 는 (도 4 의 뉴럴 네트워크 시스템 (410) 의 예로서의) 뉴럴 네트워크 시스템 (510) 및 (뉴럴 네트워크 시스템 (411) 의 예로서의) 뉴럴 네트워크 시스템 (511) 을 포함하는 시스템을 도시하는 다이어그램이다. 뉴럴 네트워크 시스템 (510) 의 I-프레임 인코더 (563) 는 이전에 재구성된 프레임들의 데이터에 기초하여 모션 보상을 수행하지 않으면서, 그 비디오 프레임으로부터의 비디오 데이터만을 사용하여 비디오 프레임을 인코딩 또는 압축하도록 훈련되는 뉴럴 네트워크 (예를 들어, 하나 이상의 CNN 을 포함함) 를 포함할 수 있다. 예

를 들어, 도 5a 에 나타난 바와 같이, I-프레임 인코더 (563) 는 입력으로서 비디오 프레임 $X_I\text{-frame}$ 을 수신하고 비디오 프레임 $X_I\text{-frame}$ 에 포함된 데이터만을 사용하여 비디오 프레임 $X_I\text{-frame}$ 을 인코딩 또는 압축한다. I-프레임 인코더 (563) 는 비디오 프레임 $X_I\text{-frame}$ 을 인코딩 또는 압축하여 중간 데이터 (예를 들어, 중간 데이터 (472)) 를 생성할 수 있다. 일부 경우들에서, 중간 데이터는 실제 입력 비디오 프레임 $X_I\text{-frame}$ 과 비디오 프레임 $X_I\text{-frame}$ 의 예측 사이의 차이를 나타내는 잔차를 나타내는 데이터를 포함할 수 있다.

[0107] 뉴럴 네트워크 시스템 (510) 의 양자화기 (564) 는 도 4 에 나타난 양자화기 (464) 의 예이다. 양자화기 (564) 는 I-프레임 인코더 (563) 로부터 출력된 중간 데이터로부터 잠재성을 생성한다. 일부 경우들에서,

잠재성은 압축된 비디오 프레임 $X_I\text{-frame}$ 의 양자화된 버전 (예를 들어, 잔차의 양자화된 버전) 을 나타내는 양자화된 코드들을 포함한다. 일부 경우들에서, 잠재성은 양자화된 코드들을 나타내는 뉴럴 네트워크 데이터 (예를 들어, 노드의 활성화 맵 또는 특징 맵) 를 포함할 수 있다. 양자화기 (564) 는 또한 I-프레임 프라이어(prior)를 포함한다. I-프레임 프라이어는 엔트로피 코드들을 나타내는 엔트로피 코드들 또는 뉴럴 네트워크 데이터 (예를 들어, 노드의 활성화 맵 또는 특징 맵) 를 생성하기 위해 잠재성에 적용되는 엔트로피 모델을 포함하거나 나타낼 수 있다. 양자화기 (564) 는 I-프레임 디코더 (568) 및/또는 나중 취출을 위한 스토리지 (예를 들어, 저장 매체 (414)) 로 출력되는 출력 데이터 (예를 들어, 출력 데이터 (474)) 를 생성한다. 예를 들어, 양자화기 (564) 로부터의 출력 데이터는 송신 매체 (418) 를 통해 디바이스 (490) 로 송신될 수 있다.

[0108] 뉴럴 네트워크 시스템 (510) 의 I-프레임 디코더 (568) 는 양자화기 (564) 로부터의 출력 데이터를 프로세싱하여 재구성된 프레임 $\hat{X}_I\text{-frame}$ 을 생성한다. 일부 구현들에서, I-프레임 인코더 (563), 양자화기 (564), 및 I-프레임 디코더 (568) 는 엔드-투-엔드로 함께 훈련되는 단일 뉴럴 네트워크 또는 다중 뉴럴 네트워크들을 사용하여 구현될 수 있다. 도 5a 에 나타난 손실 함수 $Loss_0 = \sum_t distortion(x_t, \hat{x}_t)$ 는 입력 이미지들 및 재구성된 출력 이미지들을 포함하는 훈련 세트와 지도 학습 기법들을 사용하여 단일 뉴럴 네트워크를 훈

련하는데 사용될 수 있다. 뉴럴 네트워크 훈련에 관한 추가 상세는 뉴럴 네트워크 시스템 (511) 과 관련하여 하기에서 설명된다. 일부 구현들에서, I-프레임 인코더 (563), I-프레임 디코더 (568), 및 양자화기 (564) 의 뉴럴 네트워크들은 별도로 훈련될 수 있다.

[0109] 위에서 언급한 바와 같이, 일부 구현들에서, I-프레임 인코더 (563), I-프레임 디코더 (568), 및 양자화기 (564) 는 (예를 들어, HEVC, VVC, MPEG-4, 또는 다른 비디오 코딩 표준에 따른) 전형적인 비디오 코딩 기법들을 수행하여 I-프레임을 생성할 수 있다. 예를 들어, (비디오 인코딩 및 디코딩을 수행할 수 있는) 비디오 코딩 디바이스는 HEVC 표준 또는 다른 비디오 코딩 표준에 따라 I-프레임들을 생성하기 위해 블록 기반 인트라 예측을 수행하도록 구성될 수 있다.

[0110] 뉴럴 네트워크 시스템 (511) 은 비디오 코딩 (예를 들어, 인코딩 및/또는 디코딩) 을 수행하기 위해 모션 보상 기법들을 사용하여 뉴럴 네트워크 시스템 (511) 으로 인해, 일부 경우들에서는 P-프레임 오토인코더로 지칭될 수 있다. 뉴럴 네트워크 시스템 (511) 의 인코더 (533) 는 하나 이상의 이전에 재구성된 프레임의 데이터에 기초하여 모션 보상을 사용하여 비디오 프레임을 인코딩 또는 압축하도록 훈련되는 뉴럴 네트워크 (예를 들어, 하나 이상의 CNN 을 포함함) 를 포함할 수 있다. 뉴럴 네트워크 시스템 (511) 은 2개의 시간 스텝, t 및 $t+1$ 으로 도 5a 에 나타나 있다. 도 5a 에 나타난 시간 스텝 t 는 I-프레임이 생성된 후 (예를 들어, $\hat{X}_{I-frame}$) 뉴럴 네트워크 시스템 (511) 의 동작의 제 1 시간 스텝을 도시한다. 시간 스텝 t 에서, 인코더 (533) 는 비디오 프레임 X_t 및 이전에 재구성된 프레임 $X_{I-frame}$ 을 입력으로서 수신하고, 모션 보상을 수행함으로써 비디오 프레임 X_t 을 인코딩하여 비디오 프레임 X_t 과 비디오 프레임 $X_{I-frame}$ 사이의 모션 양을 결정한다.

[0111] 인코더 (533) 에 의해 다양한 타입들의 모션 보상이 수행될 수 있다. 예를 들어, 인코더 (533) 는 블록 기반 모션 보상 (특히 HEVC, VVC, MPEG-4 와 같은 비디오 코딩 표준들에 의해 수행된 것과 유사), 광학적 흐름 기반 모션 보상, 동적 (또는 적응) 컨볼루션 기반 모션 보상, 이들의 임의의 조합, 및/또는 다른 타입들의 모션 보상을 수행할 수 있다. 다양한 모션 보상 옵션들이 도 6 과 관련하여 하기에 설명된다.

[0112] 비디오 프레임 X_t 을 인코딩하는 결과로서, 인코더 (533) 는 중간 데이터 (예를 들어, 중간 데이터 (442)) 를 생성할 수 있다. 일부 경우들에서, 중간 데이터는 비디오 프레임 X_t 과 비디오 프레임 $X_{I-frame}$ 사이의 모션 양을 나타내는 모션 데이터 및 잔차를 나타내는 데이터를 포함할 수 있다. 잔차는 실제 입력 비디오 프레임 X_t 과 비디오 프레임 X_t 의 예측 사이의 차이를 나타낸다. 비디오 프레임 X_t 과 비디오 프레임 $X_{I-frame}$ 사이의 모션 양을 나타내는 모션 데이터는 전형적인 비디오 코딩 기법들 (예를 들어, 특히 HEVC, VVC) 에서의 모션 벡터와 유사한 것으로 개념화될 수 있다.

[0113] 뉴럴 네트워크 시스템 (511) 의 양자화기 (534) 는 도 4 에 나타난 양자화기 (434) 의 예이다. 양자화기 (534) 는 인코더 (533) 로부터 출력된 중간 데이터로부터 잠재성을 생성한다. 일부 경우들에서, 잠재성은 인코딩된 비디오 프레임 X_t 의 양자화된 버전 (예를 들어, 잔차의 양자화된 버전 및 모션 데이터의 양자화된 버전) 을 나타내는 양자화된 코드들을 포함한다. 일부 경우들에서, 잠재성은 양자화된 코드들을 나타내는 뉴럴 네트워크 데이터 (예를 들어, 노드의 활성화 맵 또는 특징 맵) 를 포함할 수 있다. 양자화기 (534) 는 또한 프라이어를 포함하며, 이는 엔트로피 코드들을 나타내는 엔트로피 코드들 또는 뉴럴 네트워크 데이터 (예를 들어, 노드의 활성화 맵 또는 특징 맵) 를 생성하기 위해 잠재성에 적용되는 엔트로피 모델을 포함하거나 나타낼 수 있다. 양자화기 (534) 는 디코더 (538) 및/또는 나중 취출을 위한 스토리지 (예를 들어, 저장 매체 (414)) 로 출력되는 출력 데이터 (예를 들어, 출력 데이터 (444)) 를 생성한다. 예를 들어, 양자화기 (534) 로부터의 출력 데이터는 송신 매체 (418) 를 통해 디바이스 (490) 로 송신될 수 있다.

[0114] 뉴럴 네트워크 시스템 (511) 의 디코더 (538) 는 출력 데이터를 프로세싱하여 재구성된 프레임 $\hat{X}_t$ 을 생성한다. 디코더 (538) 는 양자화기 (534), 스토리지 (예를 들어, 저장 매체 (414)) 로부터, 또는 다른 디바이스로부터 송신 매체 (예를 들어, 송신 매체 (418)) 를 통해 출력 데이터를 수신할 수 있다. 도 5a 에 나타난 바와 같

이, 디코더 (538) 는 하나 이상의 프리-순환 디코더 계층 (570), 하나 이상의 순환 계층 (571), 및 하나 이상의 포스트-순환 계층 (572) 을 포함하는 순환 계층들 또는 컴포넌트들을 포함한다. 하나 이상의 순환 계층 (571) 은 상태 정보를 저장할 수 있고, 후속 시간 스텝들에서 (예를 들어, 도 5a 에 나타난 바와 같이, 시간 스텝 $t+1$ 에서) 디코더 (538) 및/또는 인코더 (533) 의 소정 계층들에 상태 정보를 피드백할 수 있다. 나중 시간 스텝들에서 뉴럴 네트워크 시스템 (511) 에서의 초기 계층들에 제공된 피드백은 뉴럴 네트워크 시스템 (511) 에 일종의 메모리를 제공한다. 하나 이상의 프리-순환 디코더 계층 (570) 및 하나 이상의 포스트-순환 계층 (572) 은 일 방향으로 네트워크의 다른 계층들을 통해 정보를 순방향으로 전달하는 순방향 계층을 포함할 수 있는 반면, 하나 이상의 순환 계층은 위에 언급된 바와 같이, 피드백을 통해 순방향 및/또는 역방향으로 정보를 전달할 수 있다.

[0115] 디코더 (538) 의 순환 상태는 도 5a 에서 h 로서 표기된다. 뉴럴 네트워크 시스템 (510) 이 I-프레임을 생성한 시간 스텝 $t-1$ 에 대응하는, 도 5a 에 나타난 순환 상태 h_{t-1} 는, (예를 들어, 임의의 순환 계층들을 포함하지 않는 뉴럴 네트워크 시스템 (510) 으로 인해) 하나 이상의 0 값으로 개시될 수 있다. 도 5a 에서의 순환 상태 h_t 는 타임 스텝 t 에서의 디코더 (538) 의 순환 상태를 나타낸다. 순환 상태 h_t 는 하나 이상의 순환 계층 (571) 으로부터의 뉴럴 네트워크 정보를 포함할 수 있다. 예를 들어, 뉴럴 네트워크 정보는 (예를 들어, 데이터로 승산된 하나 이상의 순환 계층 (571) 의 가중치들의 승산이 디코더 (538) 를 통과하는 것에 기초하여 생성된) 활성화 맵 또는 특징 맵을 포함할 수 있다. 도 5a 에 나타난 바와 같이, 순환 상태 h_t 는 다음 시간 스텝 $t+1$ 에서 하나 이상의 순환 계층 (571) 및 인코더 (533) 에 피드백으로서 전달된다.

[0116] 출력 데이터의 프로세싱에 기초하여, 디코더 (538) 의 하나 이상의 포스트-순환 계층들 (572) 은 재구성된 모션 추정 데이터 $\hat{f}_t$ 및 재구성된 잔차 데이터 $\hat{r}_t$ 를 출력한다. 일부 경우들에서, 재구성된 모션 추정 데이터 $\hat{f}_t$ 는 도 6 과 관련하여 하기에 설명된 바와 같이, 광학 흐름 데이터 (예를 들어, 변위 벡터), 동적 컨볼루션 데이터 (예를 들어, 데이터 컨볼루션을 위한 행렬 또는 커널), 또는 블록-기반 모션 데이터 (예를 들어, 모션 벡터) 를 포함할 수 있다. 재구성된 모션 추정 데이터 $\hat{f}_t$ 는 이전에 재구성된 프레임 $\hat{x}_{I-frame}$ 을 (예를 들어, 모션 보상을 수행함으로써) 워핑하기 위해 디코더 (538) 의 워핑 엔진 (567) 에 의해 사용되어, 워핑된 프레임의 생성을 야기한다. 예를 들어, 이전에 재구성된 프레임 $\hat{x}_{I-frame}$ 의 픽셀들은 새로운 위치들로 이동되거나 그렇지 않으면 재구성된 모션 추정 데이터 $\hat{f}_t$ 에 따라 수정될 수 있다. 워핑 엔진 (567) 에 의해 출력된 워핑된 프레임은 가산기 엔진 (569) 으로 출력된다. 가산기 엔진 (569) 은 워핑된 프레임을 재구성된 잔차 데이터 $\hat{r}_t$ 에 가산하여 현재 시간 스텝 t 에 대한 재구성된 프레임 $\hat{x}_t$ 을 생성한다.

[0117] 순환 상태 정보 (예를 들어, 순환 상태 h_t) 에 부가하여, 재구성된 프레임 $\hat{x}_t$ 은 또한 동작의 다음 시간 스텝 $t+1$ 동안 뉴럴 네트워크 시스템 (511) 으로 피드백될 수 있다. 예를 들어, 도 5a 에 나타난 바와 같이, 재구성된 프레임 $\hat{x}_t$ 및 순환 상태 h_t 는 입력으로서 인코더 (533) 에 제공된다. 일부 경우들에서, 도 5a 에 나타난 바와 같이, 이전에 재구성된 모션 추정 데이터 $\hat{f}_t$ 및 이전에 재구성된 잔차 데이터 $\hat{r}_t$ 는 또한 동작의 다음 시간 스텝 $t+1$ 에서 인코더 (533) 에 대한 입력으로서 피드백될 수 있다. 시간 스텝 $t+1$ 에서, 인코더 (533) 는 현재 입력 비디오 프레임 $\hat{x}_{t+1}$, 이전에 재구성된 프레임 $\hat{x}_t$, 순환 상태 h_t , 이전에 재구성된 모션 추정 데이터 $\hat{f}_t$, 및 이전에 재구성된 잔차 데이터 $\hat{r}_t$ 를 프로세싱하여 현재 입력 비디오 프레임 $\hat{x}_{t+1}$ 의 코딩된 버전을 생성할 수 있다. 양자화기 (534) 는 잠재성을 생성하기 위해 코딩된 비디오 데이터

를 양자화할 수 있고, 출력 데이터를 생성하기 위해 프라이어를 사용하여 잠재성에 대해 (프라이어를 사용하여) 엔트로피 코딩을 수행할 수 있다.

[0118] 디코더 (538) 의 하나 이상의 프리-순환 디코더 계층들 (570), 하나 이상의 순환 계층들 (571), 및 하나 이상의 포스트-순환 계층들 (572) 은 현재 시간 스텝 $t+1$ 에 대한 출력 데이터를 프로세싱할 수 있고 재구성된 모션 추정 데이터 $\hat{f}_{t+1}$ 및 재구성된 잔차 데이터 $\hat{r}_{t+1}$ 를 출력할 수 있다. 디코더 (538) 의 워핑 엔진 (567) 은 현재 시간 스텝 $t+1$ 에 대한 워핑된 프레임을 생성하기 위해 이전에 재구성된 프레임 $\hat{x}_t$ 을 (예를 들어, 모션 보상을 수행함으로써) 워핑할 수 있다. 가산기 엔진 (569) 은 워핑된 프레임을 재구성된 잔차 데이터 $\hat{r}_{t+1}$ 에 가산하여 현재 시간 스텝 $t+1$ 에 대한 재구성된 프레임 $\hat{x}_{t+1}$ 을 생성할 수 있다.

[0119] 재구성된 프레임 $\hat{x}_{t+1}$, 순환 상태 h_{t+1} , 재구성된 모션 추정 데이터 $\hat{f}_{t+1}$ 및 재구성된 잔차 데이터 $\hat{r}_{t+1}$ 는 동작의 다음 시간 스텝 $t+2$ 에서 인코더 (533) 에 대한 입력으로서 피드백될 수 있다. 뉴럴 네트워크 시스템 (511) 은 다음 I-프레임이 필요할 때까지 (예를 들어, 비디오에서의 소정 시점에 대한 랜덤 액세스를 위해, 누적될 수도 있는 임의의 모션 에러를 리셋하기 위해 등) 프레임 단위로 프레임들의 비디오 시퀀스에 대한 모션 추정 기반 코딩 프로세스를 계속할 수 있다. I-프레임이 필요할 때, 비디오 시퀀스에서의 비디오의 다음 프레임이 뉴럴 네트워크 시스템 (510) 에 의해 인코딩될 수 있다.

[0120] 일부 구현들에서, 인코더 (533), 양자화기 (534), 및 디코더 (538) 는 엔드-투-엔드로 함께 훈련되는 단일 뉴럴 네트워크 또는 다중 뉴럴 네트워크들을 사용하여 구현될 수 있다. 도 5b 는 뉴럴 네트워크 시스템 (511) 에 사용될 수 있는 뉴럴 네트워크 아키텍처의 예를 도시하는 다이어그램이다. 일부 구현들에서, 도 5b 에 나타낸 것과 유사한 뉴럴 네트워크 아키텍처가 뉴럴 네트워크 시스템 (510) 에 대해 사용될 수 있지만, 순환 계층들 (예를 들어, ConvGRU 계층) 없이 사용될 수 있다.

[0121] 도 5b 에 나타낸 바와 같이, 인코더 (533) 로의 입력은 인코더 (533) 에 의해 인코딩될 현재 프레임 $\hat{x}_t$, (예를 들어, 뉴럴 네트워크 시스템 (510) 으로부터 또는 뉴럴 네트워크 시스템 (511) 으로부터) 이전에 재구성된 프레임 $\hat{x}_{t-1}$, 및 이전 순환 상태 h_{t+1} 를 포함한다. 인코더에 제공될 수 있는 다른 입력들은 이전에 재구성된 모션 추정 데이터 $\hat{f}_{t-1}$, 및 이전에 재구성된 잔차 데이터 $\hat{r}_{t-1}$ 를 포함한다.

[0122] 입력들은 제 1 컨볼루션 계층 (Conv c:64 k:5 s:2 로 표기되며, 여기서 c 는 채널들 (예를 들어, 커널들 또는 필터들) 의 수이고, k 는 커널 사이즈이고, s 는 주어진 컨볼루션 계층과 연관된 스트라이드이다) 에 제공된다.

컨볼루션 계층은 예컨대, 수용 필드로 불리는 노드들 (예를 들어, 현재 프레임 $\hat{x}_t$ 의 픽셀들) 의 영역을 프로세싱함으로써, 컨볼루션 방식으로 입력 데이터를 프로세싱한다. 컨볼루션 계층은 하나 이상의 커널 (필터로서 또한 지칭됨) 로 구성될 수 있으며, 여기서 각각의 커널은 상이한 활성화 또는 특징 맵에 대응할 수 있고, 커널의 각각의 컨볼루션 반복은 컨볼루션 계층의 노드 또는 뉴런이다. 예를 들어, 각각의 컨볼루션 반복에

서 커널이 커버하는 현재 프레임 $\hat{x}_t$ 의 영역은 커널에 대한 수용 필드일 것이다. 하나의 예시적인 예에서,

프레임 $\hat{x}_t$ 이 28×28 어레이를 포함하고, 각각의 커널 (및 대응 수용 필드) 가 5×5 어레이 (여기서 $k=5$) 이면, 컨볼루션 계층에 24×24 노드가 있을 것이다. 비디오 프레임은 통상적으로 1080p 비디오에 대해 $1,920 \times 1,080$ 과 같은, 28×28 보다 큰 해상도를 가짐을 유의한다. 노드와 그 노드에 대한 수용 필드 사이의 각각의 연결은 가중치, 그리고 일부 경우들에서는, 전체 바이어스를 학습하여 각각의 노드가 입력에서의 그 특정 로컬 수용 필드를 분석하도록 학습한다. 컨볼루션 계층의 각각의 노드는 동일한 가중치들 및 일부 경우들에서는 바이어스 (공유 가중치 및 공유 바이어스로 불림) 를 가질 수 있다. 예를 들어, 커널은 입력과 동

일한 깊이 및 가중치들 (숫자들) 의 어레이를 갖는다. 일 예에서, 커널은 프레임 $\hat{x}_t$ 의 3개의 컬러 컴포넌

트 (예를 들어, 적색 (R), 녹색 (G), 및 청색 (B) 컬러 컴포넌트들) 에 따라 현재 프레임 $\hat{x}_t$ 에 대해 3 의 값을 가질 수 있다.

[0123] 컨볼루션 계층의 컨볼루션 성질은 컨볼루션 계층의 각각의 노드가 그의 대응하는 수용 필드에 적용되는 것에 기인한다. 예를 들어, 컨볼루션 계층의 커널은 현재 프레임 $\hat{x}_t$ 의 어레이의 상단-좌측 코너에서 시작할 수 있고 프레임 $\hat{x}_t$ 주위에서 컨볼브(convolve)할 수 있다. 위에 언급된 바와 같이, 커널의 각각의 컨볼루션 반복은 컨볼루션 계층의 노드 또는 뉴런으로 간주될 수 있다. 각각의 컨볼루션 반복에서, 커널의 값들은 이미지의 원래 픽셀 값들의 대응하는 수로 승산된다 (예를 들어, 5x5 커널 어레이는 프레임 $\hat{x}_t$ 어레이의 상단-좌측 코너에서 입력 픽셀 값들의 5x5 어레이로 승산된다). 각각의 컨볼루션 반복으로부터의 승산들은 그 반복 또는 노드에 대한 총 합을 획득하기 위해 함께 합산될 수 있다. 프로세스는 컨볼루션 계층에서 다음 노드의 수용 필드에 따라 프레임 $\hat{x}_t$ 의 다음 위치에서 계속된다. 예를 들어, 커널은 (도 5b 에서 s 로 표기된) 스트라이드 양만큼 다음 수용 필드로 이동될 수 있다. 스트라이드 양은 1, 2, 또는 다른 적절한 양으로 설정될 수 있다. 예를 들어, 스텝 양이 1 로 설정되는 경우, 커널은 각각의 컨볼루션 반복에서 1 픽셀만큼 우측으로 이동될 것이다. 입력 블록의 각각의 고유 위치에서 커널을 프로세싱하는 것은 그 위치에 대한 커널 결과들을 나타내는 수를 생성하여, 컨볼루션 계층의 각각의 노드에 대해 결정되는 총 합산 값을 야기한다.

[0124] 입력 계층으로부터 컨볼루션 은닉 계층 (622a) 으로의 매핑은 활성화 맵 (또는 특징 맵) 으로 지칭된다. 활성화 맵은 입력 블록의 각 위치에서의 커널 결과들을 나타내는 각각의 노드에 대한 값을 포함한다. 활성화 맵은 입력 블록 (예를 들어, 현재 프레임 $\hat{x}_t$) 상에서 커널의 각 반복으로부터 야기되는 다양한 총 합산 값들을 포함하는 어레이를 포함할 수 있다. 예를 들어, 28 x 28 입력 프레임의 각각의 픽셀 (1 의 스텝 양) 에 5 x 5 커널이 적용되는 경우 활성화 맵은 24 x 24 어레이를 포함할 것이다. 컨볼루션 계층은 입력의 다중 특징들을 식별하기 위해 여러 활성화 맵들을 포함할 수 있다. 예를 들어, 3개의 활성화 맵을 사용하여, 컨볼루션 계층은 3개의 상이한 종류의 특징들을 검출할 수 있으며, 각각의 특징은 전체 입력에 걸쳐 검출가능하다.

[0125] 제 1 컨볼루션 계층 이후에 정규화 계층 및 비선형 계층이 적용될 수 있다. 정규화 계층은 은닉된 계층 (예를 들어, 컨볼루션 계층) 에서의 입력들을 비선형성들로 정규화하기 위해 (예를 들어, 훈련 동안) 사용될 수 있다. 정규화 계층의 하나의 예시적인 예는 도 5b 의 아키텍처에서 사용된 바와 같은 BatchNorm 을 포함한다.

하나의 예시적인 예에서, BatchNorm 은 다음과 같이 은닉된 계층 z^k 을 정규화할 수 있다:

$$z^{-k} = \frac{z^k - E[z^k]}{\sqrt{V[z^k]}}$$

[0126] , 식 (1).

[0127] 여기서, $E[x]$ 는 x 의 제 1 모멘트 (예를 들어, 평균 또는 기대 값) 이고, $V[x]$ 는 x 의 제2 모멘트 (예를 들어, 분산) 이다.

[0128] 비선형 계층은 선형 연산들을 계산하고 있던 시스템에 비선형성을 도입하는데 사용될 수 있다. 비선형 계층의 하나의 예시적인 예는 도 5b 의 아키텍처에서 사용된 바와 같은, 정류된 선형 유닛 (ReLU) 이다. ReLU 계층은 함수 $f(x) = \max(0, x)$ 를 입력 블록의 모든 값들에 적용할 수 있으며, 이는 모든 음의 활성화들을 0 으로 변경한다. 따라서 ReLU 는 컨볼루션 계층들의 수용 필드들에 영향을 미치지 않으면서 뉴럴 네트워크의 비선형 특성들을 증가시킬 수 있다.

