(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0021946
(43) 공개일자 2025년02월14일

(51) 국제특허분류(Int. Cl.)
H04N 19/31 (2014.01) *G06N 3/0464* (2023.01)
G06N 3/048 (2023.01) *H04N 19/105* (2014.01)
H04N 19/124 (2014.01) *H04N 19/13* (2014.01)
H04N 19/577 (2014.01)
(52) CPC특허분류
H04N 19/31 (2015.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0160501
(22) 출원일자 2023년11월20일
심사청구일자 없음
(30) 우선권주장
1020230102601 2023년08월07일 대한민국(KR)

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
김승언
경기도 수원시 영통구 삼성로 130 ,본관동5층(삼성미래기술캠퍼스)
김영웅
경기도 수원시 영통구 청명로 25 404호
(뒷면에 계속)
(74) 대리인
특허법인 신지

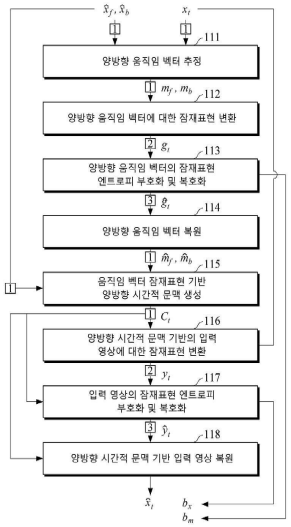
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **신경망 기반 비디오 부호화 방법 및 복호화 방법**

(57) 요 약

신경망 기반 비디오 부호화 방법이 개시된다. 일 실시예에 따르면 비디오 부호화 방법은 참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하는 단계, 상기 움직임 벡터를 잠재표현으로 변환하는 단계, 상기 움직임 벡터의 잠재표현을 시간적 계층에 따라 스케일링하는 단계, 및 상기 스케일링된 움직임 벡터의 잠재표현과 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 단계를 포함할 수 있다.

대 표 도 - 도1a



(52) CPC특허분류
G06N 3/048 (2023.01)
H04N 19/105 (2015.01)
H04N 19/124 (2015.01)
H04N 19/13 (2015.01)
H04N 19/577 (2015.01)

(72) 발명자
김휘용
대전시 유성구 은구비남로 34 열매마을 810동 201호
이원희
경기도 용인시 기흥구 금화로58번길 10, 403동 1405호 (상갈동, 상갈 메트로 파크)
박수용
경기도 용인시 기흥구 서그내로 52 희망의도시 401호
성영훈
경기도 화성시 여울로4길 67, 401호 (능동)
오도관
경기도 화성시 동탄대로시범길 235, 942동 1403호 (청계동, 동탄역 시범 금강펜테리움 센트럴파크 3)

명 세 서

청구범위

청구항 1

참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하는 단계;

상기 움직임 벡터를 잠재표현으로 변환하는 단계;

상기 움직임 벡터의 잠재표현을 상기 입력 영상의 시간적 계층에 따라 스케일링하는 단계; 및

상기 스케일링된 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 단계를 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 2

제1항에 있어서,

상기 참조 영상은

상기 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 3

제1항에 있어서,

상기 시간적 계층에 따라 스케일링하는 단계는

상기 움직임 벡터의 잠재표현을 상기 시간적 계층이 상위일수록 더 작은 양자화 구간을 사용하도록 스케일링하는, 신경망 기반 비디오 부호화 방법.

청구항 4

제1항에 있어서,

상기 시간적 계층에 따라 스케일링하는 단계는

상기 움직임 벡터의 잠재표현을 상기 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터로 나누어 스케일링하는, 신경망 기반 비디오 부호화 방법.

청구항 5

제1항에 있어서,

상기 움직임 벡터의 잠재표현을 엔트로피 부호화 및 복호화하는 단계;

상기 스케일링된 움직임 벡터의 잠재표현에 대하여 시간적 계층에 따라 리스케일링하는 단계; 및

상기 리스케일링된 움직임 벡터의 잠재표현을 기초로 움직임 벡터들을 복원하는 단계를 더 포함하고,

상기 시간적 문맥을 생성하는 단계는

상기 복원된 움직임 벡터와 상기 참조 영상을 기초로 상기 시간적 문맥을 생성하는, 신경망 기반 비디오 부호화 방법.

청구항 6

제5항에 있어서,

상기 시간적 계층에 따라 리스케일링하는 단계는

상기 스케일링된 움직임 벡터의 잠재표현에 상기 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터를 곱하

여 리스케일링하는, 신경망 기반 비디오 부호화 방법.

청구항 7

제5항에 있어서,

상기 엔트로피 부호화 및 복호화 하는 단계 이전에,

상기 스케일링된 움직임 벡터의 잠재표현을 양자화하는 단계; 및

상기 움직임 벡터의 잠재표현에 대한 확률 분포를 추정하는 단계를 더 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 8

제5항에 있어서

상기 시간적 문맥을 생성하는 단계는

상기 참조 영상을 특징 추출 신경망에 입력하여 참조 특징맵을 출력하고, 상기 복원된 움직임 벡터를 기초로 참조 특징맵에 대해 쌍선형 와핑을 수행하여 와핑된 참조 특징맵을 출력하고, 상기 와핑된 참조 특징맵을 후처리 신경망에 입력하고, 상기 후처리 신경망의 출력을 시간적 문맥 융합 신경망에 입력하여 상기 시간적 문맥을 출력하는, 신경망 기반 비디오 부호화 방법.