[0129] 제 1 ReLU 의 출력은 컨볼루션, BatchNorm, 및 ReLU 계층들의 제 2 세트에 제공된다. 제 2 ReLU 의 출력은 제 1 잔차 계층 (ResBlock B) 에 제공되며, 이는 제 2 잔차 계층 (ResBlock A) 에 그 출력을 제공한다. ResBlock A 및 ResBlock B 의 상세가 또한 도 5b 에 나타나 있다. ResBlock A 의 출력은 제 2 ReLU 의 출력과 조합 (또는 가산) 되고, 조합된 출력은 제 3 컨볼루션 계층에 제공된다. 제 3 컨볼루션 계층의 출력은 제 3 BatchNorm 계층에 제공되고, 이는 데이터를 양자화 계층 (소프트/하드 양자화로서 표기됨) 에 출력한다.

소프트/하드 양자화 계층은 하나 이상의 양자화된 (그리고 일부 경우들에서는 엔트로피 코딩된) 계수들 z_t 을 출력한다. 하나 이상의 계수 z_t 는 일부 경우들에서 저장될 수 있고, 일부 경우들에서는 디코더 (538) 에 제공될 수 있다.

[0130] 디코더 (538) 로의 입력은 하나 이상의 계수 z_t 이다. 하나 이상의 계수 z_t 는 제 1 컨볼루션 계층 (Tconv c:128 k:3 s:2 로 표기됨) 에 의해 프로세싱된 후, 제 1 BatchNorm 계층 및 제 1 ReLU 계층에 의해 프로세싱된다. 제 1 ReLU 의 출력은 제 1 잔차 계층 (ResBlock B) 에 제공되며, 이는 제 2 잔차 계층 (ResBlock A) 에 그 출력을 제공한다. ResBlock A 의 출력은 제 1 ReLU 의 출력과 조합 (또는 가산) 되고, 조합된 출력은 제 2 컨볼루션 계층 (Tconv c:64 k:5 s:2 로 표기됨) 에 제공된다. 제 2 컨볼루션 계층은 제 2 BatchNorm 계층 및 제 2 ReLU 계층에 후속하며, 이는 데이터를 제 3 컨볼루션 계층 (Tconv c:C₂ k:5 s:2 로 표기됨) 에 출력한 다음 제 3 BatchNorm 계층에 출력한다. 제 3 BatchNorm 계층의 출력은 순환 계층 (ConvGRU c:C₃ k:3 s:1 로 표기됨) 에 제공된다. 나타낸 바와 같이, 순환 계층에 대한 다른 입력은 이전 시간 스텝 t-1 에 대한 이전 순환 상태 h_{t-1} 이고, 순환 계층의 출력은 현재 시간 스텝 t 에 대한 순환 상태를 포함한다. 순환 계층의 출력은 최종 컨볼루션 계층 (Conv c:5 k:5 s:1 로 표기됨) 에 제공되고, 이는 재구성된 모션 추정 데이터 $\hat{f}_t$ 및 재구성된 잔차 데이터 $\hat{r}_t$ 를 출력한다.

[0131] 위에 언급된 바와 같이, 디코더 (538) 의 워핑 엔진 (567) 은 현재 시간 스텝 t 동안 워핑된 프레임을 생성하기 위해 이전에 재구성된 프레임 $\hat{x}_{t-1}$ 을 (예를 들어, 모션 보상을 수행함으로써) 워핑할 수 있고, 가산기 엔진 (569) 은 현재 시간 스텝 t 동안 재구성된 프레임 $\hat{x}_t$ 을 생성하기 위해 워핑된 프레임을 재구성된 잔차 데이터 $\hat{r}_t$ 에 가산할 수 있다. 일부 구현들에서, 워핑 엔진 (567) 및/또는 가산기 엔진 (569) 은 도 7 에 나타낸 인코더의 구성에서와 같은, 뉴럴 네트워크 시스템 (511) 의 일부일 수 있다 (예를 들어, 하나 이상의 컨볼루션 계층들, 하나 이상의 정규화 계층, 하나 이상의 비선형 계층, 하나 이상의 잔차 블록, 또는 다른 뉴럴 네트워크 계층들을 사용하여 구현됨). 일부 구현들에서, 워핑 엔진 (567) 및/또는 가산기 엔진 (569) 은 뉴럴 네트워크 시스템 (511) 외부에 있는 컴포넌트들일 수 있다. 예를 들어, 이러한 구현들에서, 워핑 및 가산은 학습되지 않은 오퍼레이터들이고, 이 경우 이들 동작들은 뉴럴 네트워크들 또는 다른 머신 러닝 시스템들을 사용하여 수행되지 않는다.

[0132] 뉴럴 네트워크 시스템 (511) 은 임의의 적절한 머신 러닝 훈련 기법을 사용하여 훈련될 수 있다. 하나의 예시적인 예에서, 지도 학습 기법들이 뉴럴 네트워크 시스템 (511) 을 훈련하는데 사용될 수 있다. 예를 들어, 역전파 훈련 프로세스는 뉴럴 네트워크의 노드들의 가중치들 (및 일부 경우들에서는 바이어스들과 같은 다른 파라미터들) 을 조정하기 위해 사용될 수 있다. 역전파는 순방향 패스, 손실 함수, 역방향 패스, 및 가중치 업데이트를 포함할 수 있다. 순방향 패스, 손실 함수, 역방향 패스, 및 파라미터 업데이트는 하나의 훈련 반복을 위해 수행된다. 프로세스는 뉴럴 네트워크 시스템 (511) 의 파라미터들의 가중치들이 정확하게 튜닝될 때까지 각각의 훈련 데이터의 세트에 대해 소정 수의 반복을 위해 반복될 수 있다.

[0133] 하나의 예시적인 예에서, 뉴럴 네트워크 시스템 (511) 을 훈련하기 위해 사용되는 훈련 데이터는 입력으로서 비디오 프레임들 x_t (인코딩/압축되지 않음) 및 알려진 출력으로서 재구성된 비디오 프레임들 $\hat{x}_t$ (입력 비디오 프레임들 x_t 의 재구성된 버전들) 을 포함할 수 있다. 순방향 패스는 입력 비디오 프레임을 뉴럴 네트워크를 통해 전달하는 것을 포함할 수 있다. 가중치들은 뉴럴 네트워크가 훈련되기 전에 초기에 랜덤화될 수 있다. 뉴럴 네트워크 시스템에 대한 제 1 훈련 반복에 대해, 출력은 가중치들이 아직 교정되지 않았기 때문에, 임의의 특정 출력에 대한 선호도를 부여하지 않는 값들을 포함할 수도 있다. 예를 들어, 출력은 재구성된 비디오 프레임 $\hat{x}_t$ 을 나타내는 값들을 갖는 데이터 표현 (예를 들어, 벡터, 텐서 등) 을 포함할 수 있다. 초기 가중치들을 사용한 제 1 훈련 반복 후에, 재구성된 비디오 프레임 $\hat{x}_t$ 은 대응하는 입력 프레임

x_t 과 유사할 가능성이 없을 것이다.

[0134] 손실 함수는 출력에서의 에러를 분석하는데 사용될 수 있다. 압축되지 않은 비디오 프레임들을 입력으로서 사용하고 재구성된 비디오 프레임들 $\hat{x}_t$ 을 알려진 출력으로서 사용하는 예에서, 도 5a 에 나타난 손실 함수 $Loss0 = \sum_t distortion(x_t, \hat{x}_t)$ 는 뉴럴 네트워크 시스템 (511) 을 훈련하는데 사용될 수 있다. 다른 훈련 데이터가 사용될 때 다른 손실 함수들이 사용될 수 있다. 다른 손실 함수의 일 예는

$E_{total} = \sum \frac{1}{2} (target - output)^2$ 로서 정의된, 평균 제곱 에러 (MSE) 를 포함한다. MSE 는 실제 답변 마이너스 예측된 (출력) 답변을 제곱한 것의 1/2 의 합을 계산한다.

[0135] 실제 출력 값들 (네트워크에 의해 출력된 재구성된 프레임) 이 예측된 출력 (입력 프레임의 알려진 재구성된 버전) 과 매우 상이할 수도 있기 때문에, 손실 (또는 에러) 은 제 1 훈련 비디오 프레임들에 대해 높을 수도 있다. 훈련의 목표는 예측된 출력에 대한 손실의 양을 최소화하는 것이다. 뉴럴 네트워크는 뉴럴 네트워크의 손실에 가장 기여한 입력들 (가중치들) 을 결정함으로써 역방향 패스를 수행할 수 있으며, 가중치들을 조정하여 손실이 감소하고 결국 최소화될 수 있다. 가중치들과 관련하여 손실의 도함수 (derivative) (dL/dW , 로 표기됨, 식 중 W 는 특정 계층에서의 가중치들이다) 는 뉴럴 네트워크의 손실에 가장 기여한 가중치들을 결정하기 위해 계산될 수 있다. 예를 들어, 가중치들은 업데이트될 수 있어서 그래디언트

의 반대 방향으로 변화한다. 가중치 업데이트는
$$w = w_i - \eta \frac{dL}{dw}$$
 로서 표기될 수 있으며, 식중 W 는 가중치를 표기하고, w_i 는 초기 가중치를 표기하며, η 는 러닝 레이트를 표기한다. 러닝 레이트는 임의의 적절한 값으로 설정될 수 있으며, 높은 러닝 레이트는 더 큰 가중치 업데이트들을 포함하고, 낮은 값은 더 작은 가중치 업데이트들을 표시한다. 뉴럴 네트워크 시스템 (511) 은 원하는 출력이 달성될 때까지 이러한 방식으로 계속 훈련될 수 있다. 일부 경우들에서, 뉴럴 네트워크 시스템 (510) 은 유사한 방식으로 훈련될 수 있다.

[0136] 도 6 은 모션 보상 기반 뉴럴 네트워크 시스템 (예를 들어, 뉴럴 네트워크 시스템 (411) 및/또는 뉴럴 네트워크 시스템 (511)) 에 의해 수행될 수 있는 상이한 타입들의 모션 추정들을 도시하는 다이어그램이다. 도 6 에서, 용어 x 는 (현재 프레임이 인코딩되고 있을 때 이전 프레임에 대응하는) 모션이 추정될 수 있는 프레임들을 표기하고, 용어 f 는 모션 추정을 표기하며, 용어 y 는 모션 보상된 프레임을 표기하며, 이는 다음과 같이 계산될 수 있다: $y = f(x)$.

[0137] 수행될 수 있는 모션 추정의 일 타입은 블록-기반 모션 추정 기법 (602) 이다. 블록-기반 모션 추정은 블록 단위로 수행될 수 있다. 예를 들어, (모션 보상될) 현재 프레임 y 에서의 각각의 블록에 대해, 모션 추정 f 는 이전 프레임 x 에서의 대응하는 블록의 위치를 정의한다. 하나의 예시적인 예에서, 모션 추정 f 는 이전 프레임 (x) 에서의 대응하는 블록에 대한 현재 프레임 (y) 에서의 블록의 변위 (예를 들어, 수평 및 수직 변위) 를 표시하는 모션 벡터를 포함할 수 있다. 이전 프레임 x 로부터의 블록은 블록들 사이의 유사성 (예를 들어, 픽셀 값들의 유사성) 을 결정함으로써 현재 프레임 y 에서의 블록에 대응하는 것으로 결정될 수 있다.

[0138] 수행될 수 있는 다른 타입의 모션 추정은 광학 흐름 모션 추정 기법 (604) 이다. 광학 흐름 모션 추정은 픽셀 단위로 수행될 수 있다. 예를 들어, 현재 프레임 y 에서의 각각의 픽셀에 대해, 모션 추정 f 는 이전 프레임 x 에서의 대응하는 픽셀의 위치를 정의한다. 각각의 픽셀에 대한 모션 추정 f 는 프레임들 사이의 픽셀의 움직임을 표시하는 벡터를 포함할 수 있다. 일부 경우들에서, 광학 흐름 맵들 (또한 모션 벡터 맵들로서 지칭됨) 은 프레임들 사이의 광학 흐름 벡터들의 계산에 기초하여 생성될 수 있다. 광학 흐름 맵들은 프레임에서의 각각의 픽셀에 대한 광학 흐름 벡터를 포함할 수 있으며, 여기서 각각의 벡터는 프레임들 사이의 픽셀의 이동을 표시한다. 하나의 예시적인 예에서, 픽셀에 대한 광학 흐름 벡터는 제 1 프레임으로부터 제 2 프레임으로의 픽셀의 이동을 나타내는 변위 벡터 (예를 들어, x - 및 y - 변위들과 같은 수평 및 수직 변위들을 표시함) 일 수 있다. 모션 보상 기반 뉴럴 네트워크 시스템의 인코더 (예를 들어, 인코더 (533)) 는 현재 프레임 y 에서의 각각의 픽셀에 대한 광학 흐름 벡터를 인코딩 또는 압축할 수 있거나, 광학 흐름 맵들이 사용될 때 광학 흐름 맵을 인코딩 또는 압축할 수 있다. 모션 보상 기반 뉴럴 네트워크 시스템의 디코더 (예를 들어,

디코더 (538)) 에 의해 재구성되는 모션 추정 정보 (f) 는 재구성된 광학 흐름 벡터들의 그룹 (예를

들어, 프레임에서의 각각의 픽셀에 대해 하나) 또는 재구성된 광학 흐름 맵을 포함할 수 있다.

[0139] 일부 경우들에서, 광학 흐름 맵은 프레임에서의 모든 픽셀들보다 적은 픽셀들에 대한 벡터들을 포함할 수 있다. 예를 들어, 조밀한 광학 흐름 맵에 포함될 수 있는 프레임에서의 각각의 픽셀에 대한 광학 흐름 벡터를 생성하기 위해 인접한 프레임들 사이에서 조밀한 광학 흐름이 계산될 수 있다. 다른 예에서, 광학 흐름 맵에 포함될 수 있는, 프레임에서의 일부 또는 모든 픽셀들에 대한 광학 흐름 벡터들을 생성하기 위해 인접한 프레임들 사이에서 Lucas-Kanade 광학 흐름이 계산될 수 있다. 임의의 다른 적합한 타입의 광학 흐름 기법 또는 알고리즘이 프레임들 사이의 광학 흐름을 결정하는데 사용될 수 있다. 각각의 광학 흐름 맵은 2D 벡터 필드를 포함할 수 있고, 각각의 벡터는 제 1 프레임으로부터 제2 프레임으로의 포인트들의 이동을 나타내는 변위 벡터이다.

[0140] 위에 언급된 바와 같이, 광학 흐름 벡터 또는 광학 흐름 맵들은 프레임들의 시퀀스의 인접한 프레임들 사이에서 (예를 들어, 인접한 프레임들 x_t 및 x_{t-1} 의 세트들 사이에서) 계산될 수 있다. 2개의 인접한 프레임은 연속적으로 캡처된 프레임들인 2개의 직접 인접한 프레임 또는 프레임들의 시퀀스에서 (예를 들어, 서로의 2개의 프레임 내에서, 서로 3개의 프레임 내에서, 또는 다른 적절한 거리로) 소정의 거리만큼 떨어진 2개의 프레임을 포함할 수 있다. 프레임 x_{t-1} 에서 프레임 x_t 로의 광학 흐름은 $O_{x_{t-1}, x_t} = \text{dof}(x_{t-1}, x_t)$ 에 의해 주어질 수 있으며, 식중 dof 는 조밀한 광학 흐름이다. 광학 흐름 맵들을 생성하기 위해 임의의 적합한 광학 흐름 프로세스가 사용될 수 있다. 하나의 예시적인 예에서, 프레임 x_{t-1} 에서의 픽셀 $I(x, y, t)$ 은 다음 프레임 x_t 에서의 거리 $(\Delta x, \Delta y)$ 만큼 이동할 수 있다. 픽셀들이 동일하고 강도가 프레임 x_{t-1} 과 다음 프레임 x_t 사이에서 변경되지 않는다고 가정하면, 다음의 식이 가정될 수 있다:

[0141]
$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$
 식 (2).

[0142] 식 (2) 의 우변의 테일러 급수 근사 (Taylor series approximation) 를 취한 다음, 공통 항들을 제거하고 Δt 로 나눔으로써, 광학 흐름 식이 다음과 같이 유도될 수 있다:

[0143]
$$f_x u + f_y v + f_t = 0, \text{ 식 (3),}$$

[0144] 식 중:

$$f_x = \frac{df}{dx},$$

$$f_y = \frac{df}{dy},$$

$$f_t = \frac{df}{dt},$$

[0145]
$$u = \frac{\Delta x}{\Delta t},$$
 및

[0147]
$$v = \frac{\Delta y}{\Delta t}.$$

[0148] 광학 흐름 식 (3) 을 사용하여, 이미지 그래디언트들 f_x 및 f_y 은 시간에 따른 그래디언트와 함께 발견될 수

있다 (f_t 로서 표기됨). 용어들 u 및 v 는 $I(x, y, t)$ 의 속도 또는 광학 흐름의 x 및 y 컴포넌트들이며, 알려져 있지 않다. 광학 흐름 식이 2개의 알려지지 않은 변수로 해결될 수 없는 일부 경우들에서 추정 기법이 필요할 수도 있다. 광학 흐름을 추정하기 위해 임의의 적합한 추정 기법이 사용될 수 있다. 이러한 추정 기법들의 예들은 자동 방법들 (예를 들어, Lucas-Kanade 추정, Horn-Schunck 추정, Buxton-Buxton 추정, 또는 다른 적합한 자동 방법), 위상 상관, 블록-기반 방법들, 또는 다른 적합한 추정 기법을 포함한다. 예를 들어, Lucas-Kanade 는 광학 흐름 (이미지 픽셀의 변위) 이 픽셀 I 의 로컬 이웃에서 작고 대략 일정하다고 가정하고, 최소 자승법을 사용하여 그 이웃에서의 모든 픽셀들에 대한 기본 광학 흐름 식들을 푼다.

[0149] 수행될 수 있는 다른 타입의 모션 추정은 동적 컨볼루션 (또한, 적응 컨볼루션으로 지칭됨) 모션 추정 기법 (606) 이다. 동적 컨볼루션 모션 추정은 픽셀 단위로 수행될 수 있다. 예를 들어, 현재 프레임 y 에서의 각각의 픽셀에 대해, 모션 추정 f 는 이전 프레임 x 에서의 픽셀들의 이웃의 선형 조합을 정의한다. 도 6 에 나타난 바와 같이, 현재 프레임 y 에서의 픽셀 위치에 있는 픽셀에 대해, 커널 또는 행렬은 이전 프레임 x 에서의 동일한 픽셀 위치에 있는 픽셀을 둘러싸는 픽셀들의 이웃 주위에 정의된다. 도 6 의 예시의 동적 컨볼루션 모션 추정은 현재 프레임 y 에서의 위치 (2, 7) ((행 번호, 열 번호) 에 대응함) 에서의 픽셀에 대해 수행되며, 여기서 현재 프레임 y 에서의 상단-좌측 픽셀 위치는 위치 (0, 0) 에 있다. 커널은 이전 프레임 x 에서 위치 (2, 7) 에 있는 픽셀 주위에 정의된다. 예를 들어, 나타난 바와 같이, 사이즈 5x5 의 커널은 이전 프레임 x 에서의 위치 (2, 7) 에 있는 픽셀을 둘러싸는 픽셀들의 이웃에 적용된다.

[0150] 커널 또는 행렬의 각각의 엔트리는 숫자 (도 6의 예에서 5x5 숫자들의 세트) 를 포함할 수 있다. 커널에서의 숫자들의 세트는 새롭게 재구성된 프레임 (현재 프레임 y) 에서 재구성되는 픽셀의 동일한 픽셀 위치에 대응하는 이전 프레임 x 에서의 픽셀 위치에 있는 픽셀을 둘러싸는 픽셀들의 이웃으로부터의 픽셀들을 조합하는 방법을 설명할 수 있다. 모션 보상 기반 뉴럴 네트워크 시스템의 인코더 (예를 들어, 인코더 (533)) 는 현재 프레임 y 에서의 각각의 픽셀에 대한 커널로부터의 숫자들의 표현 (예를 들어, 도 6 으로부터의 5x5 커널로부터의 25개의 숫자들) 을 압축할 수 있다. 모션 보상 기반 뉴럴 네트워크 시스템의 디코더 (예를 들어, 디코더

(538)) 에 의해 재구성되는 모션 추정 정보 ($\hat{f}$) 는 각각의 픽셀에 대해 재구성된 커널을 포함할 수 있으며, 이는 이전에 재구성된 프레임을 워핑하기 위해 데이터 컨볼루션을 수행하는데 사용될 수 있다.

[0151] 도 7 은 포스트-워핑 강화를 수행하기 위한 부가적인 포스트-워핑 네트워크 (574) 를 갖는 뉴럴 네트워크 시스템 (511) 의 예를 도시하는 다이어그램이다. 일부 경우들에서, 워핑 엔진 (567) 이 워핑을 수행하고 가산기 엔진 (569) 이 잔차들을 가산한 후에, 재구성된 프레임을 개선할 여지가 여전히 있을 수도 있다. 예를

들어, (워핑 엔진 (567) 에 의한) 워핑 및 (가산기 엔진 (569) 에 의한) 잔차의 가산 후의 재구성된 프레임 $\hat{x}_t$ 은 원래의 프레임 $\hat{x}$ 에 존재하지 않았던 시각적 아티팩트들을 가질 수도 있다. 포스트-워핑 네트워크 (574) 는 (예를 들어, 아티팩트들을 제거함으로써) 재구성된 프레임을 추가로 강화 또는 정제하는데 사용되는 뉴럴 네트워크 컴포넌트들 (예를 들어, 하나 이상의 컨볼루션 계층) 을 포함할 수 있다.

[0152] 포스트-워핑 네트워크 (574) 는 워핑 엔진 (567) 이 이전에 재구성된 프레임을 워핑한 후에 적용될 수 있다. 예를 들어, 시간 스텝 t 에서 도 7 에 나타난 바와 같이, 워핑된 프레임 (워핑이 적용된 후) 및 재구성된 잔차 $\hat{r}_t$ 는 포스트-워핑 네트워크 (574) 에 입력된다. 포스트-워핑 네트워크는 임의의 시각적 불순물들 (예

를 들어, 시각적 아티팩트들) 이 제거된 재구성된 프레임 $\hat{x}_t$ 을 생성할 수 있다. 일부 예들에서, 포스트-워핑 네트워크 (574) 는 뉴럴 네트워크 시스템 (511) 의 나머지의 뉴럴 네트워크와 독립적으로 훈련된다. 일부 예들에서, 포스트-워핑 네트워크 (574) 는 뉴럴 네트워크 시스템 (511) 의 나머지와 독립적으로 훈련되지 않으며, 이 경우 전체 뉴럴 네트워크 시스템 (511) 은 엔드-투-엔드로 훈련된다. 예를 들어, 클린 또는 압축되지 않은 프레임이 뉴럴 네트워크 시스템 (511) 의 뉴럴 네트워크에 입력될 수 있고, 네트워크는 재구성된 프레임을 출력한다. (입력 프레임과 비교하여) 재구성된 프레임과 연관된 에러 또는 손실을 평가하기 위해 역전파가 사용될 수 있고, 상술한 바와 같이, 뉴럴 네트워크 시스템 (511) 의 파라미터들이 튜닝될 수 있다.

[0153] 도 8 및 도 9 는 추가적인 모션 추정 네트워크 (575) 를 갖는 뉴럴 네트워크 시스템 (511) 의 예를 예시하는 다이어그램들이다. 모션 추정 네트워크 (575) 는 모션 추정들의 강화된 학습을 제공할 수 있다. 도 8 및 도 9 에 나타낸 바와 같이, 시간 스텝 $t+1$ 에서 뉴럴 네트워크 시스템 (511) 의 인코더 측에 (모션 추정 네트워크 (575) 에 입력되는) 이전 입력 프레임 x_t , 모션 추정 네트워크 (575) 에 의해 인코더 측에서 결정된 모션 추정 f_{t+1}^{enc} , 및 모션 추정 f_{t+1}^{enc} 에 의해 위핑된 이전 재구성된 프레임 (위핑된 재구성된 프레임 $f_{t+1}^{enc}(\hat{x}_t)$ 으로 표기됨) 을 포함하는 추가적인 입력 변수들이 제공된다.

[0154] 위에 언급된 바와 같이, 일부 구현들에서, 뉴럴 네트워크 시스템 (511) 은 엔드-투-엔드 방식으로 훈련된다. 일부 경우들에서, 뉴럴 네트워크 시스템 (511) 은 소정의 파라미터들을 계산하려고 시도할 때 훈련 동안 문제들을 가질 수 있다. 예를 들어, 뉴럴 네트워크가 모션 추정 파라미터들 (예를 들어, 광학 흐름) 을 결정하고 이들 모션 추정 파라미터들을 사용하여 이전에 재구성된 프레임을 위핑하려고 시도할 때 문제들이 발생할 수 있다. 다른 예에서, 잔차를 추정하고 위핑된 프레임을 잔차에 부가할 때 문제들이 발생할 수 있다. 이러한 문제들은 뉴럴 네트워크 시스템 (511) 이 재구성이 어떻게 획득되는지에 관계없이 가능한 최상의 재구성을 획득하려고 시도하는 것으로 인해 발생할 수 있다. 예를 들어, 가능한 원래 프레임에 가까운 재구성된 프레임을 생성하려고 시도할 때, 뉴럴 네트워크 시스템 (511) 은 모션 추정 파라미터들 (예를 들어, 광학 흐름) 을 적절하게 학습하지 않을 수도 있고, 재구성된 프레임을 생성하기 위해 잔차에만 의존할 수도 있다. 일부 경우들에서, 재구성된 광학 흐름 출력은 심지어 데이터가 없는 블랭크 텐서 또는 다른 데이터 구조일 수도 있다. 그러나, 비디오에서의 리던던시들의 대부분은 그러한 모션 데이터로부터 (예를 들어, 프레임들 사이의 광학 흐름으로부터) 을 수 있으며, 이 경우 모션 추정 파라미터들을 학습하는 것이 중요할 수 있다.

[0155] 모션 추정 네트워크 (575) 는 뉴럴 네트워크 시스템 (511) 의 인코더 측 상에 모션 추정 뉴럴 네트워크를 제공함으로써 뉴럴 네트워크 시스템 (511) 이 모션 추정 파라미터들을 학습하는 것을 도울 수 있다. 예를 들어, 모션 추정 네트워크 (575) 는 모션 추정 파라미터들 (예를 들어, 광학 흐름 벡터들, 광학 흐름 맵들, 블록 기반 모션 정보, 데이터 컨볼루션을 수행하기 위해 사용된 각각의 픽셀에 대한 커널 등) 을 추정하고, 이들 모션 추정 파라미터들을 인코더 (533) 에 제공하는 별도의 뉴럴 네트워크를 포함할 수 있다. 뉴럴 네트워크 시스템 (511) 의 인코더 측 상에서 모션 추정 파라미터들을 결정함으로써, 모션 추정 파라미터들의 실제 세트가 인코더 (533) 에 제공되어, 인코더 (533) 가 모션 추정 (예를 들어, 광학 흐름) 을 고려하게 한다.

[0156] 일부 예들에서, 모션 추정 네트워크 (575) 는 뉴럴 네트워크 시스템 (511) 의 나머지로써 독립적으로 훈련될 수 있거나 또는 전체 뉴럴 네트워크 시스템 (511)(모션 추정 네트워크 (575) 를 포함함) 은 엔드-투-엔드 방식으로 함께 훈련될 수 있다. 모션 추정 네트워크 (575) 를 고려한 다양한 손실 항들은 다음을 포함할 수 있다:

$$\text{Loss1} = \sum_t \text{distortion}(f_t^{enc}(x_{t-1}), x_t)$$

$$\text{Loss2} = \sum_t \text{distortion}(f_t^{enc}(\hat{x}_{t-1}), x_t)$$

$$\text{Loss3} = \sum_t \text{distortion}(\hat{f}_t(x_{t-1}), x_t)$$

$$\text{Loss4} = \sum_t \text{distortion}(\hat{f}_t(\hat{x}_{t-1}), x_t)$$

[0157]

[0158] 일부 구현들에서, 모션 추정 네트워크 (575) 는 (실측 자료 출력들을 사용하지 않으면서) 비지도 학습을 사용하여 훈련될 수 있다. Loss1 및 Loss2 항들은 인코더 측 상에서 사용될 수 있고, Loss3 및 Loss4 항들은 디코더 측 상에서 사용될 수 있다. 모션 추정의 예로서 광학 흐름을 사용하는 하나의 예시적인 예에서, 2개의 연속적인 프레임 x_t 및 x_{t+1} 이 주어지면, 모션 추정 네트워크 (575) 의 목표는 2개의 프레임 x_t 와 x_{t+1}

사이의 광학 흐름을 추정하는 것이다. 모션 추정 네트워크 (575) 는 추정된 광학 흐름 파라미터들을 출력할 것이며, 이는 $x_t, \hat{x}_t$ 의 재구성된 버전을 워핑하는데 사용될 것이다. 손실 (예를 들어, 상기 Loss2) 은 워핑된 프레임 $\hat{x}_t$ 과 원본 프레임 x_t 사이에서 결정될 수 있다. 그 후, 손실은 상술한 역전과 기법들을 사용하여 가중치들 (및 일부 경우들에서 다른 파라미터들) 을 튜닝하는데 사용될 수 있다. 일부 경우들에서, 손실 항 Loss1 은 Loss2 보다는 인코더 측 상에서 사용될 수 있다.

[0159] 인코더 측 상에서 추정된 모션 추정 파라미터들 (예를 들어, 광학 흐름) 을 복구하기 위해 디코더 (538) 를 훈련하기 위해 Loss3 또는 Loss4 를 사용하는 디코더 측에서 동일한 개념이 사용될 수 있다. 예를 들어, 디코더 (538) 는 현재 프레임 x_t 및 이전 프레임 x_{t-1} (Loss3 이 사용될 때) 또는 이전에 재구성된 프레임 $\hat{x}_{t-1}$ (Loss4 가 사용될 때) 을 포함하는 2개의 연속적인 프레임들사이의 모션 (예를 들어, 광학 흐름) 을 계산 또는 추정할 수 있다. 이러한 방식으로, 압축되지 않은 프레임들이 모션 추정 파라미터들을 추정하는데 사용될 수 있다. 결과의 모션 추정 파라미터들은 프레임 x_t 을 매칭하기 위한 시도로 이전 프레임 (x_{t-1} 또는 $\hat{x}_{t-1}$) 을 워핑하기 위해 사용될 수 있다. 워핑의 출력은 현재 시간 스텝에서의 입력 프레임 x_t 과 매칭되는 것이 바람직하다.

[0160] 그 후 워핑된 프레임과 입력 프레임 x_t 사이의 손실은 Loss3 또는 Loss4 를 사용하여 계산될 수 있다. 위 예 언급된 바와 같이, Loss3 은 실제 이전 프레임 x_{t-1} 의 사용을 포함하고, Loss4 는 이전에 재구성된 프레임 $\hat{x}_{t-1}$ (이전 프레임 x_{t-1} 의 압축해제 또는 디코딩된 버전) 의 사용을 포함한다. 디코더 측 상의 Loss4 (및 인코더 측 상의 Loss2) 는 시간 스텝 t-1 로부터의 실제 이전 프레임 x_{t-1} 이 다음 시간 스텝 t 에서 뉴럴 네트워크 시스템 (511) 에 이용 가능하지 않은 상황에서 사용될 수 있다. 역전과는 손실을 평가하고 뉴럴 네트워크 시스템 (511) 의 파라미터들을 조정하거나 튜닝하는데 사용될 수 있다. 이러한 훈련은 디코더 (538) 가 모션 추정 파라미터들을 학습하는 것을 돕는다.

[0161] 일부 예들에서, 손실 항들 Loss1 및 Loss3 또는 손실 항들 Loss2 및 Loss4 (실제 이전 프레임 x_{t-1} 이 시간 스텝 t 에서 이용가능한지 여부에 의존함) 은 훈련의 상이한 스테이지들에서 활성화된다. 예를 들어, 인코더 측 상의 Loss2 및 디코더 측 상의 Loss4 는 뉴럴 네트워크 시스템 (511) 의 훈련이 먼저 시작할 때 그리고 소정의 포인트까지 활성화되고 사용될 수 있으며, 그 후에 뉴럴 네트워크 시스템 (511) 은 Loss0 (도 5a 및 위에 나타난 바와 같음) 을 손실 항 (손실 항들 Loss1-Loss4 는 비활성화됨) 으로서만 사용하여 훈련된다. 훈련 동안 초기에 손실 항들 (Loss1 및 Loss3 또는 Loss2 및 Loss4) 을 사용함으로써, 뉴럴 네트워크 시스템 (511) 은 모션 추정 파라미터들 (예를 들어, 광학 흐름 벡터들 또는 맵들) 을 학습하도록 강제된다. 뉴럴 네트워크 시스템 (511) 이 모션 추정 파라미터들을 충분히 학습하면, 이것은 그 정보를 보유했을 수 있고, 추가 훈련 반복들에서 그리고 추론 동안 (훈련된 네트워크가 실제 비디오 데이터에 대한 동작을 위해 배치될 때) 모션 추정 파라미터들을 추정하는 방법을 유지할 것이다. 따라서, 부가적인 손실 항들 (Loss1-Loss4) 은 모션 추정 파라미터들이 뉴럴 네트워크 시스템 (511) 에 의해 학습되면 비활성화될 수 있고, Loss0 은 뉴럴 네트워크 시스템 (511) 을 추가로 훈련하기 위해 계속 사용될 수 있다.

[0162] 도 10 은 단일 레벨 순환 구성 및 다중 레벨 순환 구성을 포함하는, 뉴럴 네트워크 시스템 (511) 의 상이한 순환 계층 구성들의 예를 도시하는 다이어그램이다. 인코더 (533) 및 디코더 (538) 는 보틀네크 (1034) 에 의해 분리된 상이한 계층들로서 예시된다. 보틀네크 (1034) 는 양자화기 (534) 를 사용하여 양자화되는 잠재성 표현에 대응한다. 단일-레벨 순환 구성에서, 인코더 (533) 는 순방향 계층들만을 포함하고, 디코더는 4 개의 순방향 계층들 및 단일 순환 계층 (1071) 을 포함한다. 인코더 (533) 의 순방향 계층들은 하나 이상의 컨볼루션 계층들, 정규화 계층들, 비선형 계층들, 잔차 블록들, 또는 다른 뉴럴 네트워크 계층들과 같은, 도 5b

에 도시된 뉴럴 네트워크 계층들 중 하나 이상을 포함할 수 있다. 순방향 계층 (1070a), 순방향 계층 (1070b), 및 순방향 계층 (1070c) 은 도 5a 에 나타난 하나 이상의 프리-순환 디코더 계층들 (570) 에 대응한다. 하나의 예시적인 예에서, 순방향 계층들 (1070a, 1070b, 및 1070c) 은 순환 계층 (도 5b 에서 ConvGRU $c:C_3 k:3 s:1$ 로 표기됨) 이전에 발생하는 도 5b 의 디코더 (538) 의 모든 계층들 (도 5b 에서 계층들 (579) 로 표기됨) 을 포함할 수 있다. 도 10 에서 순환 계층 (1071) 은 도 5a 에 나타난 하나 이상의 순환 계층 (571) 에 대응한다. 순환 계층 (1071) 의 예는 도 5b 로부터의 ConvGRU 순환 계층 (설정들 $c:C_3 k:3 s:1$ 을 가짐) 이다. 나타낸 바와 같이, 시간 스텝 t 로부터의 순환 계층 (1071) 의 순환 상태는 다음 시간 스텝 $t+1$ 에서 순환 계층 (1071) 및 인코더 (533) 에 제공된다. 순방향 계층 (1073) 은 도 5a 에 나타난 하나 이상의 포스트-순환 계층 (572) 에 대응한다. 하나의 예시적인 예에서, 순방향 계층 (1073) 은 도 5b 에서 디코더 (538) 의 컨볼루션 계층 (Conv $c:5 k:5 s:1$ 로 표기됨) 을 지속할 수 있다.

[0163] 멀티-레벨 순환 구성은 단일-순환 구성과 동일한 순방향 계층들을 포함하지만, 디코더 (538) 에서 다중 레벨들의 순환 계층들을 포함한다. 상이한 순환 계층들의 각각은 별도의 순환 상태 (은닉된 상태로 또한 지칭됨) 를 유지하고, 각각의 순환 상태는 뉴럴 네트워크 시스템 (511) 의 그 지점까지 상이한 레벨의 특징들을 캡처한다. 도 10 에 나타낸 바와 같이, 디코더 (538) 의 순환 계층들로 피드백되는 것에 부가하여, 순환 상태들은 디코더 (538) 의 순환 계층들로부터 인코더 (533) 에서의 미러링된 위치들로 피드백된다. 다중 순환 상태들은 인코더 (533) 및 디코더 (538) 가 더 효율적인 비디오 코딩을 수행하기 위해 부가 정보를 이용하게 할 수 있다.

[0164] 도 11 은 부가적인 워핑 엔진 (576) 을 갖는 뉴럴 네트워크 시스템 (511) 의 예를 도시하는 다이어그램이다. 뉴럴 네트워크 시스템 (511) 의 다양한 컴포넌트들 (예를 들어, 양자화기 (534), 하나 이상의 프리-순환 디코더 계층들 (570) 등) 이 예시적인 목적들을 위해 도 11 로부터 생략되지만, 당업자는 이들 컴포넌트들이 뉴럴 네트워크 시스템 (511) 에 존재한다는 것을 이해할 것이다. 워핑 엔진 (576) 은 현재 시간 스텝 (예를 들어, 시간 스텝 $t+1$) 에서 순환 상태(들)을 사용하기 전에 하나 이상의 순환 계층 (571) 의 이전 시간 스텝 (예를 들어, 시간 스텝 t) 로부터 이전 순환 상태(들)(또는 숨겨진 상태)를 워핑하기 위해 사용될 수 있다. 워핑 엔진 (576) 은 현재 프레임에서 픽셀들의 위치에 순환 상태를 매칭하기 위해서 순환 상태를 워핑하기 위해 인코더 (533) 및 디코더 (538) 에 이용가능한 모션 추정 파라미터들을 사용할 수 있다.

[0165] 현재 시간 스텝에서 이전 반복 상태(들)를 사용하기 전에 이들을 워핑하는 것이 비디오에 유용할 수 있다. 예를 들어, 오브젝트들은 비디오 프레임들에 걸쳐 시간적으로 이동할 수 있고, 이러한 모션은 (예를 들어, 광학 흐름, 블록-기반 모션 추정 등을 사용하여) 상술한 모션 추정 기법들에 의해 모델링될 수 있다. 디코더 (538) 의 하나 이상의 순환 계층 (571) 의 순환 상태들이 나중 시간 스텝들에서 디코더 (538) 및 인코더 (533) 에 직접 피드될 때, 오브젝트 모션은 (예를 들어, 도 5a 에 나타낸 바와 같이, 순환 계층들 이후에 수행되는 모션 보상으로 인해) 은닉된 상태들 내에서 모델링되지 않는다. 순환 상태는 입력 프레임을 나타내며 (예를 들어, 입력 이미지의 특징 표현을 포함함), 이 경우 입력 이미지에서의 픽셀의 이동은 순환 상태의 특징 공간에서 유사한 이동을 야기해야 한다. 따라서, 모션 추정에 기초하여 순환 상태를 워핑함으로써, 입력 프레임으로부터 추정된 각각의 움직임이 순환 상태에서의 대응하는 움직임으로 전달되어, 순환 상태가 현재 프레임에서의 픽셀들의 위치와 매칭하게 한다.

[0166] 도 11 을 참조하면, 현재 시간 스텝에서의 워핑 엔진 (576) 은 이전 시간 스텝로부터 디코더 (538) 의 순환 계층의 이전 순환 또는 은닉된 상태 h 를 워핑하기 위해 사용된다. 이전의 은닉된 상태는 현재 시간 스텝에서의 순환 계층에서 재사용되기 전에 현재 시간 스텝에서 추정된 모션 추정 파라미터들 (예를 들어, 광학 흐름) 을 사용하여 워핑된다. 예를 들어, 시간 스텝 $t+1$ 에서의 워핑 엔진 (576) 은 시간 스텝 $t+1$ 에서 추정된 모션 추정 파라미터들 $\hat{f}_{t+1}$ 을 사용하여 시간 스텝 t 로부터의 이전 순환 상태 h_t 를 워핑할 수 있다. 워핑된 순환 상태는 현재 시간 스텝 t 에서의 순환 계층에서 재사용될 수 있다.

[0167] 일부 구현들에서, 부가적인 순환 계층이 워핑 엔진 (576) 이전에 제공될 수 있다. 부가적인 순환 계층은 동작의 현재 시간 스텝 $t+1$ 에 대한 모션 추정 파라미터들 $\hat{f}_{t+1}$ (예를 들어, 광학 흐름) 을 재구성하기 위해 사용될 수 있으며, 이는 시간 스텝 t 로부터 이전 순환 상태 h_t 를 워핑하기 위해 워핑 엔진 (576) 에 의해

사용될 수 있다. 그 후 위핑된 이전 순환 상태 $\hat{f}_t(\hat{h}_t)$ 는 현재 시간 스텝 t 에서 순환 계층에 입력될 수 있다. 일부 구현들에서, 현재 시간 스텝 $t+1$ 에서 모션 추정 파라미터들 (예를 들어, 광학 흐름) 을 먼저 결정하기 위해 반복 동작이 수행될 수 있으며, 이는 시간 스텝 t 로부터 이전 순환 상태 h_t 를 위핑하기 위해 위핑 엔진 (576) 에 의해 사용될 수 있다. 그 후, 위핑된 이전 순환 상태 $\hat{f}_t(\hat{h}_t)$ 는 현재 시간 스텝 t 에서 순환 계층에 입력될 수 있다. 예를 들어, 순방향 패스는 모션 추정 파라미터들을 복구 또는 재구성하기 위해 수행될 수 있고, 그 후 동작들은 이전 순환 상태를 위핑하기 위해 순환 계층으로 리턴할 수 있다. 일부 구현들에서, 위핑된 이전에 재구성된 프레임은 (시간 스텝 $t-1$ 로부터) 현재 시간 스텝 (t) 의 순환 계층에 대한 피드백으로서 이루어진다.

[0168] 도 12 는 부가적인 마스킹 기능성을 갖는 뉴럴 네트워크 시스템 (511) 의 예를 도시하는 다이어그램이다. 도 12 에 나타난 바와 같이, 하나 이상의 포스트-순환 계층들 (573) 은 재구성된 모션 추정 데이터 $\hat{f}_t$ 및 재구성된 잔차 데이터 $\hat{r}_t$ 에 부가하여 엘리먼트 단위 마스크 $\hat{m}_t \in [0, 1]$ 를 출력할 수 있다. 위핑 엔진 (567) 은 도 5a 와 관련하여 상술된 것과 유사하게, 이전에 재구성된 프레임 $\hat{x}_{t-1}$ 을 위핑함으로써 위핑된 이전 재구성된 프레임 $\hat{f}_t(\hat{x}_{t-1})$ 을 생성할 수 있다. 마스크 $\hat{m}_t$ 는 비디오 프레임에서의 각각의 픽셀에 대한 값을 갖는 픽셀당 마스크일 수 있다. 마스크 $\hat{m}_t$ 는 재구성된 프레임 $\hat{x}_{t-1}$ 을 재구성할 때 위핑된 이전 재구성된 프레임 $\hat{f}_t(\hat{x}_{t-1})$ 및 잔차 $\hat{r}_t$ 의 기여들을 마스킹 (또는 가중) 하는데 사용된다. 예를 들어, 일부 구현들에서, 마스크 $\hat{m}_t$ 는 다음의 식에 따라 적용될 수 있다:

[0169]
$$\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(\hat{x}_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t$$
 식 (4).

[0170] 식 (4) 에 따르면, 마스크 $\hat{m}_t$ 의 값이 증가함에 따라, 재구성된 프레임 $\hat{x}_t$ 에 대한 위핑된 이전 재구성된 프레임 $\hat{f}_t(\hat{x}_{t-1})$ 의 효과가 증가하고 재구성된 프레임 $\hat{x}_t$ 에 대한 잔차 데이터 $\hat{r}_t$ 의 효과는 감소된다.

다른 한편으로, 위핑된 이전 재구성된 프레임 $\hat{f}_t(\hat{x}_{t-1})$ 의 효과는 감소되고, 잔차 데이터 $\hat{r}_t$ 의 효과는 마스크 $\hat{m}_t$ 의 값이 감소함에 따라증가한다. 일부 구현들에서, $\hat{m}_t$ 는 0 과 1 사이의 값들의 범위 ([0, 1] 포함) 로 제한될 수 있다. 하나의 예시적인 예에서, 마스크가 픽셀에 대해 0 의 값을 갖는 경우, 그 픽셀에 대한 다음 프레임에서의 값은 잔차로부터 완전히 취해질 것이다.

[0171] 일부 예들에서, 마스크 $\hat{m}_t$ 는 이전 비디오 프레임에 존재하지 않았던 현재 비디오 프레임의 부분들을 마스킹하는 폐색(occlusion) 마스크일 수 있다. 이러한 시나리오는, 예를 들어 장면의 일부가 일시적으로 폐색될 때, 오브젝트가 장면에 진입할 때 등에 발생할 수 있다. 프레임의 그러한 이용가능하지 않은 부분들의 마스킹은 이들 이용 가능하지 않은 부분들에 대한 모션 정보 (및 그에 따른 위핑) 에 기인하여 수행될 수 있다. 예를 들어, 현재 프레임과 이전 프레임 사이의 모션 추정은 현재 프레임의 픽셀들이 이전 프레임에서 이용가능하지 않은 경우 그 픽셀들에 대해 수행될 수 없다. 이러한 예에서, 그러한 픽셀들을 포함하는 비디오 프레임의 부분은 위핑될 수 없다. 이러한 경우들에서, 현재 프레임의 재구성은 폐색의 양에 의존하여 잔차 데이

터 $\hat{r}_t$ 에 더 과중하게 (또는 일부 경우들에서 완전히) 의존함으로써 결정될 수 있다.

[0172]

도 13 은 부가적인 재구성 정제 기능성을 갖는 뉴럴 네트워크 시스템 (511) 의 예를 도시하는 다이어그램이다.

도 13 에 나타난 구성을 사용하여, 뉴럴 네트워크 시스템 (511) 은 뉴럴 네트워크 시스템 (511) 의 동작의 2 개의 시간 스텝들에서 각각의 비디오 프레임들을 재구성할 수 있다. 예를 들어, 현재 시간 스텝 (예를 들어,

시간 스텝 t-1) 에서 결정된 각각의 재구성된 비디오 프레임 (예를 들어, $\hat{x}_{t-1}$) 은 다음 시간 스텝 (예를 들어, 시간 스텝 t) 에서 추가로 정제될 수 있다. 이러한 경우들에서, 이전에 재구성된 비디오 프레임 $\hat{x}$ 은

2가지 역할을 할 수 있다. 예를 들어, 현재 시간 스텝 t 에서, 이전에 재구성된 비디오 프레임 $\hat{x}_{t-1}$ 은

위평되고, (도 5a 에 나타난 구성에서 또한 행해지는 바와 같이) 현재 시간 스텝 재구성 $\hat{x}_t$ 을 형성하기 위해

잔차 $\hat{r}_t$ 가 부가된다. 게다가, 잔차 $\hat{r}_{t-1}$ 가 또한 출력되고 이전에 재구성된 비디오 프레임 $\hat{x}_{t-1}$ 에

부가되어, 이전에 재구성된 비디오 프레임 $\hat{x}_{t-1}$ 과 비교하여 개선된 품질을 갖는 정제된 재구성된 프레임

$\hat{x}'_{t-1}$ 이 생성된다.

[0173]

도 13 에 나타난 바와 같이, 각각의 시간 스텝에서, 이전의 실측 자료 (압축되지 않은) 프레임 x 이 인코더 (533) 에 대한 입력으로서 부가되어, 이전의 재구성된 프레임 $\hat{x}$ 의 정제가 발생할 수 있도록 한다. 예를

들어, 시간 스텝 t 에서, 이전의 실측 자료 프레임 x_{t-1} 이 인코더 (533) 에 대한 입력으로서 부가되어 이전

의 재구성된 프레임 $\hat{x}_{t-1}$ 의 정제가 발생할 수 있도록 한다. 나타난 바와 같이, 시간 스텝 t 에서, 하나

이상의 포스트-순환 계층들 (573) 은 잔차 $\hat{r}_t$ 에 부가하여 출력 잔차 $\hat{r}'_{t-1}$ 를 제공한다. 잔차 $\hat{r}'_{t-1}$

는 이전 실측 자료 프레임 x_{t-1} 에 기초하여 생성되고, 정제된 재구성된 프레임 $\hat{x}'_{t-1}$ 을 생성하기 위해 이

전 재구성 $\hat{x}_{t-1}$ 을 정제하는데 사용된다. 예를 들어, 정제된 재구성은 다음과 같이 생성될 수 있다:

$$\hat{x}'_{t-1} = \hat{x}_{t-1} + \hat{r}'_{t-1} \quad \text{식 (5)}$$

[0174]

정제된 재구성된 프레임 $\hat{x}'_{t-1}$ 은 시간 스텝 t-1 에서 이전 프레임에 대한 최종 재구성으로서 사용될 수 있다.

이러한 경우들에서, 시스템 (500) 은 뉴럴 네트워크 시스템 (511) 의 동작의 하나의 시간 스텝만큼 출력 비디오 시퀀스를 지연시킨다. 뉴럴 네트워크 시스템 (511) 이 제 1 초기의 정제되지 않은 재구성에서 수행된 에러를 아는 것에 의해 다음 시간 스텝에서 재구성된 프레임들을 정제하도록 허용되는 것이 코딩 효율 (예를 들어, 레이트-왜곡) 의 관점에서 유리할 수 있다. 이러한 구성은 이전 시간 스텝에서 발생한 임의의 실수들이 다음 시간 스텝에서 보상될 수 있도록 하며, 이는 레이트 왜곡 최적화에 이로울 수 있다.

[0176]

도 14 는 뉴럴 네트워크 시스템 (511) 의 인코더 (533) 에 대한 입력으로서 하나 이상의 왜곡 맵을 사용하는 예를 도시하는 다이어그램이다. 왜곡 맵 (577) 및 왜곡 맵 (578) 을 포함하는 2개의 예시의 왜곡 맵이 도 14

에 나타나 있다. 왜곡 맵 (577) 은 현재 시간 스텝 t 로부터의 현재 소스 프레임 x_t 및 이전 시간 스텝

t-1 로부터의 이전 재구성 $\hat{x}_{t-1}$ 사이의 왜곡을 포함한다. 왜곡 맵 (578) 은 현재 소스 프레임 x_t 과

위평된 이전에 재구성된 비디오 프레임 $f_t^{enc}(\hat{x}_{t-1})$ 사이의 왜곡을 포함한다. 위에 언급된 바와 같이, 이전에 재구성된 비디오 프레임 $(\hat{x}_{t-1})$ 은 광학 흐름과 같은 모션 추정에 따라 위평될 수 있다. 이러한 경우, 광학 흐름은 플로우넷을 사용하는 것과 같이 인코더 측 상에서 추정될 수 있다. 플로우넷은 컨볼루션 네트워크들로 광학 흐름을 학습하는 뉴럴 네트워크 아키텍처이다.

[0177] 왜곡 맵들에 포함된 왜곡은 뉴럴 네트워크 시스템 (511) 을 훈련하기 위한 손실 함수 (예를 들어, 위에서 설명된 Loss0, Loss1, Loss2, Loss3, 및/또는 Loss4) 로서 사용되는 왜곡 (또는 에러 또는 손실) 일 수 있다. 왜곡은 특히 멀티-스케일 구조적 유사성 인덱스 (Multi-scale structural similarity Index; MS-SSIM), 픽셀당 평균 제곱 에러 (mean squared error; MSE) 와 같은 임의의 타입의 왜곡에 기초할 수 있다. 왜곡 맵은 또한 시각 왜곡 맵으로 지칭될 수 있다.