청구항 9

참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하는 단계;

상기 움직임 벡터를 잠재표현으로 변환하는 단계;

상기 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 단계; 및

상기 참조 영상, 상기 입력 영상, 상기 움직임 벡터, 상기 시간적 문맥 생성 단계의 입력, 상기 시간적 문맥 생성 단계의 출력, 및 상기 시간적 문맥 생성 단계의 세부 과정의 입력 또는 출력 중의 적어도 하나를 평활화 대상으로 하여 평활화를 수행하는 단계를 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 10

제9항에 있어서,

상기 참조 영상은

상기 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 11

제9항에 있어서,

상기 평활화하는 단계는

상기 입력 영상의 시간적 계층에 따라 상기 평활화 대상의 평활화 수준을 결정하는 단계;

상기 결정된 평활화 수준을 이용하여 미리 정의된 평활화 필터를 생성하는 단계; 및

상기 생성된 평활화 필터를 이용하여 상기 평활화 대상에 대한 평활화를 수행하는 단계를 포함하는, 신경망 기반 비디오 부호화 방법.

청구항 12

제11항에 있어서,

상기 평활화 대상의 평활화 수준을 결정하는 단계는

상기 시간적 계층이 하위일수록 더 낮은 평활화 수준을 결정하는, 신경망 기반 비디오 부호화 방법.

청구항 13

제11항에 있어서,

상기 평활화 대상의 평활화 수준을 결정하는 단계는

공간적 위치에 따라 서로 다른 평활화 수준을 결정하는, 신경망 기반 비디오 부호화 방법.

청구항 14

움직임 벡터에 대한 제1 비트스트림을 입력으로 움직임 벡터의 잠재표현을 복원하는 단계; 및

상기 복원된 움직임 벡터의 잠재표현에 대하여 시간적 계층에 따라 리스케일링하는 단계; 및

상기 리스케일링된 움직임 벡터의 잠재표현을 이용하여 움직임 벡터를 복원하는 단계;

상기 복원된 움직임 벡터와 참조 영상을 입력으로 시간적 문맥을 생성하는 단계; 및

입력 영상에 대한 제2 비트스트림 및 상기 시간적 문맥을 입력으로 상기 입력 영상을 복원하는 단계를 포함하는, 신경망 기반 비디오 복호화 방법.

청구항 15

제14항에 있어서,

상기 참조 영상은

상기 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함하는, 신경망 기반 비디오 복호화 방법.

청구항 16

제14에 있어서,

상기 시간적 계층에 따라 리스케일링하는 단계는

상기 움직임 벡터의 잠재표현에 상기 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터를 곱하여 리스케일링하는, 신경망 기반 비디오 복호화 방법.

청구항 17

제14항에 있어서,

상기 입력 영상을 복원하는 단계는

상기 입력 영상에 대한 제2 비트스트림을 입력으로 입력 영상의 잠재 표현을 복원하는 단계;

상기 입력 영상의 잠재 표현을 시간적 계층에 따라 리스케일링하는 단계; 및

상기 리스케일링된 입력 영상의 잠재 표현과 상기 시간적 문맥을 기초로 입력 영상을 복원하는 단계를 포함하는, 신경망 기반 비디오 복호화 방법.

청구항 18

움직임 벡터에 대한 제1 비트스트림을 입력으로 움직임 벡터를 복원하는 단계;

상기 움직임 벡터와 참조 영상을 입력으로 시간적 문맥을 생성하는 단계;

상기 참조 영상, 상기 움직임 벡터, 상기 시간적 문맥 생성 단계의 입력, 상기 시간적 문맥 생성 단계의 출력, 및 상기 시간적 문맥 생성 단계의 세부 과정의 입력 또는 출력 중의 적어도 하나를 평활화 대상으로 하여 평활화를 수행하는 단계; 및

입력 영상에 대한 제2 비트스트림 및 상기 시간적 문맥을 입력으로 상기 입력 영상을 복원하는 단계를 포함하는, 신경망 기반 비디오 복호화 방법.

청구항 19

제18항에 있어서,

상기 평활화를 수행하는 단계는

상기 입력 영상의 시간적 계층 및 공간적 위치 중의 적어도 하나에 기초하여 서로 다른 평활화 수준을 결정하는 단계;

상기 결정된 평활화 수준을 이용하여 미리 정의된 평활화 필터를 생성하는 단계; 및

상기 생성된 평활화 필터를 이용하여 상기 평활화 대상에 대한 평활화를 수행하는 단계를 포함하는, 신경망 기반 비디오 복호화 방법.

청구항 20

하나 이상의 명령어들, 및 참조 영상을 저장하는 메모리; 및

상기 하나 이상의 명령어들을 실행하여, 비디오 부호화를 수행하는 프로세서를 포함하고,

상기 하나 이상의 명령어들은,

참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하고, 상기 움직임 벡터를 잠재표현으로 변환하고, 상기 움직임 벡터의 잠재표현을 시간적 계층에 따라 스케일링하고, 상기 스케일링된 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 것을 포함하는, 전자장치.