[0178] 왜곡 맵들 (뉴럴 네트워크 시스템 (511) 의 외부) 을 명시적으로 계산하고 왜곡 맵들을 인코더 (533) 에 입력으로서 공급함으로써, 인코더 (533) 는 이미지의 어느 부분들이 시각적 왜곡에 기여하고 있는지 그리고 따라서 어디에 더 많은 정보를 할당할 필요가 있는지에 관해 안내될 수 있다. 예를 들어, 인코더 (533) 는 왜곡에 가장 기여하는 이미지의 영역들에 비트레이트를 포커싱할 수 있다. 이러한 이점은 뉴럴 네트워크에 의해 학습하기에 사소하지 않은, 시각적 왜곡 맵 (예를 들어, MS-SSIM) 과 같은, 소정의 왜곡 맵들에 대해 강화될 수 있다.

[0179] 도 15 는 인코더 (533) 에서 승산 상호작용을 허용하는 부가적인 선형 뉴럴 네트워크 계층을 갖는 뉴럴 네트워크 시스템 (511) 의 예를 도시하는 다이어그램이다. 디폴트 방식으로 적용되는 컨볼루션 뉴럴 네트워크 아키텍처는 승산 상호작용을 허용하지 않는다. 승산 상호작용은 입력에 대해 수행되고 있는 계산이 입력 자체에 의존할 때이다. 예를 들어, 전형적인 컨볼루션 뉴럴 네트워크들은 훈련 동안 가중치들을 학습한다. 가중치들은 동일하고 뉴럴 네트워크에 입력되고 있는 입력 값들에 관계없이 임의의 입력에 적용된다.

[0180] 도 15 를 참조하면, 승산 상호작용은 인코더 (533) 의 선형 계층 (581)(선형 계층 (1) 로 도 15 에 나타냄) 을 입력의 각각의 공간 포지션에 독립적으로 적용함으로써 인코더 (533) 에서 인에이블된다. 선형 계층 (581) 의 가중치들은 선형 계층 (582)(선형 계층 (2) 로 도 15 에 나타냄) 에 의해 출력된 가중치들 (583) 에 의해 결정된다. 예를 들어, 가중치들 (583) 은 선형 계층 (581) 에 대한 가중치들로서 사용될 수 있다. 동일한 입력 (580) 이 선형 계층 (581) 및 선형 계층 (582) 에 인가된다. 선형 계층 (581) 의 가중치들을 정의하기 위해 선형 계층 (582) 을 사용하는 것은 선형 계층 (581) 에 의해 수행되는 계산이 입력 (580) 자체의 값(들)에 의존할 수 있도록 한다 (따라서, 입력 (580) 에 적용되는 계산은 입력 자체에 의존한다). 예를 들어, 뉴럴 네트워크 시스템 (511) 은 이전 재구성된 프레임과 현재 프레임 사이의 왜곡은, 왜곡이 상당한 영역들과 비교하여 무시할 수 있는 영역들에서 상이한 계산을 수행하는 것을 학습할 수도 있다.

[0181] 일부 구현들에서, 도 15 의 구성은 도 16 에 나타낸 구성과 조합으로 사용될 수 있다. 예를 들어, 하나 이상의 왜곡 맵이 입력 (580) 에 부가하여 선형 계층 (582) 에 입력될 수 있으며, 이는 가중치들 (583) 의 계산을 강화하는데 도움이 될 수 있다. 일 예에서, 상이한 공간 위치들에서의 가중치들은 왜곡 맵에 의해 표시된 바와 같이, 상이한 공간 위치들에서의 왜곡의 값에 따라 상이하게 설정될 수 있다.

[0182] 비디오 압축을 위해 본 명세서에 설명된 순환-기반뉴럴 네트워크 시스템 (예를 들어, 시스템 (400) 및/또는 시스템 (500)) 및 관련 기법들은 최신 레벨들에서 수행할 수 있는 비디오 코딩 시스템을 제공하여, 고 품질의 bpp (bits-per-pixel) 값들을 제공한다. 예를 들어, 도 16 및 도 17 에 제공된 그래프들에 나타낸 바와 같이, 이러한 기법들은 1080p 해상도 (이는 실제로 관심있는 bpp 영역 - 예를 들어, 근사치 0.09-0.12bpp 를 사용하는 Netflix 스트림들 1080p 비디오임) 에 대해 >0.14bpp 의 범위에서, HEVC 및 AVC 를 포함하는 현재 산업 표준 코덱들을 능가한다. 도 16 및 도 17 에서 "RecurrentAE" 는 본 명세서에서 설명된 순환-기반뉴럴 네트워크 시스템을 사용하는 결과들을 지칭하며, 도 16 및 도 17 에서 "RDAE" 는 다른 머신 러닝 기반 비디오 압축 기법을 지칭한다.

[0183] 본 명세서에 설명된 순환-기반뉴럴 네트워크 시스템 (예를 들어, 시스템 (400) 및/또는 시스템 (500)) 은 저-레이턴시 동작을 제공한다. 예를 들어, 시스템은 (예를 들어, 픽처들의 그룹들 (GoP들) 에서의) 여러 프레임들의 बै치(batch)들에서보다는, 프레임 단위로 비디오의 생성 및 송신을 허용하고, 따라서 송신될 프레임들의 누적을 대기할 필요가 없다. 또한, 시스템은 비디오의 특성들에 의존하여 런타임 시 GoP 사이즈 (즉, 잔차 정보이기보다는, 전체 프레임을 송신하는 빈도) 를 조정하는 능력을 제공하며, 따라서 레이트-왜곡 공간에서 우

수한 동작 포인트를 유도한다. 다양한 다른 이점들이 또한 본 명세서에 설명된 순환-기반 뉴럴 네트워크 시스템에 의해 제공된다.

- [0184] 도 18 은 본 명세서에 설명된 순환-기반 머신 러닝 기법들 중 하나 이상을 사용하여 비디오를 프로세싱하는 프로세스 (1800) 의 예를 도시하는 플로우차트이다. 블록 (1802) 에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 적어도 하나의 순환 상태 데이터를 획득하는 것을 포함한다. 일부 경우들에서, 순환 상태 데이터는 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함한다. 일부 예들에서, 도 5a 에 나타난 바와 같이, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝에 의해 생성된 이전에 재구성된 비디오 프레임과 같은, 다른 입력 데이터가 인코더 부분에 의해 획득될 수 있다.
- [0185] 블록 (1804) 에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 것을 포함한다. 출력 데이터는 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 적어도 하나의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성된다. 출력 데이터는 일부 경우들에서, 도 5a 에 나타난 바와 같이, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝에 의해 생성된 이전에 재구성된 비디오 프레임과 같은, 다른 입력 데이터에 기초하여 생성될 수 있다.
- [0186] 블록 (1806) 에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 것을 포함한다. 예를 들어, 도 5a 에 나타난 바와 같이, 동작의 현재 시간 스텝 $t+1$ 에 대해, 디코더 (538) 는 양자화기 (534) 로부터 출력 데이터를 획득할 수 있고, 디코더 (538) 의 순환 계층은 동작의 이전 시간 스텝 t 로부터 순환 상태 데이터를 그 자신 내로 피드백할 수 있다. 일부 경우들에서, 뉴럴 네트워크 시스템의 디코더 부분은 복수의 순환 계층을 포함하고, 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공한다.
- [0187] 블록 (1808) 에서, 프로세서 (1800) 는 동작의 이전 시간 스텝으로부터의 적어도 하나의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함한다. 재구성된 비디오 프레임은 입력 비디오 프레임을 나타낸다. 예를 들어, 재구성된 비디오 프레임은 입력 비디오 프레임의 재구성된 버전이다. 재구성된 비디오 프레임은 디스플레이를 위해, 저장을 위해, 및/또는 송신을 위해 출력될 수 있다. 예를 들어, 일부 예들에서, 프로세서 (1800) 는 출력 데이터를 저장 매체에 저장하는 것을 포함한다. 일부 예들에서, 프로세서 (1800) 는 재구성된 비디오 프레임을 저장 매체에 저장하는 것을 포함한다. 일부 예들에서, 프로세서 (1800) 는 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 전송하는 것을 포함한다. 일부 경우들에서, 재구성된 비디오 프레임은 디스플레이, 저장, 및/또는 송신을 위해 출력되기 전에 (예를 들어, 재구성 프로세스로 인한 시각적 아티팩트들을 제거하기 위해, 또는 다른 프로세싱을 위해) 추가로 프로세싱될 수 있다.
- [0188] 일부 예들에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 적어도 하나의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세서 (1800) 는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것, 및 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것을 포함한다.
- [0189] 일부 예들에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 적어도 하나의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세서 (1800) 는 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다.
- [0190] 일부 예들에서, 프로세서 (1800) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터

터의 적어도 하나의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하는 것을 포함할 수 있다. 일부 경우들에서, 재구성된 비디오 프레임은 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 기초하여 동작의 현재 시간 스텝에 대해 생성된다.

[0191] 일부 예들에서, 프로세스 (1800) 는 뉴럴 네트워크 시스템의 모션 추정 네트워크에 의해, 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임 또는 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다. 프로세스 (1800) 는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임을 뉴럴 네트워크 시스템의 인코더 부분에 전송하는 것을 포함할 수 있다. 일부 경우들에서, 동작의 현재 시간 스텝에 대한 생성된 출력 데이터는 동작의 상기 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 뉴럴 네트워크 시스템의 인코더 부분에 의해 생성된다.

[0192] 일부 예들에서, 프로세스 (1800) 는 하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것을 포함한다. 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정한다. 예를 들어, 제 1 손실 함수는 상술한 손실 함수들 (Loss1, Loss2, Loss3, 및/또는 Loss4) 중 임의의 것을 포함할 수 있다. 하나의 예시적인 예에서, Loss1 은 인코더 (533) 를 훈련하기 위해 사용될 수 있고, Loss3 은 디코더 (538) 를 훈련하기 위해 사용될 수 있다. 다른 예시적인 예에서, Loss2 는 인코더 (533) 를 훈련하기 위해 사용될 수 있고, Loss4 는 디코더 (538) 를 훈련하기 위해 사용될 수 있다. 이러한 예들에서, 프로세스 (1800) 는 제 1 손실 함수를 비활성화시키는 것 (예를 들어, Loss1, Loss2, Loss3, 및/또는 Loss4 를 비활성화시키는 것) 을 포함할 수 있다. 예를 들어, 제 1 손실 함수는, 도 8 및 도 9 와 관련하여 상술한 바와 같이, 뉴럴 네트워크 시스템이 모션 추정 파라미터들을 학습하도록 훈련된다고 결정하는 것에 응답하여 비활성화될 수 있다. 이러한 예들에서, 프로세스 (1800) 는 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것을 포함할 수 있다. 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정한다. 예를 들어, 제 2 손실 함수는 상술한 Loss0 를 포함할 수 있다.

[0193] 일부 예들에서, 프로세스 (1800) 는 동작의 현재 시간 스텝 또는 동작의 이전 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 적어도 하나의 순환 상태 데이터를 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 것을 포함한다. 이러한 예들에서, 동작의 현재 시간 스텝에 대한 생성된 재구성된 비디오 프레임은 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성된다.

[0194] 일부 예들에서, 프로세스 (1800) 는 디코더 부분에 의해, 재구성된 마스크를 생성하는 것을 포함한다. 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 재구성된 마스크, 동작의 현재 시간 스텝에 대해 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함한다. 재구성된 마스크 가중치들은 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중한다. 하나의 예시적인 예에서, 도 12 와 관련하여 위에서 설명된 바와 같이, 마스크는 식 (4)
$$(\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(x_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t)$$
 를 사용하여 구현될 수 있다.

[0195] 일부 예들에서, 프로세스 (1800) 는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 동작의 이전 시간 스텝 동안, 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것을 포함한다. 이러한

예들에서, 프로세스 (1800) 는 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하는 것을 포함할 수 있다. 이러한 예들에서, 프로세스 (1800) 는 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임들 부가함으로써 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다. 프로세스 (1800) 는 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하는 것을 포함할 수 있다. 이러한 예들에서, 프로세스 (1800) 는, 동작의 현재 시간 스텝 동안, 제 2 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 것, 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 것을 포함할 수 있다.

예를 들어, 도 13 과 관련하여 위에 설명된 바와 같이, 수정된 이전에 재구성된 비디오 프레임 $\hat{x}_t$ 은 시간 스텝 $t+1$ 에서 재구성된 비디오 프레임 $\hat{x}_{t+1}$ 을 출력하는 대신에 시간 스텝 $t+1$ 에서 (예를 들어, 디스플레이, 저장, 송신 및/또는 다른 동작을 위해) 출력될 수 있다. 이러한 예에서, 재구성된 비디오 프레임 $\hat{x}_{t+1}$ 은 다음 시간 스텝 $t+2$ 에서 수정을 위해 출력될 수 있다.

[0196] 일부 예들에서, 프로세스 (1800) 는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵을 획득하는 것, 및 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 출력 데이터를 생성하는 것을 포함한다. 일부 경우들에서, 도 14 와 관련하여 위에 설명된 바와 같이, 하나 이상의 왜곡 맵은 동작의 현재 시간 스텝에 대한 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함할 수 있다. 일부 경우들에서, 도 14 와 관련하여 위에 설명된 바와 같이, 하나 이상의 왜곡 맵은 동작의 현재 시간 스텝에 대한 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함할 수 있다 (여기서, 워핑된 재구성된 비디오 프레임은 상술한 바와 같이, 동작의 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 생성될 수 있다).

[0197] 일부 예들에서, 프로세스 (1800) 는 뉴럴 네트워크 시스템의 인코더 부분의 제 1 계층에 의해, 입력 데이터를 프로세싱하는 것을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 인코더 부분의 제 1 계층에 의해, 인코더 부분의 제 1 계층에 대한 복수의 가중치 값을 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1800) 는 뉴럴 네트워크 시스템의 인코더 부분의 제 2 계층의 가중치들을 복수의 가중치 값들로 설정하는 것, 및 인코더 부분의 제 2 계층에 의해, 복수의 가중치 값들로 설정된 가중치들을 사용하여 입력 데이터를 프로세싱하는 것을 포함한다. 예를 들어, 도 15 와 관련하여 위에 설명된 바와 같이, 선형 계층 (581) 의 가중치들은 선형 계층 (582) 에 의해 출력된 가중치들 (583) 에 의해 결정될 수 있으며, 이는 선형 계층 (581) 에 의해 수행되는 계산이 입력 (580) 자체의 값(들)에 의존할 수 있게 할 수 있다.

[0198] 도 19 는 본 명세서에 설명된 순환-기반 머신 러닝 기법들 중 하나 이상을 사용하여, 비디오를 프로세싱하는 프로세스 (1900) 의 다른 예를 도시하는 플로우차트이다. 블록 (1902) 에서, 프로세스 (1900) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 것을 포함한다. 예시의 목적으로 도 5a 의 예를 사용하면, 동작의 현재 시간 스텝 $t+1$ 에 대해, 디코더 (538) 는 양자화기 (534) 로부터 출력 데이터를 획득할 수 있고, 디코더 (538) 의 순환 계

층은 동작의 이전 시간 스텝 t 로부터 순환 상태 데이터 h_t 를 그 자신 내로 피드백할 수 있다. 일부 경우들에서, 뉴럴 네트워크 시스템의 디코더 부분은 복수의 순환 계층을 포함하고, 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공한다.

[0199] 블록 (1904) 에서, 프로세스 (1900) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 적어도 하나의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 것을 포함한다. 블록 (1906) 에서, 프로세스 (1900) 는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것을 포함한다. 블록 (1908) 에서, 프로세스 (1900) 는 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비

오 프레임을 생성하는 것을 포함한다.

[0200] 블록 (1910) 에서, 프로세스 (1900) 는 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함한다. 재구성된 비디오 프레임은 뉴럴 네트워크 시스템의 인코더 부분 또는 다른 뉴럴 네트워크 시스템의 인코더에 의해 프로세싱된 입력 비디오 프레임을 나타낸다. 예를 들어, 재구성된 비디오 프레임은 입력 비디오 프레임의 재구성된 버전이다. 재구성된 비디오 프레임은 디스플레이를 위해, 저장을 위해, 및/또는 송신을 위해 출력될 수 있다. 예를 들어, 일부 예들에서, 프로세스 (1800) 는 출력 데이터를 저장 매체에 저장하는 것을 포함한다. 일부 예들에서, 프로세스 (1800) 는 재구성된 비디오 프레임을 저장 매체에 저장하는 것을 포함한다. 일부 예들에서, 프로세스 (1800) 는 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 전송하는 것을 포함한다. 일부 경우들에서, 재구성된 비디오 프레임은 디스플레이, 저장, 및/또는 송신을 위해 출력되기 전에 (예를 들어, 재구성 프로세스로 인한 시각적 아티팩트들을 제거하기 위해, 또는 다른 프로세싱을 위해) 추가로 프로세싱될 수 있다.

[0201] 일부 예들에서, 프로세스 (1900) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다.

[0202] 일부 예들에서, 프로세스 (1900) 는 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 뉴럴 네트워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 동작의 현재 시간 스텝 동안, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 것을 포함할 수 있다. 프로세스 (1900) 는 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 것을 포함할 수 있다.

[0203] 일부 예들에서, 프로세스 (1900) 는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 상기 워핑된 재구성된 비디오 프레임을 프로세싱하는 것을 포함할 수 있다. 프로세스 (1900) 는 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다.

[0204] 일부 예들에서, 프로세스 (1900) 는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 또는 동작의 이전 시간 스텝에 대한 결정된 모션 추정을 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다.

[0205] 일부 예들에서, 프로세스 (1900) 는 디코더 부분에 의해, 재구성된 마스크를 생성하는 것을 포함한다. 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 동작의 현재 시간 스텝에 대해 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다. 재구성된 마스크 가중치들은 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중한다.

하나의 예시적인 예에서, 도 12 와 관련하여 위에서 설명된 바와 같이, 마스크는 식 (4)

$$(\hat{x}_t = \hat{m}_t \cdot \hat{f}_t(x_{t-1}) + (1 - \hat{m}_t) \cdot \hat{r}_t)$$

를 사용하여 구현될 수 있다.

[0206] 일부 예들에서, 프로세스 (1900) 는 뉴럴 네트워크 시스템의 인코더 부분을 사용하여, 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스

템으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 것을 포함한다. 일부 경우들에서, 순환 상태 데이터는 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함한다. 일부 예들에서, 도 5a 에 나타낸 바와 같이, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝에 의해 생성된 이전에 재구성된 비디오 프레임과 같은, 다른 입력 데이터가 인코더 부분에 의해 획득될 수 있다. 프로세스 (1900) 는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 출력 데이터를 생성하는 것을 포함할 수 있고, 출력 데이터는 상기 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 상기 이전 시간 스텝으로부터의 상기 순환 상태 데이터에 적어도 부분적으로 기초하여 생성된다. 출력 데이터는 일부 경우들에서, 도 5a 에 나타낸 바와 같이, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝에 의해 생성된 이전에 재구성된 비디오 프레임과 같은, 다른 입력 데이터에 기초하여 생성될 수 있다.

[0207]

일부 예들에서, 프로세스 (1900) 는 하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것을 포함한다. 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정한다. 예를 들어, 제 1 손실 함수는 상술한 손실 함수들 (Loss1, Loss2, Loss3, 및/또는 Loss4) 중 임의의 것을 포함할 수 있다. 하나의 예시적인 예에서, Loss1 은 인코더 (533) 를 훈련하기 위해 사용될 수 있고, Loss3 은 디코더 (538) 를 훈련하기 위해 사용될 수 있다. 다른 예시적인 예에서, Loss2 는 인코더 (533) 를 훈련하기 위해 사용될 수 있고, Loss4 는 디코더 (538) 를 훈련하기 위해 사용될 수 있다. 이러한 예들에서, 프로세스 (1900) 는 제 1 손실 함수를 비활성화시키는 것 (예를 들어, Loss1, Loss2, Loss3, 및/또는 Loss4 를 비활성화시키는 것) 을 포함할 수 있다. 예를 들어, 제 1 손실 함수는, 도 8 및 도 9 와 관련하여 상술한 바와 같이, 뉴럴 네트워크 시스템이 모션 추정 파라미터들을 학습하도록 훈련된다고 결정하는 것에 응답하여 비활성화될 수 있다. 이러한 예들에서, 프로세스 (1900) 는 하나 이상의 훈련 반복 후 수행된 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것을 포함할 수 있다. 제 2 손실 함수는 적어도 하나의 출력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정한다. 예를 들어, 제 2 손실 함수는 상술한 Loss0 를 포함할 수 있다.

[0208]

일부 예들에서, 프로세스 (1900) 는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 동작의 이전 시간 스텝 동안, 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 것을 포함한다. 이러한 예들에서, 프로세스 (1900) 는 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하는 것을 포함할 수 있다. 이러한 예들에서, 프로세스 (1900) 는 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것을 포함할 수 있다. 프로세스 (1900) 는 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하는 것을 포함할 수 있다. 이러한 예들에서, 프로세스 (1900) 는, 동작의 현재 시간 스텝 동안, 제 2 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 것, 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 것을 포함할 수 있다.

예를 들어, 도 13 과 관련하여 위에 설명된 바와 같이, 수정된 이전에 재구성된 비디오 프레임 $\hat{x}_t$ 은 시간 스텝 $t+1$ 에서 재구성된 비디오 프레임 $\hat{x}_{t+1}$ 을 출력하는 대신에 시간 스텝 $t+1$ 에서 (예를 들어, 디스플레이, 저장, 송신 및/또는 다른 동작을 위해) 출력될 수 있다. 이러한 예에서, 재구성된 비디오 프레임 $\hat{x}_{t+1}$ 은 다음 시간 스텝 $t+2$ 에서 수정을 위해 출력될 수 있다.

[0209]

일부 예들에서, 본 명세서에 설명된 프로세스들 (예를 들어, 프로세스 (1800), 프로세스 (1900), 및/또는 본 명세서에 설명된 다른 프로세스) 은 도 20 에 나타낸 컴퓨팅 디바이스 아키텍처 (2000) 를 갖는 컴퓨팅 디바이스와 같은, 컴퓨팅 디바이스 또는 장치에 의해 수행될 수도 있다. 일부 예들에서, 컴퓨팅 디바이스는 모바일 디바이스, 웨어러블 디바이스, 확장 현실 디바이스 (예를 들어, 가상 현실 (VR) 디바이스, 증강 현실 (AR) 디바이스, 또는 혼합 현실 (MR) 디바이스), 개인용 컴퓨터, 랩탑 컴퓨터, 비디오 서버, 텔레비전, 또는 다른 디바이스를 포함할 수 있다. 일 예에서, 프로세스 (1800) 및/또는 프로세스 (1900) 는 (도 5a 내지 도 15 에 나타

낸 아키텍처들 중 임의의 것 또는 이들의 임의의 조합을 사용하여) 시스템 (400) 또는 시스템 (500) 을 구현하는 컴퓨팅 디바이스 아키텍처 (2000) 를 갖는 컴퓨팅 디바이스에 의해 수행될 수 있다. 컴퓨팅 디바이스는 임의의 적절한 디바이스, 예컨대 자율 차량, 로봇 디바이스, 모바일 디바이스 (예를 들어, 모바일 폰), 데스크톱 컴퓨팅 디바이스, 태블릿 컴퓨팅 디바이스, 웨어러블 디바이스, 및/또는 프로세서 (1800) 를 포함하여, 본 명세서에 설명된 프로세스들을 수행하기 위한 리소스 능력들을 갖는 임의의 다른 컴퓨팅 디바이스를 포함할 수 있다. 일부 경우들에서, 컴퓨팅 디바이스 또는 장치는 다양한 컴포넌트들, 예컨대 하나 이상의 입력 디바이스, 하나 이상의 출력 디바이스, 하나 이상의 프로세서, 하나 이상의 마이크로프로세서, 하나 이상의 마이크로컴퓨터들, 하나 이상의 송신기, 수신기 또는 조합된 송신기-수신기 (예를 들어, 트랜시버로 지칭됨), 하나 이상의 카메라, 하나 이상의 센서, 및/또는 본 명세서에 설명된 프로세스들의 단계들을 수행하도록 구성되는 다른 컴포넌트(들) 을 포함할 수도 있다. 일부 예들에서, 컴퓨팅 디바이스는 디스플레이, 데이터를 통신 및/또는 수신하도록 구성된 네트워크 인터페이스, 이들의 임의의 조합, 및/또는 다른 컴포넌트(들)을 포함할 수도 있다. 네트워크 인터페이스는 인터넷 프로토콜 (IP) 기반 데이터 또는 다른 타입의 데이터를 통신 및/또는 수신하도록 구성될 수도 있다.

[0210] 컴퓨팅 디바이스의 컴포넌트들은 회로부에서 구현될 수 있다. 예를 들어, 컴포넌트들은 하나 이상의 프로그램가능한 전자 회로 (예를 들어, 마이크로프로세서, 그래픽스 프로세싱 유닛 (GPU), 디지털 신호 프로세서 (DSP), 중앙 프로세싱 유닛 (CPU), 뉴럴 프로세싱 유닛 (NPU), 및/또는 다른 적절한 전자 회로) 를 포함할 수 있고 및/또는 이를 사용하여 구현될 수 있으며, 및/또는 본 명세서에 설명된 다양한 동작들을 수행하기 위해 컴퓨터 소프트웨어, 펌웨어, 또는 이들의 임의의 조합을 포함하고 및/또는 이를 사용하여 구현될 수 있다.

[0211] 프로세서 (1800) 및 프로세서 (1900) 는 논리 플로우 다이어그램으로서 예시되고, 그 동작은 하드웨어, 컴퓨터 명령들, 또는 이들의 조합으로 구현될 수 있는 동작들의 시퀀스를 나타낸다. 컴퓨터 명령들의 컨텍스트에서, 그 동작들은, 하나 이상의 프로세서들에 의해 실행될 때, 열거된 동작들을 수행하는 하나 이상의 컴퓨터 판독가능 저장 매체들 상에 저장된 컴퓨터 실행가능 명령들을 표현한다. 일반적으로, 컴퓨터 실행가능 명령들은 특정의 기능들을 수행하거나 또는 특정의 데이터 타입들을 구현하는 루틴들, 프로그램들, 오브젝트들, 컴포넌트들, 데이터 구조들 등을 포함한다. 동작들이 설명되는 순서는 제한으로서 해석되도록 의도되지 않으며, 임의의 수의 설명된 동작들은 프로세스들을 구현하기 위해 임의의 순서로 및/또는 병렬로 조합될 수 있다.