발명의 설명

기 술 분 야

[0001] 신경망 기반 비디오 부호화 방법에 관한 것으로, 양방향 예측 및 시간적 계층 정보를 활용한 신경망 기반의 비디오 부호화 기술과 관련된다.

배 경 기 술

[0002] 최근 인터넷 동영상 시장 규모가 지속적으로 증가하고 있다. 비디오는 음성, 문자, 사진 등의 미디어에 비해 데이터 양이 월등하기 때문에 네트워크 대역폭에 따라 서비스 형태나 품질이 제약 받을 수 있어 보다 높은 비디오 부호화 기술이 요구된다. 이를 위해 일반적인 비디오 부호화 기술의 구성 요소들을 신경망으로 대체하여 종단간(end-to-end) 학습이 가능하도록 한 신경망 기반의 비디오 부호화 기술이 제안되었다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 미국 등록특허공보 US 11,399,198 (2022.7.26)

발명의 내용

해결하려는 과제

[0004] 양방향 예측 및 시간적 계층 정보를 활용한 신경망 기반의 비디오 부호화 및 복호화 방법이 제시된다.

과제의 해결 수단

[0005] 일 양상에 따르면, 신경망 기반 비디오 부호화 방법은 참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하는 단계, 움직임 벡터를 잠재표현으로 변환하는 단계, 움직임 벡터의 잠재표현을 상기 입력 영상의 시간적 계층에 따라 스케일링하는 단계, 및 스케일

링된 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 단계를 포함할 수 있다.

[0006] 참조 영상은 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함할 수 있다.

[0007] 시간적 계층에 따라 스케일링하는 단계는 움직임 벡터의 잠재표현을 상기 시간적 계층이 상위일수록 더 작은 양자화 구간을 사용하도록 스케일링할 수 있다.

[0008] 시간적 계층에 따라 스케일링하는 단계는 움직임 벡터의 잠재표현을 상기 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터로 나누어 스케일링할 수 있다.

[0009] 신경망 기반 비디오 부호화 방법은 움직임 벡터의 잠재표현을 엔트로피 부호화 및 복호화하는 단계, 스케일링된 움직임 벡터의 잠재표현에 대하여 시간적 계층에 따라 리스케일링하는 단계 및 리스케일링된 움직임 벡터의 잠재표현을 기초로 움직임 벡터들을 복원하는 단계를 더 포함하고, 시간적 문맥을 생성하는 단계는 복원된 움직임 벡터와 상기 참조 영상을 기초로 상기 시간적 문맥을 생성할 수 있다.

[0010] 시간적 계층에 따라 리스케일링하는 단계는 스케일링된 움직임 벡터의 잠재표현에 상기 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터를 곱하여 리스케일링 할 수 있다.

[0011] 엔트로피 부호화 및 복호화 하는 단계 이전에, 스케일링된 움직임 벡터의 잠재표현을 양자화하는 단계, 및 움직임 벡터의 잠재표현에 대한 확률 분포를 추정하는 단계를 더 포함할 수 있다.

[0012] 시간적 문맥을 생성하는 단계는 참조 영상을 특징 추출 신경망에 입력하여 참조 특징맵을 출력하고, 복원된 움직임 벡터를 기초로 참조 특징맵에 대해 쌍선형 와핑을 수행하여 와핑된 참조 특징맵을 출력하고, 상기 와핑된 참조 특징맵을 후처리 신경망에 입력하고, 상기 후처리 신경망의 출력을 시간적 문맥 융합 신경망에 입력하여 상기 시간적 문맥을 출력할 수 있다.

[0013] 일 양상에 따르면, 신경망 기반 비디오 부호화 방법은 참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하는 단계, 움직임 벡터를 잠재표현으로 변환하는 단계, 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 상기 입력 영상에 대한 시간적 문맥을 생성하는 단계, 및 참조 영상, 상기 입력 영상, 상기 움직임 벡터, 상기 시간적 문맥 생성 단계의 입력, 상기 시간적 문맥 생성 단계의 출력, 및 상기 시간적 문맥 생성 단계의 세부 과정의 입력 또는 출력 중의 적어도 하나를 평활화 대상으로 하여 평활화를 수행하는 단계를 포함할 수 있다.

[0014] 이때, 참조 영상은 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함할 수 있다.

[0015] 평활화를 수행하는 단계는 입력 영상의 시간적 계층에 따라 상기 평활화 대상의 평활화 수준을 결정하는 단계, 결정된 평활화 수준을 이용하여 미리 정의된 평활화 필터를 생성하는 단계, 및 생성된 평활화 필터를 이용하여 상기 평활화 대상에 대한 평활화를 수행하는 단계를 포함할 수 있다.

[0016] 평활화 대상의 평활화 수준을 결정하는 단계는 시간적 계층이 하위일수록 더 낮은 평활화 수준을 결정할 수 있다.

[0017] 평활화 대상의 평활화 수준을 결정하는 단계는 공간적 위치에 따라 서로 다른 평활화 수준을 결정할 수 있다.