[0212] 부가적으로, 본 명세서에 설명된 프로세스들 (프로세서 (1800), 프로세서 (1900), 및/또는 본 명세서에 설명된 다른 프로세스) 은 실행가능 명령들로 구성된 하나 이상의 컴퓨터 시스템들의 제어 하에서 수행될 수도 있고, 집합적으로 하나 이상의 프로세서 상에서 실행하는 코드 (예를 들어, 실행가능 명령들, 하나 이상의 컴퓨터 프로그램들, 또는 하나 이상의 애플리케이션들) 로서, 하드웨어에 의해, 또는 이들의 조합으로 구현될 수도 있다. 위에 언급된 바와 같이, 코드는 컴퓨터 판독가능 또는 머신 판독가능 저장 매체 상에, 예를 들어, 하나 이상의 프로세서들에 의해 실행가능한 복수의 명령들을 포함하는 컴퓨터 프로그램의 형태로 저장될 수도 있다. 컴퓨터 판독가능 또는 머신 판독가능 저장 매체는 비일시적일 수도 있다.

[0213] 도 20 은 본 명세서에 설명된 다양한 기법들을 구현할 수 있는 예시의 컴퓨팅 디바이스의 예시의 컴퓨팅 디바이스 아키텍처 (2000) 를 도시한다. 일부 예들에서, 컴퓨팅 디바이스는 모바일 디바이스, 웨어러블 디바이스, 확장 현실 디바이스 (예를 들어, 가상 현실 (VR) 디바이스, 증강 현실 (AR) 디바이스, 또는 혼합 현실 (MR) 디바이스), 개인용 컴퓨터, 랩톱 컴퓨터, 비디오 서버, 또는 다른 디바이스를 포함할 수 있다. 예를 들어, 컴퓨팅 디바이스 아키텍처 (2000) 는 시스템 (400) 및/또는 시스템 (500)(도 5a 내지 도 15 에 나타난 아키텍처 중 임의의 하나 또는 이들의 임의의 조합을 포함함) 을 구현할 수 있다. 컴퓨팅 디바이스 아키텍처 (2000) 의 컴포넌트들은 버스와 같은 커넥션 (2005) 을 사용하여 서로 전기적으로 통신하는 것으로 나타낸다. 예시의 컴퓨팅 디바이스 아키텍처 (2000) 는 프로세싱 유닛 (CPU 또는 프로세서)(2010), 및 판독 전용 메모리 (ROM)(2020) 및 랜덤 액세스 메모리 (RAM)(2025)와 같은 컴퓨팅 디바이스 메모리 (2015) 를 포함하는 다양한 컴퓨팅 디바이스 컴포넌트들을 프로세서 (2010) 에 커플링하는 컴퓨팅 디바이스 커넥션 (2005) 을 포함한다.

[0214] 컴퓨팅 디바이스 아키텍처 (2000) 는 프로세서 (2010) 와 직접 연결되거나, 이에 근접하거나, 또는 이의 일부로서 통합된 고속 메모리의 캐시를 포함할 수 있다. 컴퓨팅 디바이스 아키텍처 (2000) 는 프로세서 (2010) 에 의한 빠른 액세스를 위해 메모리 (2015) 및/또는 저장 디바이스 (2030) 로부터 캐시 (2012) 로 데이터를 복사할 수 있다. 이러한 방식으로, 캐시는 데이터를 대기하면서 프로세서 (2010) 지연들을 회피하는 성능 부스트를 제공할 수 있다. 이들 및 다른 모듈들은 다양한 액션들을 수행하기 위해 프로세서 (2010) 를 제어하거나 이

를 제어하도록 구성될 수 있다. 다른 컴퓨팅 디바이스 메모리 (2015) 가 또한 이용가능할 수도 있다. 메모리 (2015) 는 상이한 성능 특성들을 갖는 다수의 상이한 타입들의 메모리를 포함할 수 있다. 프로세서 (2010) 는 소프트웨어 명령들이 프로세서 설계에 통합되는 특수 목적 프로세서 뿐만 아니라 프로세서 (2010) 를 제어하도록 구성된, 저장 디바이스 (2030) 에 저장된, 서비스 1 (2032), 서비스 2 (2034) 및 서비스 3 (2036) 과 같은, 하드웨어 또는 소프트웨어 서비스 및 임의의 범용 프로세서를 포함할 수 있다. 프로세서 (2010) 는 다중 코어들 또는 프로세서들, 버스, 메모리 제어기, 캐시 등을 포함하는 독립형 시스템일 수도 있다. 멀티-코어 프로세서는 대칭 또는 비대칭일 수도 있다.

[0215] 컴퓨팅 디바이스 아키텍처 (2000) 와의 사용자 상호작용을 가능하게 하기 위해, 입력 디바이스 (2045) 는 스피치를 위한 마이크론, 제스처 또는 그래픽 입력을 위한 터치-감지 스크린, 키보드, 마우스, 모션 입력, 스피치 등과 같은 임의의 수의 입력 메커니즘을 나타낼 수 있다. 출력 디바이스 (2035) 는 또한 디스플레이, 프로젝터, 텔레비전, 스피커 디바이스 등과 같은, 당업자에게 알려진 다수의 출력 메커니즘 중 하나 이상일 수 있다. 일부 경우들에서, 멀티모달 컴퓨팅 디바이스들은 사용자가 컴퓨팅 디바이스 아키텍처 (2000) 와 통신하기 위해 다중 타입들의 입력을 제공하는 것을 가능하게 할 수 있다. 통신 인터페이스 (2040) 는 일반적으로 사용자 입력 및 컴퓨팅 디바이스 출력을 통제하고 관리할 수 있다. 임의의 특정 하드웨어 배열 상에서 동작하는 것에 대한 제한은 없으며 따라서 여기서 기본 특징들은 이들이 개발됨에 따라 개선된 하드웨어 또는 펌웨어 배열들에 용이하게 대체될 수도 있다.

[0216] 저장 디바이스 (2030) 는 비휘발성 메모리이고, 자기 카세트, 플래시 메모리 카드, 고체 상태 메모리 디바이스, 디지털 다기능 디스크, 카트리지, 랜덤 액세스 메모리 (RAM)(2025), 판독 전용 메모리 (ROM)(2020), 및 이들의 하이브리드와 같은, 컴퓨터에 의해 액세스가능한 데이터를 저장할 수 있는 하드 디스크 또는 다른 타입들의 컴퓨터 판독가능 매체들일 수 있다. 저장 디바이스 (2030) 는 프로세서 (2010) 를 제어하기 위한 서비스들 (2032, 2034, 2036) 을 포함할 수 있다. 다른 하드웨어 또는 소프트웨어 모듈들이 고려된다. 저장 디바이스 (2030) 는 컴퓨팅 디바이스 커넥션 (2005) 에 연결될 수 있다. 일 양태에서, 특정 기능을 수행하는 하드웨어 모듈은 기능을 수행하기 위해 프로세서 (2010), 커넥션 (2005), 출력 디바이스 (2035) 등과 같은, 필요한 하드웨어 컴포넌트들과 관련하여 컴퓨터 판독가능 매체에 저장된 소프트웨어 컴포넌트를 포함할 수 있다.

[0217] 본 개시의 양태들은 하나 이상의 액티브 깊이 감지 시스템을 포함하거나 이에 커플링되는 임의의 적합한 전자 디바이스 (예컨대, 보안 시스템, 스마트폰, 태블릿, 랩탑 컴퓨터, 차량, 드론, 또는 다른 디바이스) 에 적용가능하다. 하기에서는 하나의 광 프로젝터를 갖거나 이에 커플링된 디바이스와 관련하여 설명되지만, 본 개시의 양태들은 임의의 수의 광 프로젝터를 갖는 디바이스들에 적용가능하며, 따라서 특정 디바이스들에 제한되지 않는다.

[0218] 용어 "디바이스"는 하나 또는 특정 수의 물리적 오브젝트 (예컨대, 하나의 스마트폰, 하나의 제어기, 하나의 프로세싱 시스템 등) 로 제한되지 않는다. 본 명세서에서 사용된 바와 같이, 디바이스는 본 개시의 적어도 일부 부분들을 구현할 수도 있는 하나 이상의 부분을 갖는 임의의 전자 디바이스일 수도 있다. 하기 설명 및 예들은 본 개시의 다양한 양태들을 설명하기 위해 용어 "디바이스"를 사용하지만, 용어 "디바이스"는 특정 구성, 타입 또는 오브젝트들의 수로 제한되지 않는다. 부가적으로, 용어 "시스템"은 다중 컴포넌트들 또는 특정 실시형태들에 제한되지 않는다. 예를 들어, 시스템은 하나 이상의 인쇄 회로 보드들 또는 다른 기관들 상에 구현될 수도 있고, 이동가능한 또는 정적 컴포넌트들을 가질 수도 있다. 하기 설명 및 예들은 본 개시의 다양한 양태들을 설명하기 위해 용어 "시스템"을 사용하지만, 용어 "시스템"은 특정 구성, 타입 또는 오브젝트들의 수로 제한되지 않는다.

[0219] 특정 상세들은 본 명세서에 제공된 실시형태들 및 예들의 철저한 이해를 제공하기 위해 위의 설명에서 제공된다. 그러나, 실시형태들은 이들 특정 상세들 없이도 실시될 수도 있음이 당업자에 의해 이해될 것이다. 설명의 명확성을 위해, 일부 인스턴스들에서, 본 기술은 소프트웨어, 또는 하드웨어와 소프트웨어의 조합들로 구현되는 방법에서 단계들 또는 루틴들, 디바이스 컴포넌트들, 디바이스들을 포함하는 기능 블록들을 포함한 개별 기능 블록들을 포함하는 것으로서 제시될 수도 있다. 도면들에 나타내고 및/또는 본 명세서에 설명된 것들 이외의 부가적인 컴포넌트들이 사용될 수도 있다. 예를 들어, 회로, 시스템, 네트워크, 프로세스 및 다른 컴포넌트들은, 실시형태들을 불필요한 상세들에서 불분명하게 하지 않기 위하여 블록 다이어그램 형태의 컴포넌트들로서 보여질 수도 있다. 다른 예들에서, 잘 알려진 회로들, 프로세스들, 알고리즘, 구조들, 및 기법들은 실시형태들을 모호하게 하는 것을 회피하기 위하여 불필요한 상세없이 나타낼 수도 있다.

- [0220] 개별 실시형태들은 플로우차트, 플로우 다이어그램, 데이터 플로우 다이어그램, 구조 다이어그램, 또는 블록 다이어그램으로서 도시되는 프로세스 또는 방법으로서 위에서 설명될 수도 있다. 플로우차트는 순차적 프로세스로서 동작들을 설명할 수도 있지만, 많은 동작들은 병렬로, 또는 동시에 수행될 수 있다. 또한, 동작들의 순서는 재배열될 수도 있다. 프로세스는 그의 동작들이 완료되면 종료되지만, 도면에 포함되지 않은 추가의 단계들을 가질 수 있다. 프로세스는 방법, 함수, 절차, 서브루틴, 서브프로그램 등에 대응할 수도 있다. 프로세스가 함수에 대응할 경우, 그 종료는 그 함수의 호출 함수 또는 메인 함수로의 복귀에 대응할 수 있다.
- [0221] 상술한 예들에 따른 프로세스들 및 방법들은 컴퓨터 판독가능 매체들로부터 저장되거나 그렇지 않으면 이용가능한 컴퓨터 실행가능 명령들을 사용하여 구현될 수 있다. 이러한 명령들은 예를 들어, 범용 컴퓨터, 특수 목적 컴퓨터, 또는 프로세싱 디바이스가 소정의 기능 또는 기능들의 그룹을 수행하게 하거나 그렇지 않으면 수행하도록 이들을 구성하는 명령들 및 데이터를 포함할 수 있다. 사용된 컴퓨터 리소스들의 부분들은 네트워크를 통해 액세스가능할 수 있다. 컴퓨터 실행가능 명령들은, 예를 들어 바이너리들, 중간 포맷 명령들, 예컨대 어셈블리 언어, 펌웨어, 소스 코드 등일 수도 있다.
- [0222] 용어 "컴퓨터 판독가능 매체"는 휴대용 또는 고정식 저장 디바이스들, 광학 저장 디바이스들, 및 명령(들) 및/또는 데이터를 저장, 포함, 또는 반송할 수 있는 다양한 다른 매체들을 포함하지만 이에 한정되지 않는다. 컴퓨터 판독 가능 매체는 데이터가 저장될 수 있고 캐리어 파 (carrier wave) 및/또는 무선 또는 유선 연결을 통해 전파되는 일시적인 전자 신호를 포함하지 않는 비일시적인 매체를 포함할 수도 있다. 비일시적 매체의 예들은, 특히 자기 디스크 또는 테이프, 광학 저장 매체들, 예컨대 플래시 메모리, 메모리 또는 메모리 디바이스들, 자기 또는 광학 디스크들, 플래시 메모리, 비휘발성 메모리가 제공된 USB 디바이스들, 네트워크된 저장 디바이스들, 콤팩트 디스크 (CD) 또는 디지털 다기능 디스크 (DVD), 또는 이들의 임의의 적절한 조합을 포함할 수도 있지만, 이에 제한되지 않는다. 컴퓨터 판독가능 매체는 절차, 함수, 서브프로그램, 프로그램, 루틴, 서브루틴, 모듈, 소프트웨어 패키지, 클래스, 또는 명령들, 데이터 구조들, 또는 프로그램 스테이트먼트들의 임의의 조합을 나타낼 수도 있는 코드 및/또는 머신 실행가능 명령들을 저장할 수도 있다. 코드 세그먼트는, 정보, 데이터, 인수들, 파라미터들, 또는 메모리 콘텐츠를 전달 및/또는 수신함으로써 다른 코드 세그먼트 또는 하드웨어 회로에 커플링될 수도 있다. 정보, 인수들, 파라미터들, 데이터 등은 메모리 공유, 메시지 전달, 토큰 전달, 네트워크 송신 등을 포함한 임의의 적합한 수단을 통해 전달, 포워딩, 또는 송신될 수도 있다.
- [0223] 일부 실시형태들에서, 컴퓨터 판독가능 저장 디바이스들, 매체들, 및 메모리들은 비트 스트림 등을 포함하는 케이블 또는 무선 신호를 포함할 수 있다. 그러나, 언급될 때, 비일시적 컴퓨터 판독가능 저장 매체는 에너지, 캐리어 신호들, 전자기 파들, 및 신호 그 자체와 같은 매체들을 명시적으로 배제한다.
- [0224] 이들 개시들에 따른 프로세스들 및 방법들을 구현하는 디바이스들은 하드웨어, 소프트웨어, 펌웨어, 미들웨어, 마이크로코드, 하드웨어 디스크립션 언어들, 또는 이들의 임의의 조합을 포함할 수 있고, 다양한 폼 팩터들 중 임의의 것을 취할 수 있다. 소프트웨어, 펌웨어, 미들웨어 또는 마이크로코드로 구현될 때, 필요한 작업들을 수행하기 위한 프로그램 코드 또는 코드 세그먼트들 (예를 들어, 컴퓨터 프로그램 제품)은 컴퓨터 판독가능 또는 머신 판독가능 매체에 저장될 수도 있다. 프로세서(들)은 필요한 작업들을 수행할 수도 있다. 폼 팩터들의 통상의 예들은 랩탑들, 스마트폰들, 모바일 폰들, 태블릿 디바이스들 또는 다른 소형 폼 팩터 개인용 컴퓨터들, 개인용 디지털 보조기들, 랩마운트 디바이스들, 독립형 디바이스들 등을 포함한다. 본 명세서에 설명된 기능성은 또한 주변기기 또는 애드-인 카드 (add-in card) 로 구현될 수 있다. 이러한 기능성은 또한, 추가적인 예로서, 단일 디바이스에서 실행되는 상이한 칩들 또는 상이한 프로세스들 사이의 회로 보드 상에서 구현될 수 있다.
- [0225] 명령들, 이러한 명령들을 전달하기 위한 매체들, 이들을 실행하기 위한 컴퓨팅 리소스들, 및 이러한 컴퓨팅 리소스들을 지원하기 위한 다른 구조들은 본 개시에서 설명된 기능들을 제공하기 위한 예시적인 수단이다.
- [0226] 전술한 설명에서, 본원의 양태들은 이들의 구체적 실시형태를 참조하여 설명되었지만, 당업자는 본 출원이 이에 제한되지 않는다는 것을 인식할 것이다. 따라서, 본원의 예시적인 실시형태들이 본 명세서에 상세하게 설명되었지만, 본 발명의 개념은 다르게는 다양하게 구체화되고 채용될 수도 있으며, 첨부된 청구 범위는 선행 기술에 의해 제한되는 것을 제외하고는 그러한 변형을 포함하는 것으로 해석되도록 의도된다. 상술한 출원의 다양한 특징들 및 양태들은 개별적으로 또는 공동으로 사용될 수도 있다. 또한, 실시형태들은 본 명세서의 더 넓은 사상 및 범위를 벗어나지 않으면서 본원에 기재된 것 이외의 임의의 수의 환경 및 애플리케이션들에서 이용될 수 있다. 따라서, 명세서 및 도면들은 제한적이 아닌 예시적인 것으로 간주되어야 한다. 예시의 목적을 위해, 방법들이 특정 순서로 설명되었다. 대안의 실시형태에서, 그 방법들은 설명된 것과 상이한 순

서로 수행될 수도 있다는 것을 이해해야 한다.

- [0227] 당업자는 본 명세서에서 사용된 미만 (" $<$ ") 및 초과 (" $>$ ") 기호들 또는 용어가 본 개시의 범위를 벗어나지 않으면서 이하 (" $\leq$ ") 및 이상 (" $\geq$ ") 기호들로 각각 대체될 수 있음을 알 것이다.
- [0228] 컴포넌트들이 특정 동작을 수행하도록 "구성된" 것으로 기술되는 경우, 그러한 구성은 예를 들어, 전자 회로 또는 다른 하드웨어를 설계하여 그 동작을 수행하는 것에 의해, 프로그램 가능한 전자 회로 (예를 들어, 마이크로 프로세서 또는 다른 적절한 전자 회로) 를 프로그램하여 그 동작을 수행하는 것에 의해 또는 이들의 임의의 조합에 의해, 달성될 수 있다.
- [0229] 구절 "커플링된" 은 직접 또는 간접적으로 다른 컴포넌트에 물리적으로 연결되는 임의의 컴포넌트, 및/또는 직접 또는 간접적으로 다른 컴포넌트와 통신하는 (예를 들어, 유선 또는 무선 연결, 및/또는 다른 적절한 통신 인터페이스를 통해 다른 컴포넌트에 연결되는) 임의의 컴포넌트를 지칭한다.
- [0230] 세트 중 "적어도 하나" 및/또는 세트 중 "하나 이상"을 인용하는 청구항 언어 또는 다른 언어는 세트 중 하나의 멤버 또는 세트 중 다중 멤버들 (임의의 조합으로) 이 청구항을 만족한다는 것을 나타낸다. 예를 들어, 청구항 언어가 "A 및 B 중 적어도 하나" 또는 "A 또는 B 중 적어도 하나"를 인용하는 것은 A, B, 또는 A 및 B 를 의미한다. 다른 예에서, 청구항 언어가 "A, B, 및 C 중 적어도 하나" 또는 "A, B, 또는 C 중 적어도 하나"를 인용하는 것은 A, B, C, 또는 A 및 B, 또는 A 및 C, 또는 B 및 C, 또는 A 및 B 및 C를 의미한다. 세트 중 "적어도 하나" 및/또는 세트 중 "하나 이상" 의 언어는 세트에 리스트된 항목들에 대한 세트를 제한하지 않는다. 예를 들어, 청구항 언어가 "A 및 B 중 적어도 하나" 또는 "A 또는 B 중 적어도 하나"를 인용하는 것은 A, B, 또는 A 및 B 를 의미할 수 있고, A 및 B 의 세트에 리스트되지 않은 아이템들을 부가적으로 포함할 수 있다.
- [0231] 본 명세서에 개시된 실시형태들과 관련하여 설명된 다양한 예시적인 논리 블록, 모듈, 회로, 및 알고리즘 단계들이 전자 하드웨어, 컴퓨터 소프트웨어, 펌웨어, 또는 이들의 조합으로 구현될 수도 있다. 하드웨어 및 소프트웨어의 이러한 교환가능성을 명백하게 예증하기 위하여, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들, 및 단계들이 일반적으로 그들의 기능의 관점에서 설명되었다. 그러한 기능이 하드웨어 또는 소프트웨어로 구현될지 여부는, 전체 시스템에 부과된 특정 애플리케이션 및 설계 제약에 달려 있다. 당업자는 설명된 기능성을 각각의 특정 애플리케이션에 대해 다른 방식으로 구현할 수도 있지만, 이러한 구현 결정이 본 출원의 범위를 벗어나게 하는 것으로 해석되지 않아야 한다.
- [0232] 본 명세서에 기재된 기법들은 또한 전자 하드웨어, 컴퓨터 소프트웨어, 펌웨어 또는 이들의 임의의 조합으로 구현될 수도 있다. 그러한 기법들은 범용 컴퓨터, 무선 통신 디바이스 핸드세트, 또는 무선 통신 디바이스 핸드세트 및 다른 디바이스들에서의 애플리케이션을 포함하는 다수의 사용들을 갖는 집적 회로 디바이스들과 같은 다양한 디바이스들 중의 임의의 것에서 구현될 수도 있다. 모듈들 또는 컴포넌트들로서 설명된 임의의 특징들은 집적 로직 디바이스 (integrated logic device) 에서 함께 구현되거나 또는 이산이지만 상호동작가능한 로직 디바이스들로서 별도로 구현될 수도 있다. 소프트웨어로 구현되면, 기법들은, 실행될 때, 위에서 설명된 방법들 중 하나 이상을 수행하는 명령들을 포함하는 프로그램 코드를 포함하는 컴퓨터 판독가능 데이터 저장 매체에 의해 적어도 부분적으로 실현될 수도 있다. 컴퓨터 판독가능 데이터 저장 매체는 패키징 재료들을 포함할 수도 있는 컴퓨터 프로그램 제품의 일부를 형성할 수도 있다. 컴퓨터 판독가능 매체는 메모리 또는 데이터 저장 매체, 예컨대 동기식 랜덤 액세스 메모리 (SDRAM) 와 같은 랜덤 액세스 메모리 (DRAM), 판독 전용 메모리 (ROM), 비휘발성 랜덤 액세스 메모리 (NVRAM), 전기적 소거가능 프로그램가능 판독 전용 메모리 (EEPROM), 플래시 메모리, 자기 또는 광학 데이터 저장 매체 등을 포함할 수도 있다. 부가적으로 또는 대안으로 기법들은, 전파된 신호들 또는 파들과 같은, 컴퓨터에 의해 실행되고 및/또는, 액세스될 수 있는 그리고 명령들 또는 데이터 구조들의 형태로 프로그램 코드를 반송 또는 통신하는 컴퓨터 판독가능 통신 매체에 의해 적어도 부분적으로 실현될 수도 있다.
- [0233] 프로그램 코드는, 하나 이상의 디지털 신호 프로세서들 (DSP들), 범용 마이크로 프로세서들, 주문형 집적 회로들 (ASIC들), 필드 프로그램가능 로직 어레이들 (FPGA들), 또는 다른 균등한 집적된 또는 별개의 로직 회로부와 같은 하나 이상의 프로세서들을 포함할 수도 있는 프로세서에 의해 실행될 수도 있다. 그러한 프로세서는 본 개시에 기재된 기법들 중의 어느 것을 수행하도록 구성될 수도 있다. 범용 프로세서는 마이크로프로세서일 수도 있지만, 대안으로, 프로세서는 임의의 종래의 프로세서, 제어기, 마이크로제어기, 또는 상태 머신일 수도 있다. 또한, 프로세서는 컴퓨팅 디바이스들의 조합, 예를 들어 DSP 와 마이크로프로세서의 조합, 복수의

마이크로프로세서, DSP 코어와 조합한 하나 이상의 마이크로프로세서, 또는 임의의 다른 이러한 구성으로서 구현될 수도 있다. 따라서, 본원에 사용된 용어 "프로세서"는 임의의 전술한 구조, 전술한 구조의 임의의 조합, 또는 본원에 설명된 기술들의 구현에 적합한 임의의 다른 구조 또는 장치를 지칭할 수도 있다.

[0234] 본 개시의 예시적인 예들은 다음을 포함한다:

[0235] 양태 1: 비디오 데이터를 프로세싱하는 방법으로서, 방법은, 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 단계로서, 출력 데이터는 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 단계; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계로서, 재구성된 비디오 프레임은 상기 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하는 단계를 포함한다.

[0236] 양태 2: 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계; 및 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태 1의 방법.

[0237] 양태 3: 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 및 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태 2의 방법.

[0238] 양태 4: 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하는 단계; 및 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태들 2 및 3 중 임의의 것의 방법.

[0239] 양태 5: 뉴럴 네트워크 시스템의 모션 추정 네트워크에 의해, 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임 또는 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 및 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임을 뉴럴 네트워크 시스템의 인코더 부분에 전송하는 단계를 더 포함하고, 동작의 현재 시간 스텝에 대한 결정된 출력 데이터는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 뉴럴 네트워크 시스템의 인코더 부분에 의해 생성되는, 양태들 1 내지 4 중 임의의 것의 방법.

[0240] 양태 6: 하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 단계로서, 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 단계; 제 1 손실 함

수를 비활성화하는 단계; 및 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 단계로서, 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 2 손실 함수를 사용하여 상기 뉴럴 네트워크 시스템을 훈련하는 단계를 더 포함하는, 양태 5의 방법.

[0241] 양태 7: 순환 상태 데이터는 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함하는, 양태들 1 내지 6 중 임의의 것의 방법.

[0242] 양태 8: 뉴럴 네트워크 시스템의 디코더 부분은 복수의 순환 계층을 포함하고, 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공하는, 양태들 1 내지 7 중 임의의 것의 방법.

[0243] 양태 9: 동작의 이전 시간 스텝 또는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 단계를 더 포함하고, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임은 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 양태들 1 내지 8 중 임의의 것의 방법.

[0244] 양태 10: 디코더 부분에 의해, 재구성된 마스크를 생성하는 단계로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 단계; 및 재구성된 마스크, 동작의 현재 시간 스텝에 대한 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태들 1 내지 9 중 임의의 것의 방법.