[0018] 일 양상에 따르면, 신경망 기반 비디오 복호화 방법은 움직임 벡터에 대한 제1 비트스트림을 입력으로 움직임 벡터의 잠재표현을 복원하는 단계, 복원된 움직임 벡터의 잠재표현에 대하여 시간적 계층에 따라 리스케일링하는 단계, 상기 리스케일링된 움직임 벡터의 잠재표현을 이용하여 움직임 벡터를 복원하는 단계, 복원된 움직임 벡터와 참조 영상을 입력으로 시간적 문맥을 생성하는 단계 및 입력 영상에 대한 제2 비트스트림 및 상기 시간적 문맥을 입력으로 상기 입력 영상을 복원하는 단계를 포함할 수 있다.

[0019] 참조 영상은 입력 영상을 기준으로 시간적으로 선후인 양방향 참조 영상을 포함할 수 있다.

[0020] 이때, 시간적 계층에 따라 리스케일링하는 단계는 움직임 벡터의 잠재표현에 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터를 곱하여 리스케일링 할 수 있다.

[0021] 입력 영상을 복원하는 단계는 입력 영상에 대한 제2 비트스트림을 입력으로 입력 영상의 잠재 표현을 복원하는 단계, 입력 영상의 잠재 표현을 시간적 계층에 따라 리스케일링하는 단계, 및 리스케일링된 입력 영상의 잠재표현과 상기 시간적 문맥을 기초로 입력 영상을 복원하는 단계를 포함할 수 있다.

[0022] 일 양상에 따르면, 움직임 벡터에 대한 제1 비트스트림을 입력으로 움직임 벡터를 복원하는 단계, 움직임 벡터

와 참조 영상을 입력으로 시간적 문맥을 생성하는 단계, 참조 영상, 상기 움직임 벡터, 상기 시간적 문맥 생성 단계의 입력, 상기 시간적 문맥 생성 단계의 출력, 및 상기 시간적 문맥 생성 단계의 세부 과정의 입력 또는 출력 중의 적어도 하나를 평활화 대상으로 하여 평활화를 수행하는 단계, 및 입력 영상에 대한 제2 비트스트림 및 상기 시간적 문맥을 입력으로 상기 입력 영상을 복원하는 단계를 포함할 수 있다.

[0023] 평활화를 수행하는 단계는 입력 영상의 시간적 계층 및 공간적 위치 중의 적어도 하나에 기초하여 서로 다른 평활화 수준을 결정하는 단계, 결정된 평활화 수준을 이용하여 미리 정의된 평활화 필터를 생성하는 단계, 및 생성된 평활화 필터를 이용하여 상기 평활화 대상에 대한 평활화를 수행하는 단계를 포함할 수 있다.

[0024] 일 양상에 따르면, 전자 장치는 하나 이상의 명령어들, 및 참조 영상을 저장하는 메모리, 및 하나 이상의 명령어들을 실행하여, 비디오 부호화를 수행하는 프로세서를 포함하고, 하나 이상의 명령어들은, 참조 영상, 입력 영상 및 입력 영상의 시간적 계층을 입력으로, 상기 입력 영상과 상기 참조 영상 사이의 움직임 벡터를 추정하고, 움직임 벡터를 잠재표현으로 변환하고, 상기 움직임 벡터의 잠재표현을 시간적 계층에 따라 스케일링하고, 상기 스케일링된 움직임 벡터의 잠재표현과 상기 참조 영상을 기초로 입력 영상에 대한 시간적 문맥을 생성하는 것을 포함할 수 있다.

발명의 효과

[0025] 시간적 계층 정보 및 양방향 정보를 활용하여 비디오 부호화 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0026] 도 1a는 일 실시예에 따른 신경망 기반의 비디오 부호화 방법의 흐름도이다.

도 1b는 일 실시예에 따른 신경망 기반의 비디오 복호화 방법의 흐름도이다.

도 2a는 다른 실시예에 따른 신경망 기반의 비디오 부호화 방법의 흐름도이다.

도 2b는 일 실시예에 따른 양방향 시간적 문맥 생성 단계의 흐름도이다.

도 3a 및 도 3b는 시간적 계층에 따른 스케일링 및 리스케일링을 설명하기 위한 도면이다.

도 4는 일 실시예에 따른 양방향 시간적 문맥 평활화 단계의 흐름도이다.

도 5는 시간적 계층에 따른 양방향 시간적 문맥 평활화를 설명하기 위한 도면이다.

도 6a는 일 실시예에 따른 입력 영상의 부호화 단계의 흐름도이다.

도 6b는 일 실시예에 따른 입력 영상의 복호화 단계의 흐름도이다.

도 7은 일 실시예에 따른 전자장치의 블록도이다.

도 8a 내지 도 8i는 일 실시예에 따른 비디오 부호화 과정의 전체 동작을 설명하기 위한 도면이다.

도 9a는 시간적 문맥 융합 신경망의 일 예이다.

도 9b는 확률분포를 추정하는 엔트로피 모델의 일 예이다.

도 10은 다른 실시예에 따른 전자장치의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0027] 기타 실시예들의 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다. 기재된 기술의 이점 및 특징, 그리고 그것들을 달성하는 방법은 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[0028] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의

결합으로 구현될 수 있다.

[0029] 도 1a는 일 실시예에 따른 신경망 기반의 비디오 부호화 방법의 흐름도이다.

[0030] 도 1a를 참조하여, 신경망 기반 비디오 부호화 장치에 의해 수행되는 양방향 예측 및/또는 시간적 계층 정보를 활용한 비디오 부호화 방법의 일 실시예를 설명한다. 양방향 예측을 활용하는 경우, 단방향 예측을 활용하는 경우에 비해 더 높은 부호화 성능을 달성할 수 있다. 양방향 예측을 활용한 비디오 부호화 방법의 흐름을 설명하고, 이후 시간적 계층 정보를 활용하는 방법에 대해 설명한다.

[0031] 먼저, 양방향 예측 기반 비디오 부호화 장치는 양방향 참조 영상($\hat{x}_f, \hat{x}_b$) 및 입력 영상(x_t)을 이용하여 양방향 움직임 벡터(m_f, m_b)를 추정할 수 있다(111). 입력 영상(x_t)은 비디오를 구성하는 연속된 영상 중에서 현재 시점(t)에서 부호화할 영상을 의미한다. 참조 영상은 입력 영상(x_t)을 부호화 또는 복호화 하기 위해 참조하는 영상으로 이미 복호화된 영상을 의미한다. 양방향 참조 영상은 입력 영상(x_t)의 시점(t) 대비 선/후 영상으로서 각각 순방향 참조 영상($\hat{x}_f$)과 역방향 참조 영상($\hat{x}_b$)이라 한다. 순방향 참조 영상과 역방향 참조 영상은 하나 이상일 수 있으나 이하 설명의 편의를 위해 각각 하나인 경우로 한정하여 설명한다.

[0032] 그 다음, 양방향 움직임 벡터(m_f, m_b)는 합성곱 신경망(Convolutional Neural Network) 등을 이용하여 움직임 벡터의 잠재표현(g_t)으로 변환할 수 있다(112). 여기서, 잠재표현(latent representation)은 입력 영상이나 움직임 정보를 입력으로 하는 신경망의 출력으로서 잠재특징(latent feature), 잠재벡터(latent vector) 등을 통칭한다.

[0033] 그 다음, 양방향 움직임 벡터의 잠재표현(g_t)은 양자화 및 엔트로피 부호화를 통해 양방향 움직임 벡터의 비트스트림(b_m)으로 표현되어 복호화기에 전송되고, 움직임 벡터의 비트스트림(b_m)으로부터 엔트로피 복호화를 통해 양자화된 움직임 벡터의 잠재표현($\hat{g}_t$)을 복원할 수 있다(113).

[0034] 그 다음, 양자화된 움직임 벡터의 잠재표현($\hat{g}_t$)을 이용하여 양방향 움직임 벡터($\hat{m}_f, \hat{m}_b$)를 복원할 수 있다(114).

[0035] 그 다음, 복원된 양방향 움직임 벡터($\hat{m}_f, \hat{m}_b$)와 양방향 참조 영상($\hat{x}_f, \hat{x}_b$)을 이용하여 양방향 시간적 문맥(C_t)을 생성할 수 있다(115). 양방향 시간적 문맥(C_t)은 단일 해상도를 갖는 특징맵 또는 둘 이상의 해상도를 갖는 특징맵일 수 있으며, 입력 영상(x_t)에 대한 예측 정보로서 이후 입력 영상(x_t)의 부호화 및 복호화에 이용될 수 있다.

[0036] 그 다음, 입력 영상(x_t)은 양방향 시간적 문맥(C_t) 기반의 잠재 표현 변환을 통해 입력 영상의 잠재 표현(y_t)으로 변환될 수 있다(116).

[0037] 그 다음, 입력 영상의 잠재 표현(y_t)은 엔트로피 부호화를 통해 입력 영상의 비트스트림(b_x)으로 표현되어 복호화기에 전송될 수 있으며, 입력 영상의 비트스트림(b_x) 엔트로피 복호화를 통해 양자화된 입력 영상의 잠재표현($\hat{y}_t$)을 복원할 수 있다(117).

[0038] 그 다음, 양자화된 입력 영상의 잠재표현($\hat{y}_t$)을 이용하여 입력 영상($\hat{x}_t$)을 복원할 수 있다(118). 이때, 추가적인 입력으로 양방향 시간적 문맥(C_t)을 사용할 수 있다. 복원된 영상($\hat{x}_t$)은 이후의 입력 영상에 대한 참조 영상으로 활용될 수 있다.

[0039] 본 실시예에서 양방향 예측을 활용한 비디오 부호화 방법은 시간적 계층(l_t)을 더 이용할 수 있다. 시간적 계층을 이용하는 방법은 시간적 계층에 따른 스케일링과 리스케일링하는 방법 및/또는 시간적 계층에 따른 평활화 방법을 포함할 수 있다.

[0040] 도 1a를 참조하면, '2' 위치에서 양방향 움직임 벡터의 잠재표현(g_t) 및/또는 입력 영상의 잠재표현(y_t)에 대해 시간적 계층에 따른 스케일링을 수행하고, '3' 위치에서 양방향 움직임 벡터의 잠재표현($\hat{g}_t$) 및/또는 입력 영상의 잠재표현($\hat{y}_t$)에 대해 리스케일링을 수행할 수 있다.