[0245] 양태 11: 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계; 동작의 현재 시간 스텝 동안, 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 제 1 재구성된 잔차 데이터를 결정하는 단계; 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계; 동작의 상기 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하는 단계; 동작의 현재 시간 스텝 동안, 제 2 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 단계; 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 단계를 더 포함하는, 양태들 1 내지 10 중 임의의 것의 방법.

[0246] 양태 12: 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵을 획득하는 단계; 및 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 출력 데이터를 생성하는 단계를 더 포함하는, 양태들 1 내지 11 중 임의의 것의 방법.

[0247] 양태 13: 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 동작의 상기 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함하는, 양태 12의 방법.

[0248] 양태 14: 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함하고, 상기 워핑된 재구성된 비디오 프레임은 동작의 상기 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 생성되는, 양태들 12 및 13 중 임의의 것의 방법.

[0249] 양태 15: 뉴럴 네트워크 시스템의 인코더 부분의 제 1 계층에 의해, 입력 데이터를 프로세싱하는 단계; 인코더 부분의 제 1 계층에 의해, 인코더 부분의 제 1 계층에 대한 복수의 가중치 값을 결정하는 단계; 뉴럴 네트워크 시스템의 인코더 부분의 제 2 계층의 가중치들을 복수의 가중치 값으로 설정하는 단계; 및 인코더 부분의 제 2 계층에 의해, 복수의 가중치 값으로 설정된 가중치들을 사용하여 입력 데이터를 프로세싱하는 단계를 더 포함하는, 양태들 1 내지 14 중 임의의 것의 방법.

- [0250] 양태 16: 출력 데이터를 저장 매체에 저장하는 단계를 더 포함하는, 양태들 1 내지 15 중 임의의 것의 방법.
- [0251] 양태 17: 재구성된 비디오 프레임을 저장 매체에 저장하는 단계를 더 포함하는, 양태들 1 내지 16 중 임의의 것의 방법.
- [0252] 양태 18: 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 전송하는 단계를 더 포함하는, 양태들 1 내지 17 중 임의의 것의 방법.
- [0253] 양태 19: 비디오 데이터를 프로세싱하기 위한 장치. 장치는 메모리 (예를 들어, 회로부로 구현됨) 및 메모리에 커플링된 프로세서 (또는 다중 프로세서들) 를 포함한다. 프로세서 (또는 프로세서들) 는, 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 뉴럴 네트워크 시스템의 인코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 출력 데이터는 입력 비디오 프레임, 동작의 상기 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하고; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하며; 그리고 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 비디오 프레임은 상기 입력 비디오 프레임을 나타내는, 상기 재구성된 비디오 프레임을 생성하도록 구성된다.
- [0254] 양태 20: 프로세서는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하며; 그리고 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하도록 구성되는, 양태 19 의 장치.
- [0255] 양태 21: 프로세서는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 양태 20 의 장치.
- [0256] 양태 22: 프로세서는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하며; 그리고 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 양태들 20 및 21 중 임의의 것의 장치.
- [0257] 양태 23: 프로세서는 뉴럴 네트워크 시스템의 모션 추정 네트워크에 의해, 입력 비디오 프레임과 동작의 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 또는 동작의 이전 시간 스텝으로부터의 이전 입력 비디오 프레임 중 적어도 하나에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임 또는 이전 입력 비디오 프레임 중 적어도 하나의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하며; 그리고 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임을 뉴럴 네트워크 시스템의 인코더 부분에 전송하도록 구성되고, 동작의 현재 시간 스텝에 대한 결정된 출력 데이터는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 뉴럴 네트워크 시스템의 인코더 부분에 의해 생성되는, 양태들 19 내지 22 중 임의의 것의 장치.

- [0258] 양태 24: 프로세서는 하나 이상의 훈련 반복에 대해, 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것으로서, 제 1 손실 함수는 하나 이상의 입력 비디오 프레임과 하나 이상의 워핑된 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 1 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하고; 하나 이상의 훈련 반복 후에 수행되는 하나 이상의 후속 훈련 반복에 대해, 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하는 것으로서, 제 2 손실 함수는 적어도 하나의 입력 비디오 프레임과 적어도 하나의 재구성된 비디오 프레임 사이의 손실을 결정하는, 상기 제 2 손실 함수를 사용하여 뉴럴 네트워크 시스템을 훈련하도록 구성되는, 양태 23 의 장치.
- [0259] 양태 25: 순환 상태 데이터는 적어도 하나의 순환 계층의 노드들의 하나 이상의 값을 포함하는, 양태들 19 내지 24 중 임의의 것의 장치.
- [0260] 양태 26: 뉴럴 네트워크 시스템의 디코더 부분은 복수의 순환 계층을 포함하고, 복수의 순환 계층의 각각의 순환 계층은 개개의 순환 상태 데이터를 제공하는, 양태들 19 내지 25 중 임의의 것의 장치.
- [0261] 양태 27: 프로세서는 동작의 이전 시간 스텝 또는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하도록 구성되고, 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임은 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 양태들 19 내지 26 중 임의의 것의 장치.
- [0262] 양태 28: 프로세서는 디코더 부분에 의해, 재구성된 마스크를 생성하는 것으로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하고; 그리고 재구성된 마스크, 동작의 현재 시간 스텝에 대한 생성된 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임 생성하도록 구성되는, 양태들 19 내지 27 중 임의의 것의 장치.
- [0263] 양태 29: 프로세서는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고; 동작의 현재 시간 스텝 동안, 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하고; 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 상기 현재 시간 스텝에 대한 제 19 재구성된 잔차 데이터를 결정하고; 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하고; 동작의 상기 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 제 2 재구성된 잔차 데이터를 결정하고; 동작의 현재 시간 스텝 동안, 제 2 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 단계; 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하도록 구성되는, 양태들 19 내지 28 중 임의의 것의 장치.
- [0264] 양태 30: 프로세서는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵을 획득하고; 그리고 뉴럴 네트워크 시스템의 인코더 부분에 의해, 하나 이상의 왜곡 맵에 적어도 부분적으로 기초하여 출력 데이터를 생성하도록 구성되는, 양태들 19 내지 29 중 임의의 것의 장치.
- [0265] 양태 31: 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 동작의 상기 이전 시간 스텝으로부터의 이전에 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 1 왜곡 맵을 포함하는, 양태 30 의 장치.
- [0266] 양태 32: 하나 이상의 왜곡 맵은 동작의 상기 현재 시간 스텝에 대한 상기 입력 비디오 프레임과 워핑된 재구성된 비디오 프레임 사이의 왜곡을 표시하는 제 2 왜곡 맵을 포함하고, 상기 워핑된 재구성된 비디오 프레임은 동작의 상기 이전 시간 스텝으로부터 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 생성되는, 양태들 30 및 31 중 임의의 것의 장치.
- [0267] 양태 33: 프로세서는 뉴럴 네트워크 시스템의 인코더 부분의 제 1 계층에 의해, 입력 데이터를 프로세싱하고; 인코더 부분의 제 1 계층에 의해, 인코더 부분의 제 1 계층에 대한 복수의 가중치 값을 결정하고; 뉴럴 네트워크 시스템의 인코더 부분의 제 2 계층의 가중치들을 복수의 가중치 값으로 설정하며; 그리고 인코더 부분의 제

2 계층에 의해, 복수의 가중치 값으로 설정된 가중치들을 사용하여 입력 데이터를 프로세싱하도록 구성되는, 양태들 19 내지 32 중 임의의 것의 장치.

- [0268] 양태 34: 프로세서는 출력 데이터가 메모리에 저장되게 하도록 구성되는, 양태들 19 내지 33 중 임의의 것의 장치.
- [0269] 양태 35: 프로세서는 재구성된 비디오 프레임이 메모리에 저장되게 하도록 구성되는, 양태들 19 내지 34 중 임의의 것의 장치.
- [0270] 양태 36: 재구성된 비디오 프레임을 송신 매체를 통해 적어도 하나의 디바이스에 송신하도록 구성된 송신기를 더 포함하는, 양태들 19 내지 35 중 임의의 것의 장치.
- [0271] 양태 37: 프로세서는 뉴럴 프로세싱 유닛 (NPU) 을 포함하는, 양태들 19 내지 36 중 임의의 것의 장치.
- [0272] 양태 38: 장치는 모바일 디바이스를 포함하는, 양태들 19 내지 37 중 임의의 것의 장치.
- [0273] 양태 39: 장치는 확장 현실 디바이스를 포함하는, 양태들 19 내지 37 중 임의의 것의 장치.
- [0274] 양태 40: 장치는 텔레비전을 포함하는, 양태들 19 내지 37 중 임의의 것의 장치.
- [0275] 양태 41: 디스플레이를 더 포함하는, 양태들 19 내지 39 중 임의의 것의 장치.
- [0276] 양태 42: 장치는 하나 이상의 비디오 프레임을 캡처하도록 구성된 카메라를 포함하는, 양태들 19 내지 41 중 임의의 것의 장치.
- [0277] 양태 43: 비디오 데이터를 프로세싱하는 방법으로서, 방법은 뉴럴 네트워크 시스템의 디코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하는 단계; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하는 단계; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하는 단계; 및 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 포함한다.
- [0278] 양태 44: 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 및 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태 43 의 방법.
- [0279] 양태 45: 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 뉴럴 네트워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 결정하는 단계; 동작의 현재 시간 스텝 동안, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하는 단계; 및 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하는 단계를 더 포함하는, 양태들 43 및 44 중 임의의 것의 방법.
- [0280] 양태 46: 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하는 단계; 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하는 단계; 및 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태들 43 내지 45 중 임의의 것의 방법.
- [0281] 양태 47: 동작의 이전 시간 스텝에 대한 결정된 모션 추정 또는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하는 단계; 및 워핑된 순환 상태 데이터에 적어도

부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태들 43 내지 46 중 임의의 것의 방법.

[0282] 양태 48: 디코더 부분에 의해, 재구성된 마스크를 생성하는 단계로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하는 단계; 및 재구성된 마스크, 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 단계로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하는 단계를 더 포함하는, 양태들 43 내지 47 중 임의의 것의 방법.

[0283] 양태 49: 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하는 단계; 및 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 단계로서, 출력 데이터는 상기 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하는 단계를 더 포함하는, 양태들 43 내지 48 중 임의의 것의 방법.

[0284] 양태 50: 비디오 데이터를 프로세싱하기 위한 장치. 장치는 메모리(예를 들어, 회로부로 구현됨) 및 메모리에 커플링된 프로세서(또는 다중 프로세서들)를 포함한다. 프로세서(또는 프로세서들)는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한 출력 데이터 및 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 모션 추정 데이터를 결정하고; 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임을 획득하고; 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 상기 이전에 재구성된 비디오 프레임의 하나 이상의 픽셀을 수정함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임을 생성하며; 그리고 워핑된 재구성된 비디오 프레임에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성된다.

[0285] 양태 51: 프로세서는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 워핑된 재구성된 비디오 프레임을 부가함으로써 적어도 부분적으로 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 양태 50의 장치.

[0286] 양태 52: 동작의 현재 시간 스텝 동안 디코더 부분에 의해, 동작의 현재 시간 스텝 동안 뉴럴 네트워크 시스템의 인코더 부분에 의해 프로세싱된 이전 입력 비디오 프레임에 적어도 부분적으로 기초하여 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 결정하고; 동작의 현재 시간 스텝 동안, 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터를 사용하여 이전에 재구성된 비디오 프레임을 수정하며; 그리고 수정된 이전에 재구성된 비디오 프레임을 동작의 현재 시간 스텝에 대한 최종 출력 비디오 프레임으로서 출력하도록 구성되는, 양태들 50 및 51 중 임의의 것의 장치.

[0287] 양태 53: 프로세서는 뉴럴 네트워크 시스템의 디코더 부분에 의해, 동작의 이전 시간 스텝으로부터의 순환 상태 데이터 및 출력 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 잔차 데이터를 결정하고; 디코더 부분의 포스트-워핑 네트워크에 의해, 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임을 프로세싱하고; 포스트-워핑 네트워크에 의한 재구성된 잔차 데이터 및 워핑된 재구성된 비디오 프레임의 프로세싱에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 양태들 50 내지 52 중 임의의 것의 장치.

[0288] 양태 54: 프로세서는 동작의 이전 시간 스텝에 대한 결정된 모션 추정 또는 동작의 현재 시간 스텝에 대한 결정된 모션 추정 데이터를 사용하여 동작의 이전 시간 스텝으로부터 순환 상태 데이터를 수정함으로써 적어도 부분

적으로 동작의 현재 시간 스텝에 대한 워핑된 순환 상태 데이터를 생성하고; 워핑된 순환 상태 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하도록 구성되는, 양태들 50 내지 53 중 임의의 것의 장치.

[0289] 양태 55: 프로세서는 디코더 부분에 의해, 재구성된 마스크를 생성하는 것으로서, 재구성된 마스크는 동작의 이전 시간 스텝 동안 생성된 이전에 재구성된 비디오 프레임에서의 데이터의 존재를 표시하는 하나 이상의 값을 포함하는, 상기 재구성된 마스크를 생성하고; 그리고 재구성된 마스크, 동작의 현재 시간 스텝에 대한 워핑된 재구성된 비디오 프레임, 및 동작의 현재 시간 스텝에 대한 결정된 재구성된 잔차 데이터에 적어도 부분적으로 기초하여 동작의 현재 시간 스텝에 대한 재구성된 비디오 프레임을 생성하는 것으로서, 재구성된 마스크는 워핑된 재구성된 비디오 프레임 및 재구성된 잔차 데이터의 기여들을 가중하는, 상기 재구성된 비디오 프레임을 생성하도록 구성되는, 양태들 50 내지 54 중 임의의 것의 장치.

[0290] 양태 56: 프로세서는 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 모션 추정 데이터, 뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터 뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의 순환 계층으로부터의 순환 상태 데이터를 획득하고; 뉴럴 네트워크 시스템의 인코더 부분에 의해, 동작의 현재 시간 스텝에 대한 출력 데이터를 생성하는 것으로서, 출력 데이터는 상기 입력 비디오 프레임, 동작의 이전 시간 스텝으로부터의 상기 재구성된 모션 추정 데이터, 동작의 이전 시간 스텝으로부터의 상기 재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터의 순환 상태 데이터에 적어도 부분적으로 기초하여 생성되는, 상기 출력 데이터를 생성하도록 구성되는, 양태들 50 내지 55 중 임의의 것의 장치.

[0291] 양태 57: 프로세서는 재구성된 비디오 프레임이 메모리에 저장되게 하도록 구성되는, 양태들 50 내지 56 중 임의의 것의 장치.

[0292] 양태 58: 프로세서는 재구성된 비디오 프레임이 디스플레이되게 하도록 구성되는, 양태들 50 내지 57 중 임의의 것의 장치.

[0293] 양태 59: 재구성된 비디오 프레임을 디스플레이하도록 구성된 디스플레이를 더 포함하는, 양태들 50 내지 58 중 임의의 것의 장치.

[0294] 양태 60: 프로세서는 뉴럴 프로세싱 유닛 (NPU) 을 포함하는, 양태들 50 내지 59 중 임의의 것의 장치.

[0295] 양태 61: 장치는 모바일 디바이스를 포함하는, 양태들 50 내지 60 중 임의의 것의 장치.

[0296] 양태 62: 장치는 확장 현실 디바이스를 포함하는, 양태들 50 내지 60 중 임의의 것의 장치.

[0297] 양태 63: 장치는 텔레비전을 포함하는, 양태들 50 내지 60 중 임의의 것의 장치.

[0298] 양태 64: 장치는 하나 이상의 비디오 프레임을 캡처하도록 구성된 카메라를 포함하는, 양태들 50 내지 63 중 임의의 것의 장치.

[0299] 양태 65: 명령들이 저장된 컴퓨터 판독가능 저장 매체로서, 명령들은 실행될 때, 하나 이상의 프로세서로 하여금, 양태들 1 내지 18 의 동작들 중 임의의 것을 수행하게 한다.

[0300] 양태 66: 양태들 1 내지 18 의 동작들 중 임의의 것을 수행하는 수단을 포함하는 장치.

[0301] 양태 67: 명령들이 저장된 컴퓨터 판독가능 저장 매체로서, 명령들은 실행될 때, 하나 이상의 프로세서로 하여금, 양태들 43 내지 49 의 동작들 중 임의의 것을 수행하게 한다.

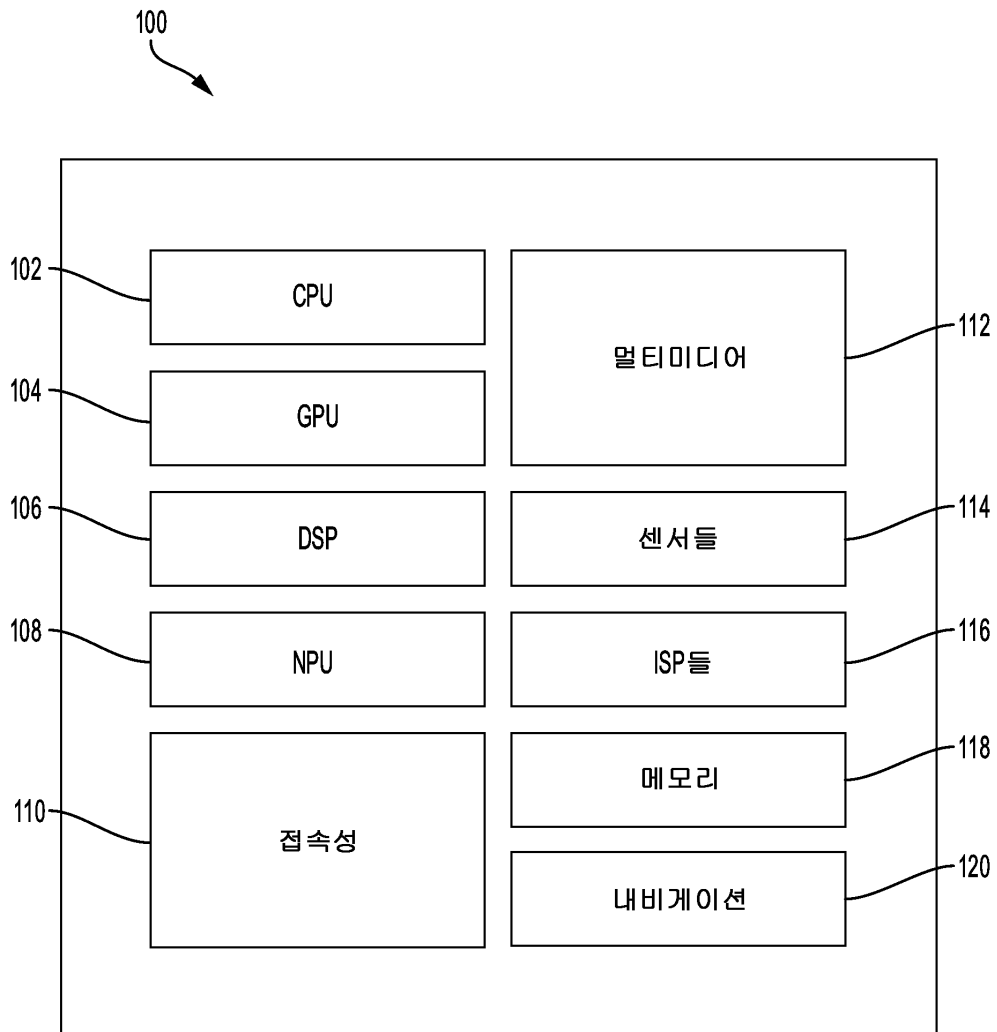
[0302] 양태 68: 양태들 43 내지 49 의 동작들 중 임의의 것을 수행하는 수단을 포함하는 장치.

[0303] 양태 69: 명령들이 저장된 컴퓨터 판독가능 저장 매체로서, 명령들은 실행될 때, 하나 이상의 프로세서로 하여금, 양태들 1 내지 18 의 동작들 중 임의의 것 및 양태들 43 내지 49 의 동작들 중 임의의 것을 수행하게 한다.

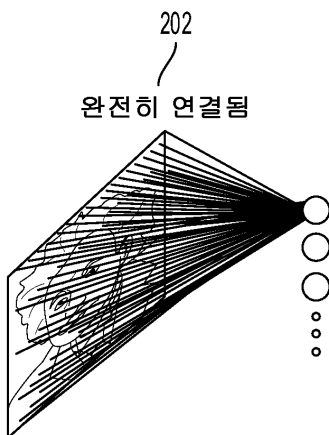
[0304] 양태 70: 양태들 1 내지 18 의 동작들 중 임의의 것 및 양태들 43 내지 49 의 동작들 중 임의의 것을 수행하는 수단을 포함하는 장치.