[0041] 일반적인 비디오 부호화 방법에서는 시간적 계층에 따라 서로 다른 부호화 방법과 신경망 모델을 사용한다. 이러한 방법은 시간적 계층에 따라 여러 신경망 모델을 저장해야 하기 때문에 메모리 효율 등 복잡도 관점에서 비효율적이다. 본 실시예에서는 시간적 계층에 대해 단일 신경망 모델을 사용하여 시간적 계층에 따른 스케일링 및 리스케일링을 수행할 수 있다. 이때, 시간적 계층에 따라 잠재표현을 다르게 스케일링할 수 있다. 시간적 계층에 따른 스케일링은 잠재표현에 대한 균등 양자화를 수행하기 전에 잠재표현을 시간적 계층에 따라 다르게 결정되는 스케일 파라메터로 나누어 스케일링하고, 엔트로피 부호화 및 복호화 이후에 스케일된 잠재표현을 다시 앞서 결정된 스케일 파라메터로 곱해주는 것으로 리스케일링 할 수 있다.

[0042] 도 1a를 참조하면, '1' 위치들 중의 전체 또는 일부 위치에서 시간적 계층에 따른 평활화를 수행할 수 있다. 도 1a에는 시간적 계층에 따른 스케일링과 평활화를 모두 수행하는 것으로 표시되어 있으나, 설명의 편의를 위한 것일 뿐 이에 한정되지 않으며 스케일링 과정과 평활화 과정이 선택적으로 수행될 수 있다. 평활화가 수행되는 위치에 따라 다양한 평활화 대상에 대해 평활화를 수행할 수 있다.

[0043] 예를 들어, 양방향 움직임 벡터 추정 단계(111)와, 양방향 움직임 벡터에 대한 잠재 표현 변환 단계(112) 사이에서 평활화를 수행할 경우 양방향 움직임 벡터(m_f, m_b)가 평활화 대상이 된다. 이에 한정되지 않으며, 도 1a에 표시된 평활화 위치 이외에도 움직임 벡터 잠재표현 기반 양방향 시간적 문맥 생성 단계(115)에서 사용되는 여러 신경망들 사이에서도 시간적 계층에 따른 평활화를 수행할 수 있다.

[0044] 일반적으로 양방향 시간적 문맥을 생성하기 위해 이용되는 신경망은 학습할 입력 영상에 대해 평균적으로 부호화 성능을 향상시킬 수 있는 양방향 시간적 문맥을 생성하도록 학습된다. 하지만 이렇게 학습된 양방향 시간적 문맥 생성을 위한 신경망들은 움직임 벡터가 부정확하게 추정된 경우 또는 복잡한 움직임이 존재하는 영상에 대해서는 오히려 개별 입력 영상의 부호화 성능에 악영향을 줄 가능성이 있다. 특히 양방향 예측 구조(Hierarchical B-frame coding)에서는 먼 거리의 참조 영상을 이용하게 되는 경우가 단방향 예측 구조에 비해 많이 발생하므로, 결과적으로 시간적 계층 별로 최적화되지 못한 양방향 시간적 문맥을 사용함에 따른 심각한 부호화 성능 하락을 야기할 수 있다.

[0045] 이에 반해, 본 실시예의 시간적 계층에 따른 평활화는 시간적 계층별로 최적화되지 못하여 발생한 부정확한 시간적 문맥에 대한 규제 방법으로 시간적 문맥 또는 시간적 문맥을 생성하기 위한 정보들에 대해 시간적 계층에 따라 달리 평활화(smoothing)를 적용하여 시간적 계층에 따라 달리 평활화 된 시간적 문맥을 생성할 수 있다. 상위 시간적 계층일수록 높은 평활화 수준을 사용하고 하위 시간적 계층일수록 낮은 평활화 수준을 사용할 수 있다. 이를 통해, 부정확한 시간적 문맥에 대한 규제가 가능해져 일반적인 기술 대비 높은 부호화 성능을 달성할 수 있다.

[0046] 도 1b는 일 실시예에 따른 신경망 기반의 비디오 복호화 방법의 흐름도이다.

[0047] 도 1b를 참조하여 신경망 기반의 비디오 복호화 장치에 의해 수행되는 양방향 예측 및/또는 시간적 계층 정보를 활용한 비디오 복호화 방법의 일 실시예를 설명한다. 비디오 복호화 방법의 단계들에 대한 자세할 설명은 명세서 전체적으로 비디오 부호화 방법에 대해 설명되는 대응하는 단계를 참조할 수 있다.

[0048] 먼저, 비디오 복호화 장치는 양방향 움직임 벡터에 대한 비트스트림(b_m)을 입력으로 양방향 움직임 벡터의 잠재표현($\hat{g}_t$)을 복원할 수 있다(121). 양방향 움직임 벡터에 대한 비트스트림(b_m)은 전술한 바와 같이 비디오 부호화 장치에 의해 양자화 및/또는 엔트로피 부호화를 통해 생성될 수 있다. 움직임 벡터의 비트스트림(b_m)으로부터 엔트로피 복호화를 통해 움직임 벡터의 잠재표현을 복원할 수 있고, 이때, 엔트로피 부호화시 사용한 평

균과 표준편차를 이용해 엔트로피 복호화를 수행할 수 있다. 엔트로피 복호화는 산술 복호화일 수 있다.

[0049] 그 다음, 양방향 움직임 벡터의 잠재표현($\hat{g}_t$)을 이용하여 양방향 움직임 벡터($\hat{m}_f, \hat{m}_b$)를 복원할 수 있다(122).

[0050] 그 다음, 양방향 움직임 벡터의 잠재표현($\hat{g}_t$)을 이용하여 시간적 문맥을 생성할 수 있다(123). 이때, 양방향 참조 영상($\hat{x}_f, \hat{x}_b$)을 활용하여 양방향 시간적 문맥(C_t)을 생성할 수 있다.

[0051] 그 다음. 입력 영상의 비트스트림(b_x)을 입력으로 입력 영상의 잠재표현($\hat{y}_t$)을 복원할 수 있다(124). 이때, 입력 영상의 비트스트림(b_x)은 비디오 부호화 과정에서 양자화 및/또는 엔트로피 부호화를 통해 생성될 수 있다. 입력 영상의 비트스트림(b_x)으로부터 엔트로피 복호화를 통해 입력 영상의 잠재표현을 복원할 수 있고, 이때, 입력 영상의 비트스트림(b_x)을 생성하기 위해 엔트로피 부호화를 수행할 때 사용한 평균과 표준편차를 이용해 엔트로피 복호화를 수행할 수 있다. 엔트로피 복호화는 산술 복호화일 수 있다.

[0052] 그 다음, 입력 영상의 잠재 표현과 시간적 문맥(C_t)를 이용하여 양방향 시간적 문맥 기반 입력 영상($\hat{x}_t$)을 복원할 수 있다(125). 복원된 영상($\hat{x}_t$)은 이후의 입력 영상에 대한 참조 영상으로 활용될 수 있다.

[0053] 본 실시예에서 양방향 예측을 활용한 비디오 복호화 방법은 시간적 계층을 더 이용할 수 있다. 시간적 계층을 이용하는 방법은 시간적 계층에 따른 스케일링과 리스케일링하는 방법과 시간적 계층에 따른 평활화 방법을 포함하며, 비디오 복호화 과정에서 모두 다 수행되거나 선택적으로 어느 하나만 수행될 수 있다.

[0054] 도 1b를 참조하면, 비디오 부호화시 스케일링이 수행된 경우(도 1a의 '2' 위치), '3' 위치에서 양방향 움직임 벡터의 잠재표현 및/또는 입력 영상의 잠재표현에 대해 리스케일링을 수행할 수 있다. 엔트로피 복호화 이후에 스케일된 잠재표현을 비디오 부호화 과정에서 사용한 스케일 파라메터(양자화 구간 결정 파라메터)로 곱해주는 방식으로 리스케일링 할 수 있다.

[0055] 시간적 계층에 따른 평활화는 참조 영상($\hat{x}_f, \hat{x}_b$), 움직임 벡터, 시간적 문맥 생성 단계(123)의 입력, 상기 시간적 문맥 생성 단계(123)의 출력, 및 상기 시간적 문맥 생성 단계(123)의 세부 과정의 입력 또는 출력 중의 적어도 하나를 평활화 대상으로 하여 평활화를 수행할 수 있다. 예를 들어, 도 1b를 참조하면, '1' 위치들 중의 전체 또는 일부 위치에서 시간적 계층에 따른 평활화를 수행할 수 있다. 평활화가 수행되는 위치에 따라 다양한 평활화 대상에 대해 평활화를 수행할 수 있다.

[0056] 도 2a는 다른 실시예에 따라 신경망 기반 비디오 부호화 장치에 의해 수행되는 신경망 기반의 비디오 부호화 방법의 흐름도로서, 평활화가 움직임 벡터 잠재표현 기반 양방향 시간적 문맥을 생성하는 단계(150)의 출력 위치에서 수행되는 경우를 일 예로 하여 비디오 부호화 방법의 흐름도를 나타낸 것이다. 평활화 위치를 움직임 벡터 잠재표현 기반 양방향 시간적 문맥을 생성하는 단계(150)의 출력 위치에 한정하여 설명하나 이에 제한되는 것은 아니며 다른 평활화 위치에 대해서도 쉽게 적용할 수 있다. 즉, 평활화는 참조 영상, 상기 입력 영상, 상기 움직임 벡터, 상기 시간적 문맥 생성 단계의 입력, 상기 시간적 문맥 생성 단계의 출력, 및 상기 시간적 문맥 생성 단계의 세부 과정의 입력 또는 출력에 대해 수행될 수 있다.

[0057] 이하 비디오 부호화(video encoding)를 중심으로 설명하나 복호화(decoding)에도 동일하게 적용될 수 있다.

[0058] 먼저, 비디오 부호화 장치는 양방향 참조 영상($\hat{x}_f, \hat{x}_b$) 및 입력 영상의 시간적 계층(l_t)을 이용하여 입력 영상(x_t)에 대한 양방향 시간적 문맥(C_t) 및/또는 양방향 움직임 벡터에 대한 비트스트림(b_m)을 출력할 수 있다(200).