도면

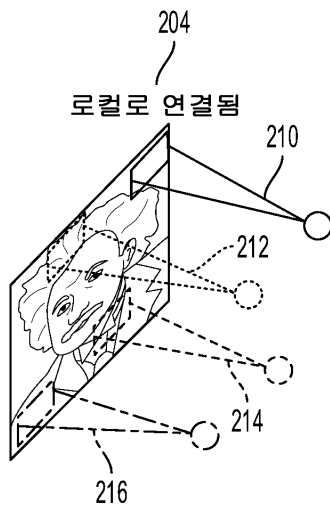
도면1



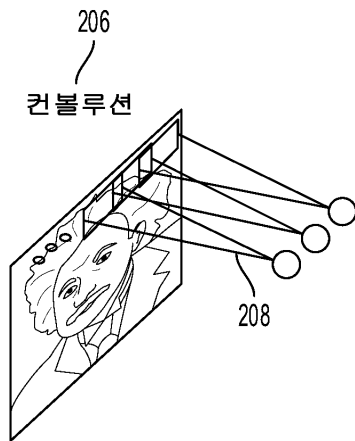
도면2a



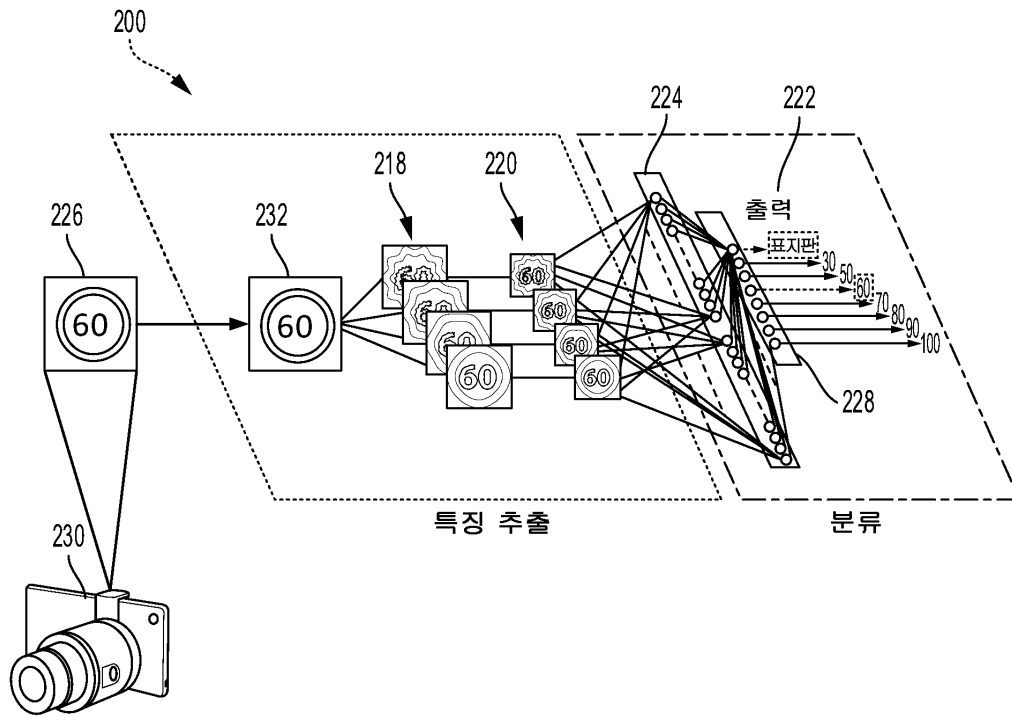
도면2b



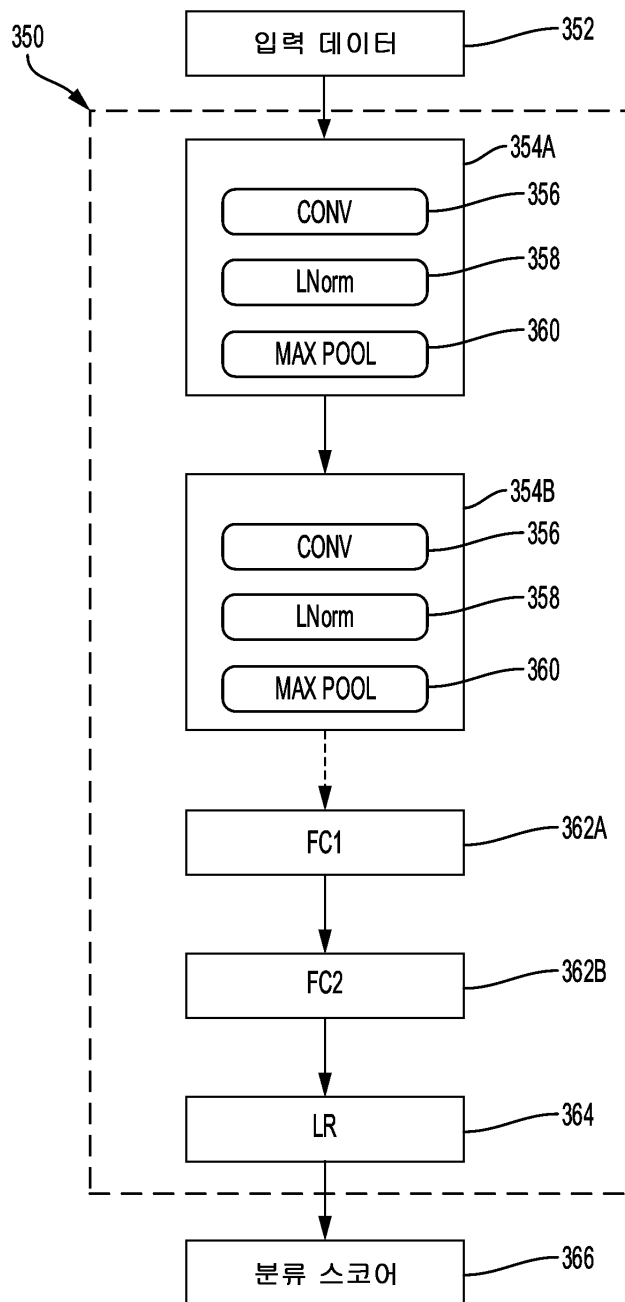
도면2c



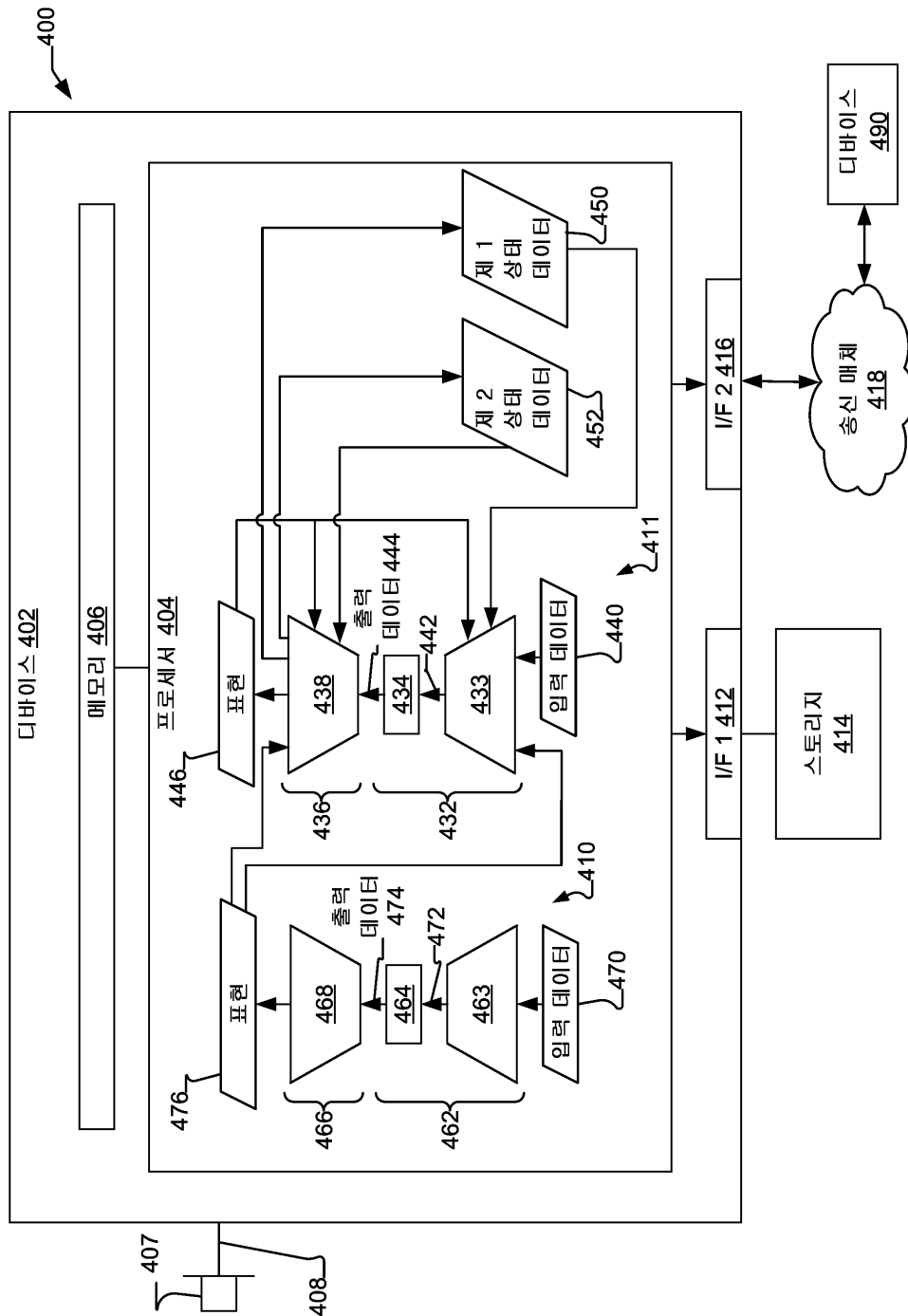
도면2d



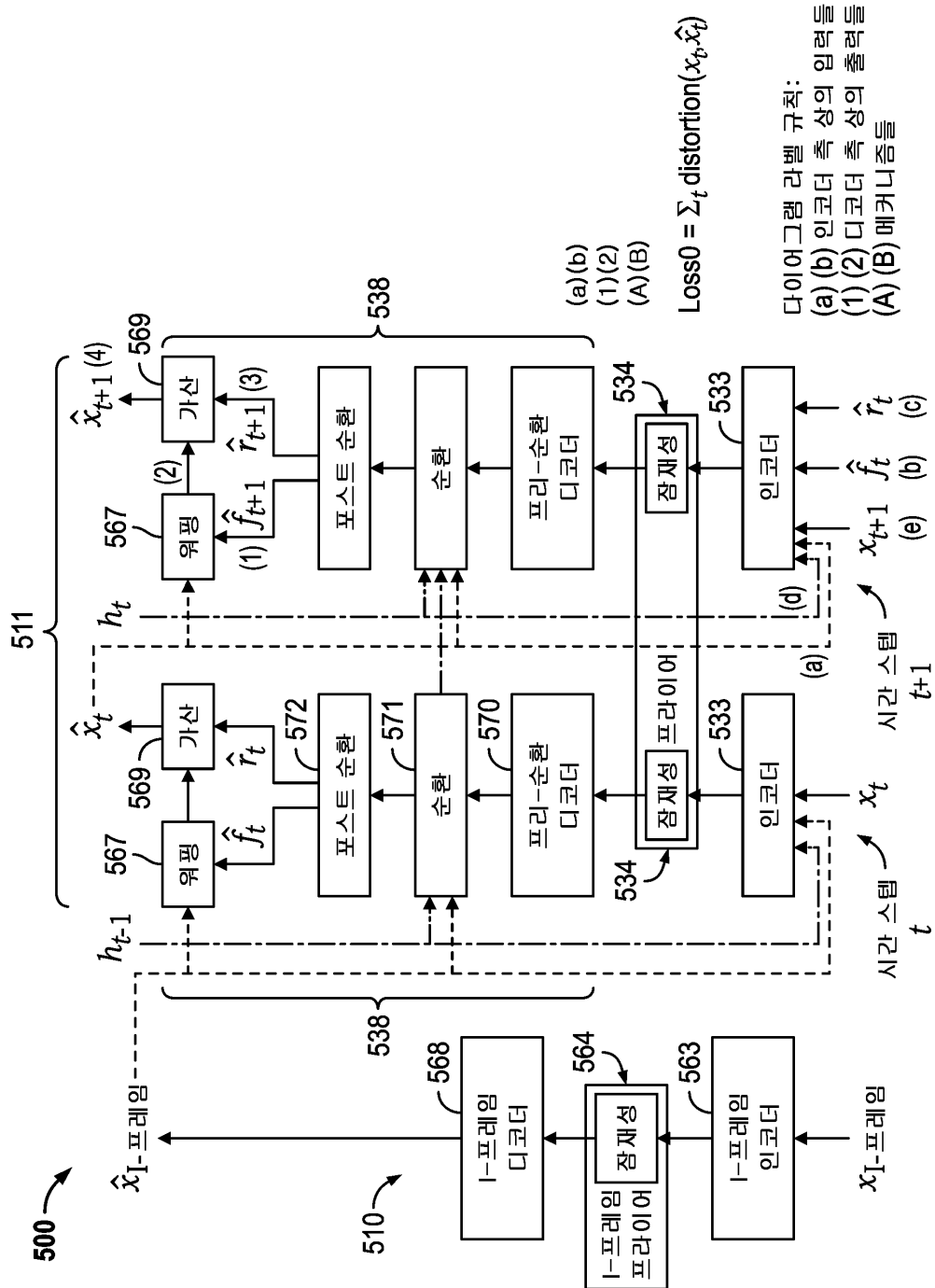
도면3



도면4



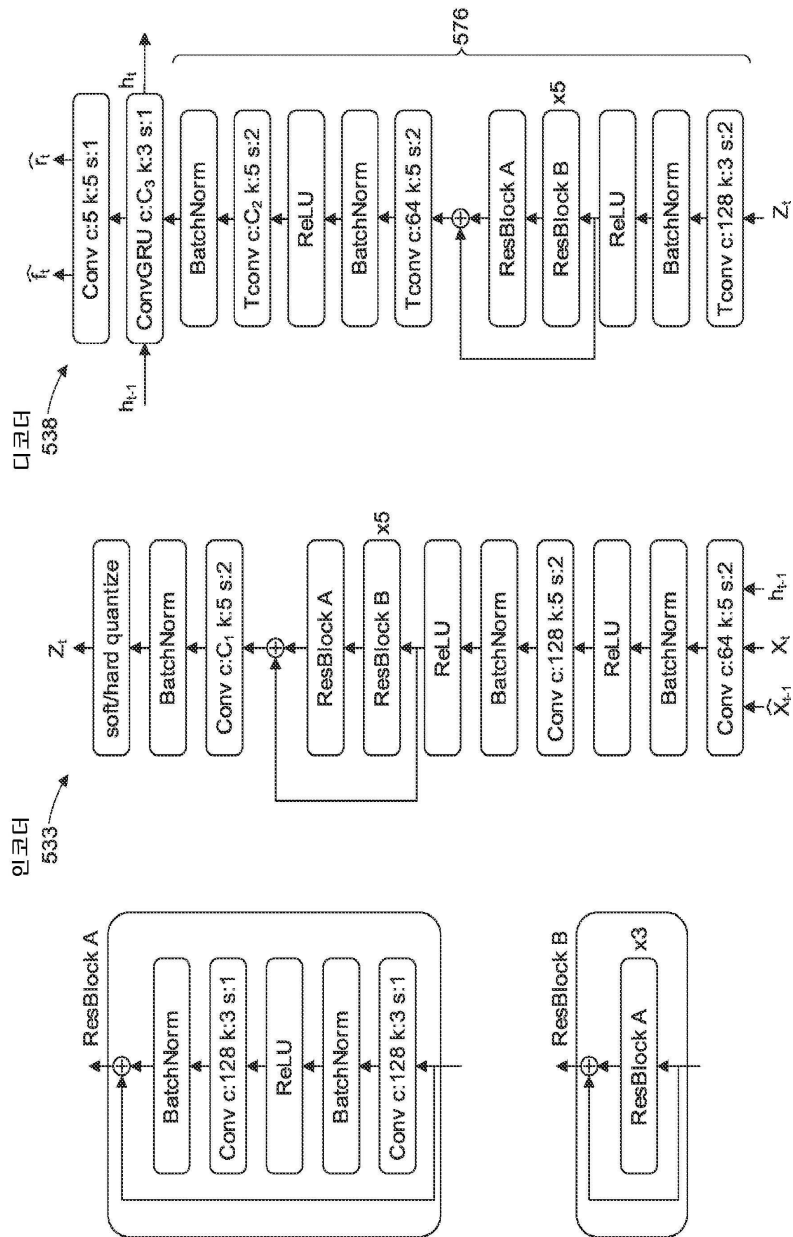
도면 5a



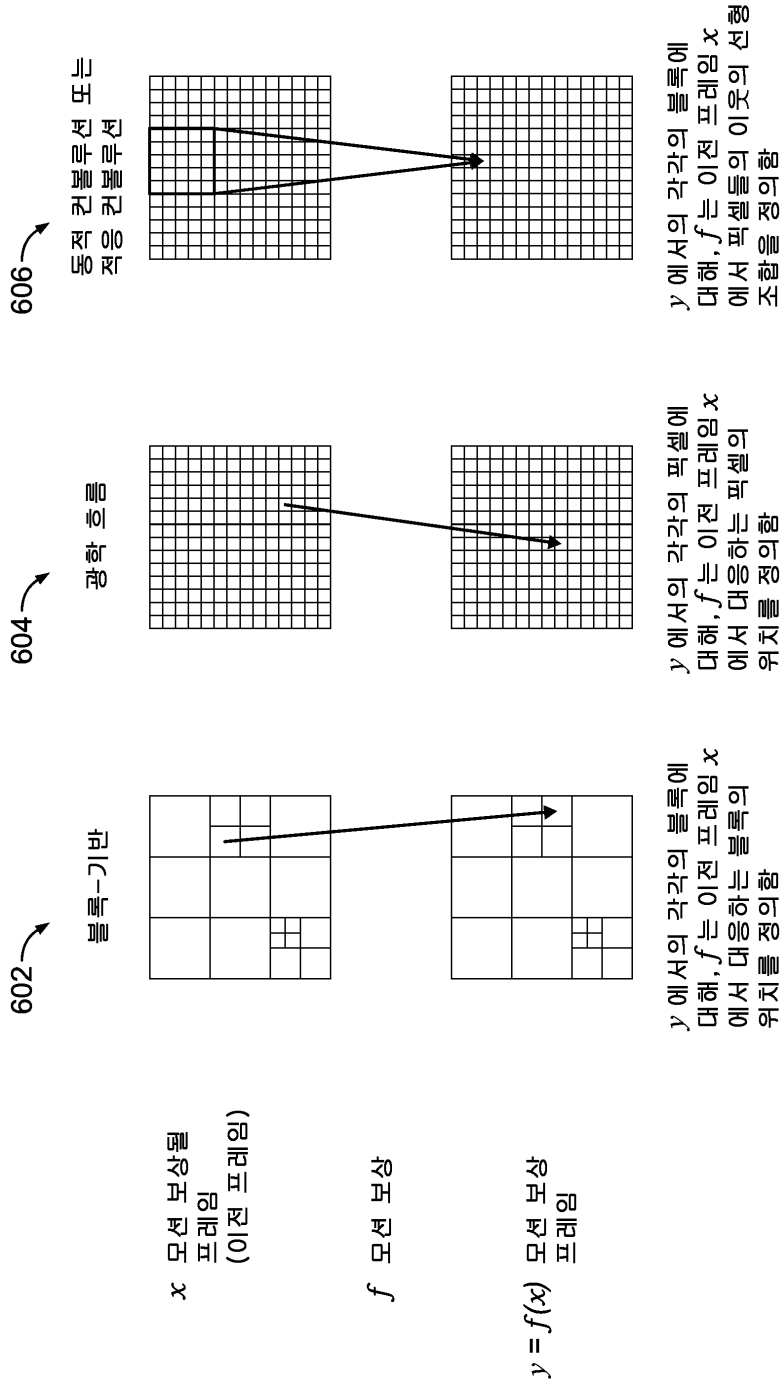
다이어그램 라벨 규칙:

(a) (b) 인코더 출력 상의 입력
(1) (2) 디코더 출력 상의 출력
(A) (B) 메커니즘들

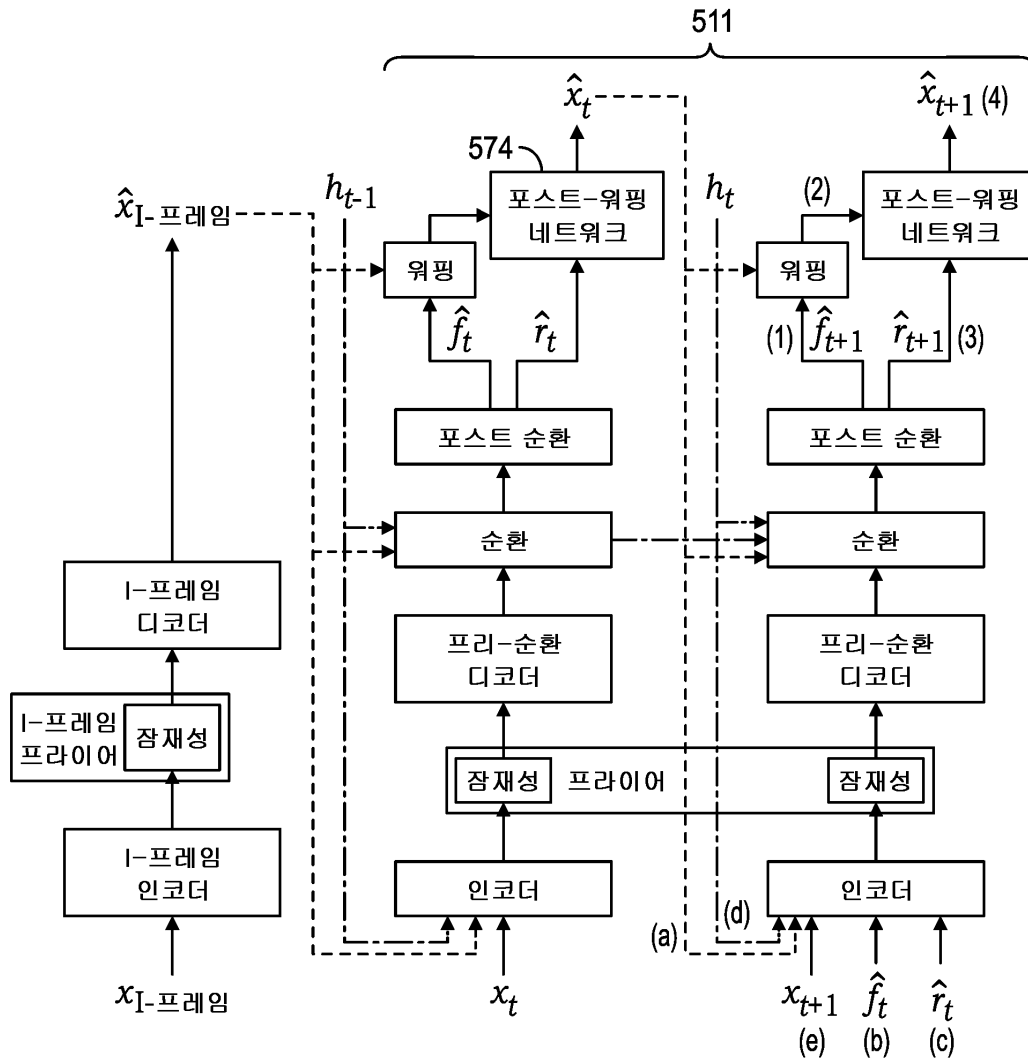
도면5b



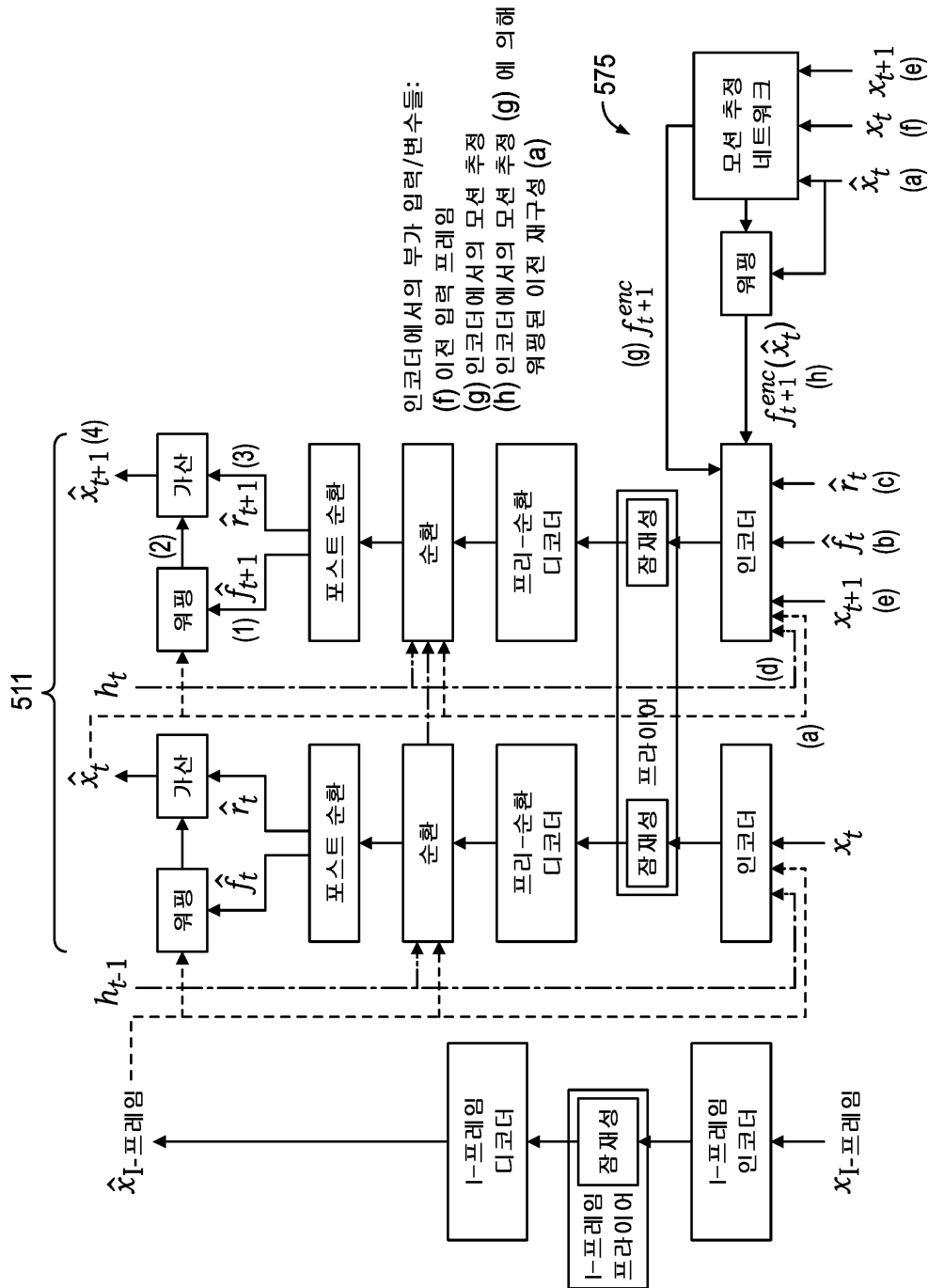
도면6



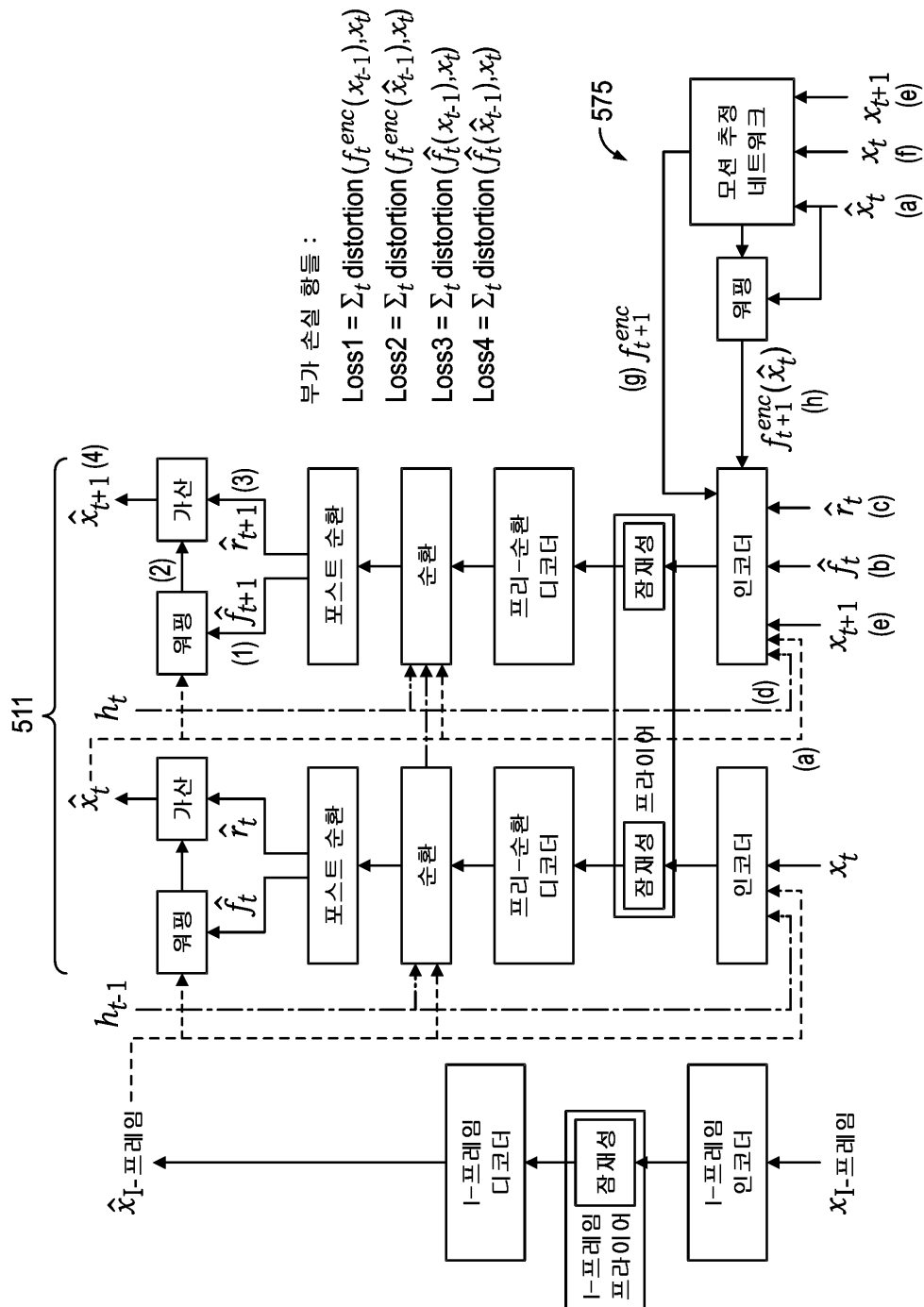
도면7



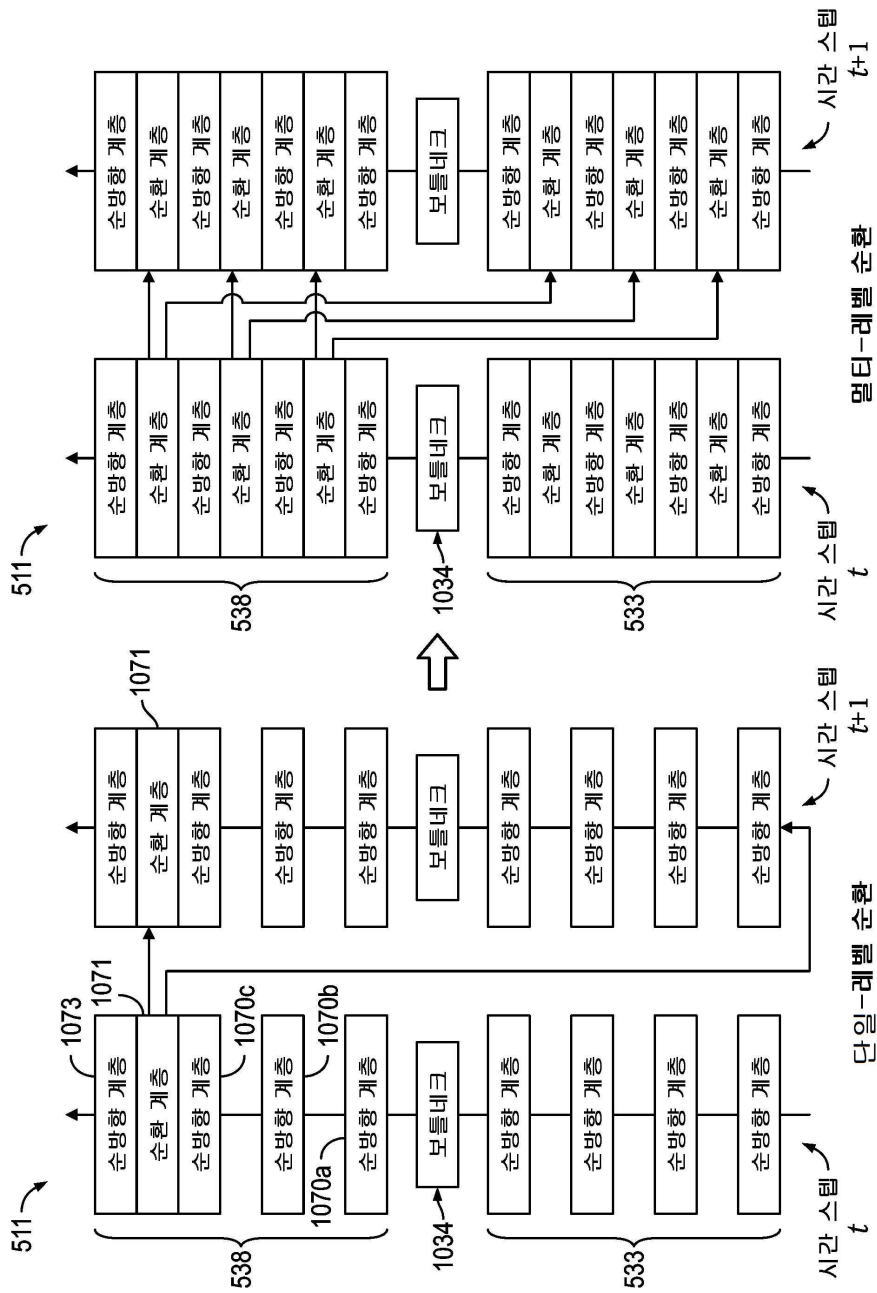
도면8



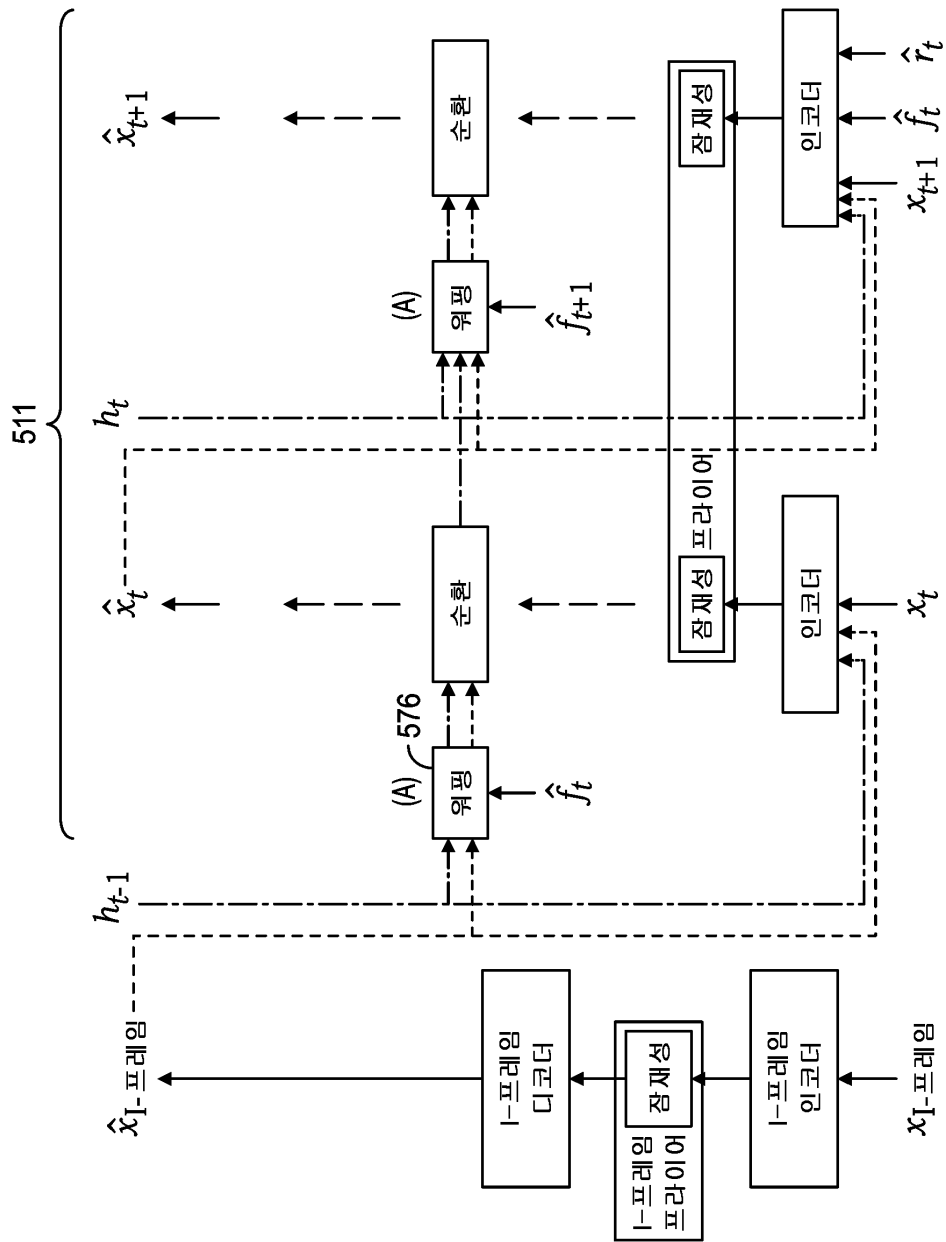
도면9



도면10

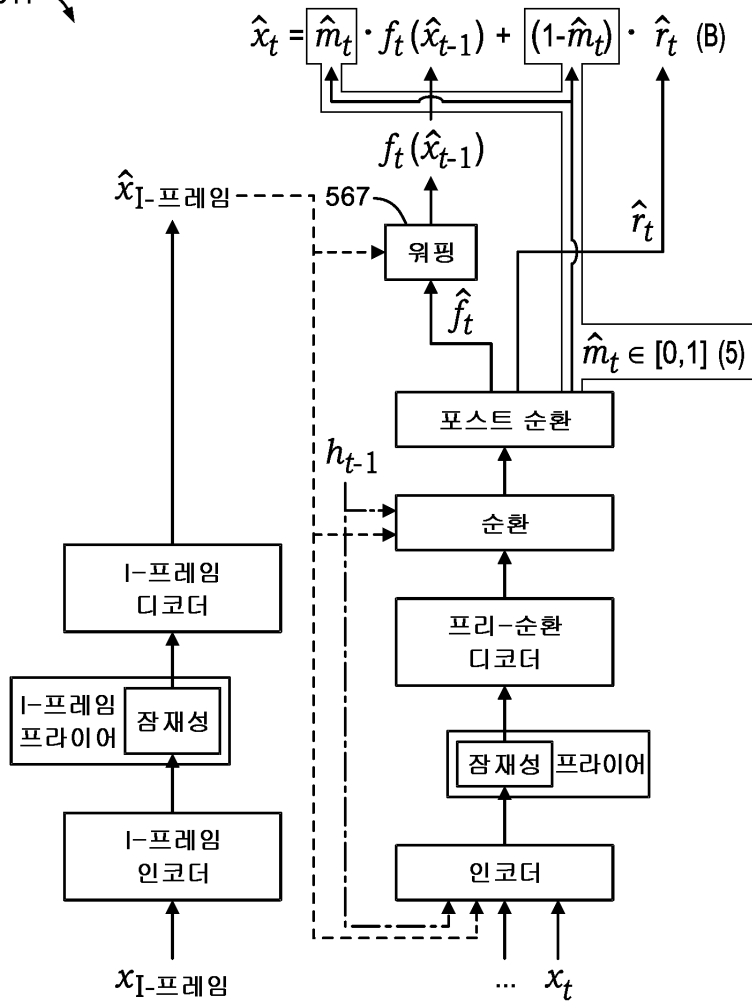


도면11

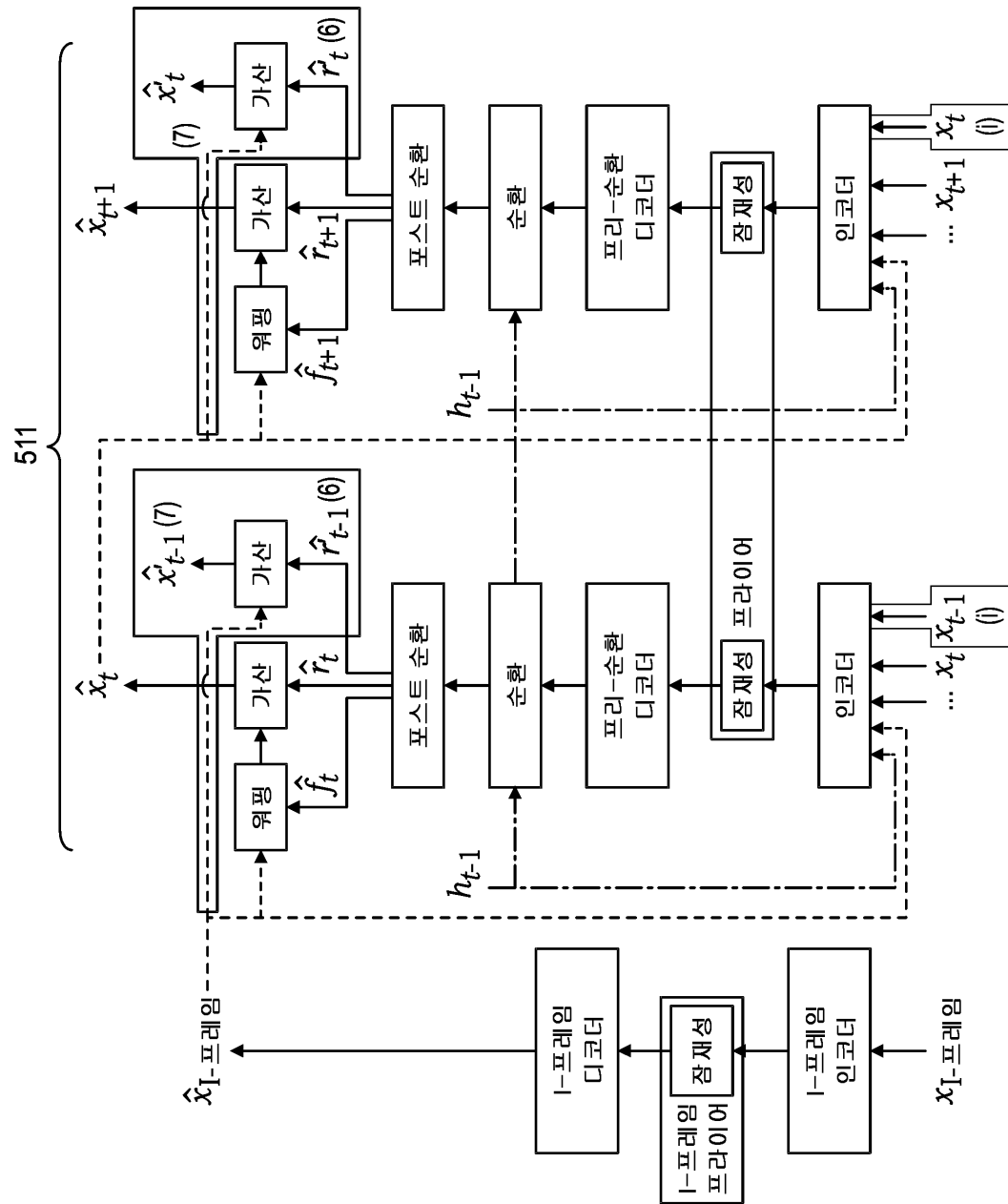


도면12

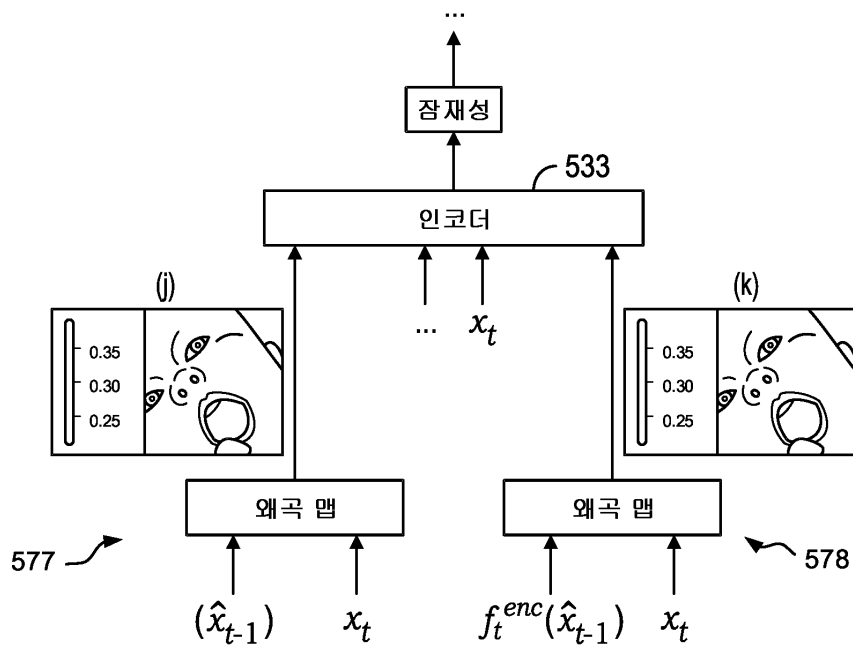
511 ↗



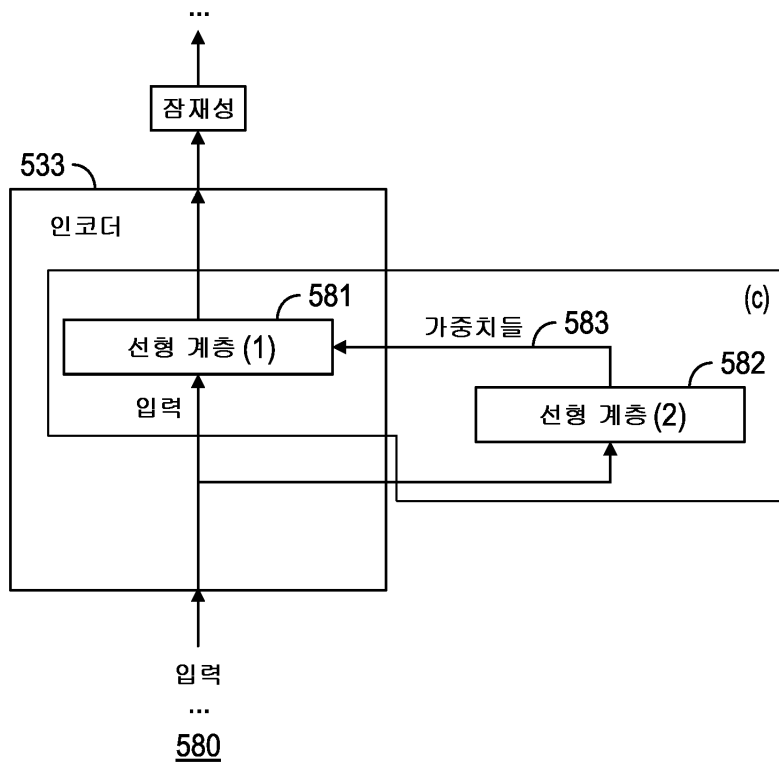
도면13



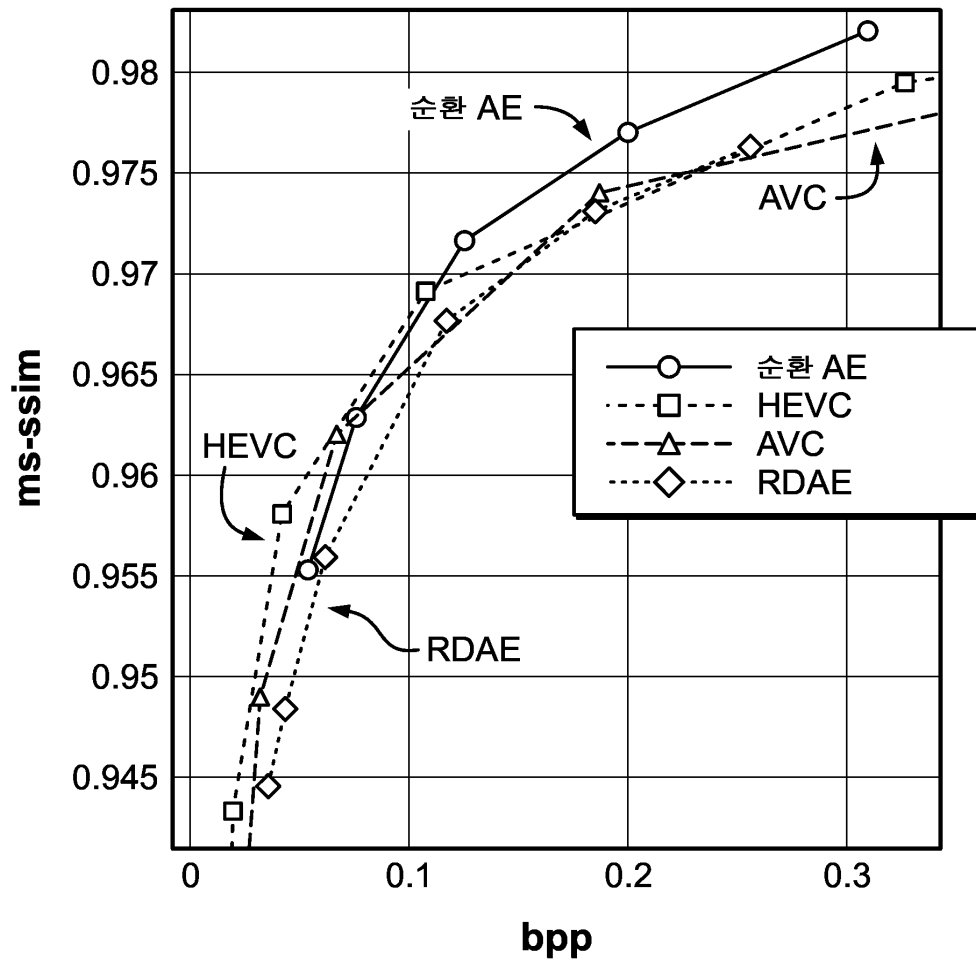
도면14



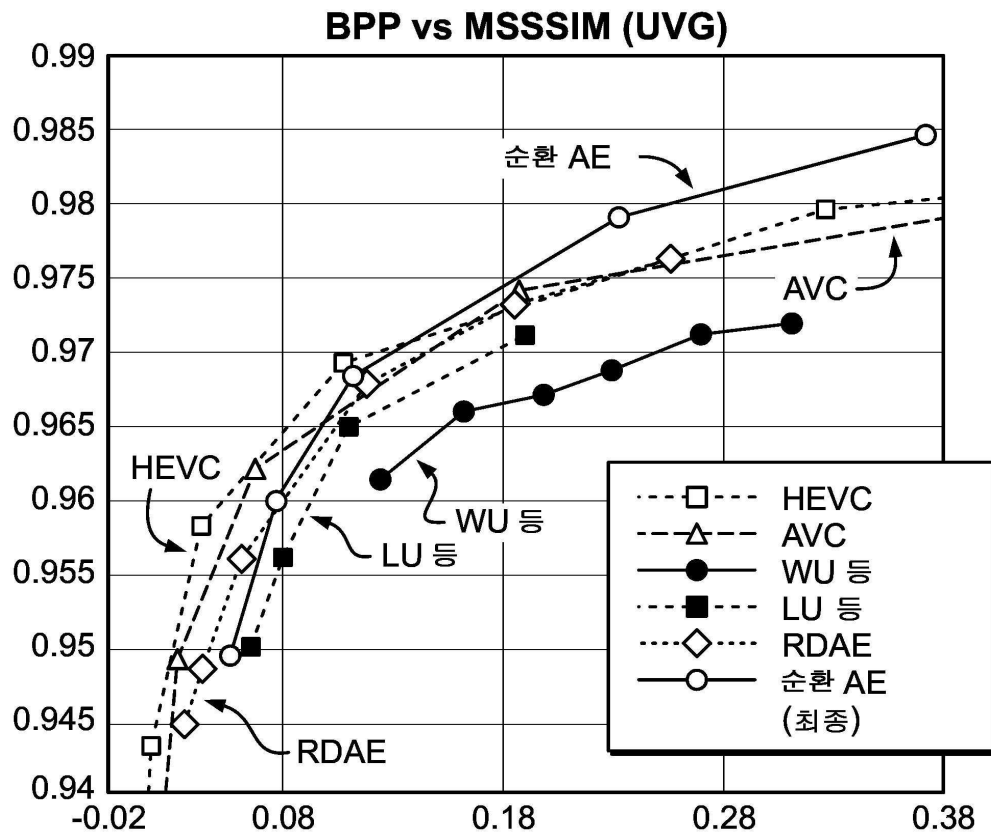
도면15



도면16



도면17



도면18

1800

뉴럴 네트워크 시스템의 인코더 부분에 의해,
뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한
입력 비디오 프레임, 뉴럴 네트워크 시스템의 동작의
이전 시간 스텝으로부터의 재구성된 모션 추정 데이터,
뉴럴 네트워크 시스템의 동작의 이전 시간 스텝으로부터의
재구성된 잔차 데이터, 및 동작의 이전 시간 스텝으로부터
뉴럴 네트워크 시스템의 디코더 부분의 적어도 하나의
순환 계층으로부터의 적어도 하나의 순환 상태 데이터를 획득함

1802

뉴럴 네트워크 시스템의 인코더 부분에 의해,
뉴럴 네트워크 시스템의 동작의 현재 시간 스텝에 대한
출력 데이터를 생성함

1804

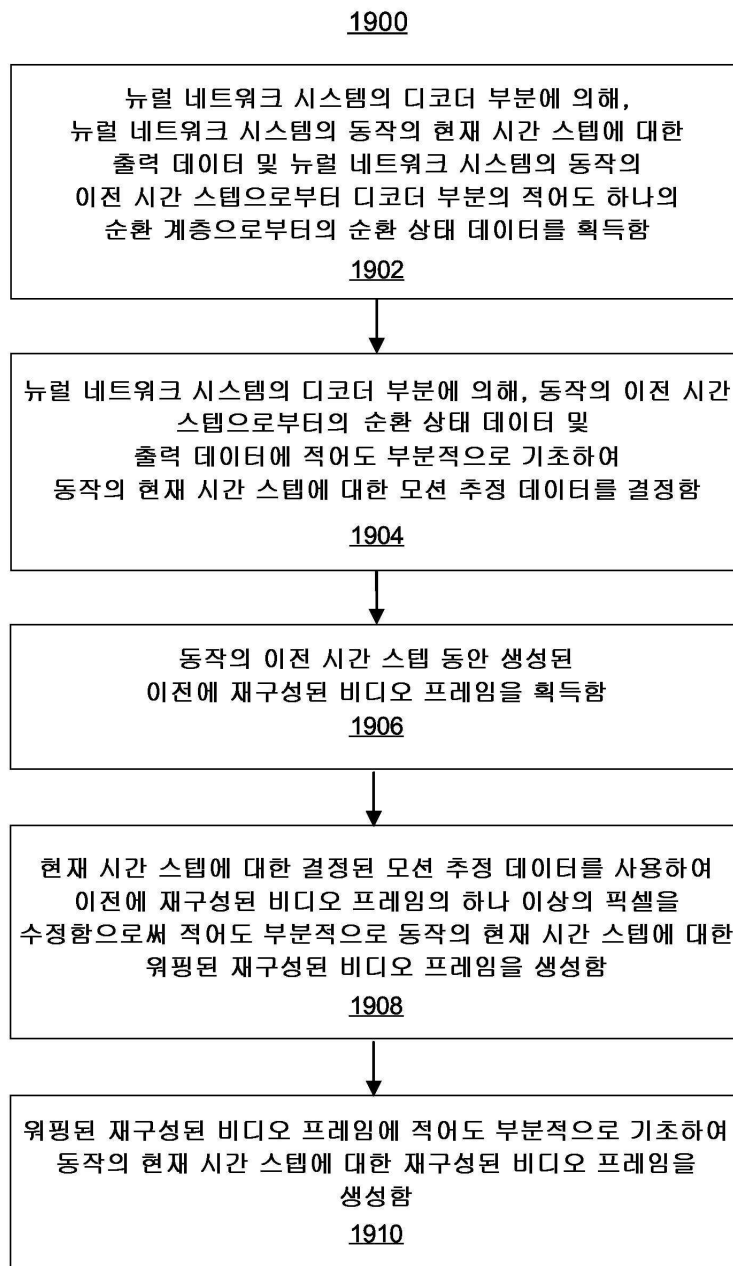
뉴럴 네트워크 시스템의 디코더 부분에 의해,
동작의 현재 시간 스텝에 대한 출력 데이터 및 동작의
이전 시간 스텝으로부터 디코더 부분의 적어도 하나의
순환 계층으로부터의 적어도 하나의 순환 상태 데이터를 획득함

1806

동작의 이전 시간 스텝으로부터의 적어도 하나의
순환 상태 데이터 및 출력 데이터에 적어도 부분적으로
기초하여 디코더 부분에 의해, 동작의 현재 시간 스텝에 대한
재구성된 비디오 프레임을 생성함

1808

도면19



도면20