[0059] 입력 영상(x_t)은 비디오를 구성하는 연속된 영상 중에서 현재 시점(t)에서 부호화할 영상을 의미한다. 참조 영상은 입력 영상(x_t)을 부호화 또는 복호화 하기 위해 참조하는 영상으로 이미 복호화된 영상을 의미한다. 양방

향 참조 영상은 입력 영상(x_t)의 시점(t) 대비 선/후 영상으로서 각각 순방향 참조 영상($\hat{x}_f$)과 역방향 참조 영상($\hat{x}_b$)이라 한다. 순방향 참조 영상과 역방향 참조 영상은 하나 이상일 수 있으나 이하 설명의 편의를 위해 각각 하나인 경우로 한정하여 설명한다. 양방향 시간적 문맥(C_t)은 단일 해상도를 갖는 특징맵 또는 둘 이상의 해상도를 갖는 특징맵일 수 있다. 시간적 계층(l_t)은 하위 계층의 영상을 복호화하지 않고도 상위 계층의 영상을 부호화할 수 있도록 비디오 내의 영상을 시간적으로 계층화할 때의 각 계층을 의미한다.

[0060] 그 다음, 양방향 시간적 문맥(C_t)과 입력 영상의 시간적 계층(l_t)을 이용하여 시간적 계층(l_t)에 따라 양방향 시간적 문맥(C_t)을 평활화하고, 평활화된 양방향 시간적 문맥($\tilde{C}_t$)을 출력할 수 있다(400). 이때, 시간적 계층별로 최적화되지 못하여 발생한 부정확한 시간적 문맥에 대한 규제를 위해 양방향 시간적 문맥을 시간적 계층 및/또는 공간적 위치에 따라 평활화 수준을 다르게 적용할 수 있다.

[0061] 그 다음, 평활화된 양방향 시간적 문맥($\tilde{C}_t$)과 입력 영상(x_t)을 입력으로 하여 입력 영상에 대한 부호화를 수행하여 비트스트림(b_x)을 출력하고, 필요에 따라 평활화된 양방향 시간적 문맥($\tilde{C}_t$) 및 비트스트림(b_x)을 기반으로 복호화를 수행하여 복원 영상($\hat{x}_t$)을 출력할 수 있다(600). 복원 영상($\hat{x}_t$)은 이후의 입력 영상을 부호화하기 위한 참조 영상으로 사용될 수 있다.

[0062] 도 2b는 일 실시예에 따른 양방향 시간적 문맥 생성 단계(200)의 흐름도이다. 도 3a 및 도 3b는 시간적 계층에 따른 스케일링 및 리스케일링을 설명하기 위한 도면이다. 도 2b, 도 3a 및 도 3b를 참조하여 양방향 시간적 문맥을 생성하는 단계(200)의 일 실시예를 설명한다. 전술한 바와 같이 스케일링 및 리스케일링은 선택적으로 생략이 가능하다.

[0063] 먼저, 비디오 부호화 장치는 입력 영상(x_t)과 양방향 참조 영상($\hat{x}_f, \hat{x}_b$)을 이용하여, 각 참조 영상($\hat{x}_f, \hat{x}_b$)과 입력 영상(x_t) 사이의 양방향 움직임 정보를 추정할 수 있다(210). 입력 영상(x_t)과 양방향 참조 영상($\hat{x}_f, \hat{x}_b$)는 신경망에 입력되고, 움직임 벡터(m_f, m_b) 및/또는 픽셀 단위의 광학 흐름(optical flow)을 움직임 정보로서 출력할 수 있다.

[0064] 그 다음, 양방향 움직임 벡터(m_f, m_b)를 양방향 움직임 벡터에 대한 잠재표현(g_t)으로 변환할 수 있다(220). 양방향 움직임 벡터(m_f, m_b)는 합성곱 신경망(Convolutional Neural Network)에 입력되어 잠재표현(g_t)으로 변환될 수 있다. 여기서, 잠재표현(latent representation)은 입력 영상이나 움직임 정보를 입력으로 하는 신경망의 출력으로서 잠재특징(latent feature), 잠재벡터(latent vector) 등을 통칭한다.

[0065] 그 다음, 시간적 계층(l_t)에 따른 양방향 움직임 벡터(m_f, m_b)에 대한 잠재표현(g_t) 및 입력 영상의 시간적 계층(l_t)을 입력으로, 시간적 계층(l_t)에 따른 잠재표현 스케일링을 수행하고, 스케일링된 양방향 움직임 벡터의 잠재표현(g_t^s)을 출력할 수 있다(230).

[0066] 양방향 움직임 벡터(m_f, m_b) 각각에 대하여 동일한 단일 신경망 모델을 이용할 수 있으며, 시간적 계층에 따라 서로 다르게 스케일링할 수 있다. 본 실시예에서 잠재표현 스케일링은 단일 신경망을 사용하므로 시간적 계층에 따라 여러 신경망 모델을 사용하는 일반적인 방법에 비해 메모리 효율 관점에서 더 효율적이다.

[0067] 예를 들어, 양방향 움직임 벡터에 대한 잠재표현(g_t)을 시간적 계층에 따라 미리 정의된 스케일 파라메터로 나누는 방식으로 스케일링할 수 있다. 시간적 계층에 따른 잠재표현 스케일링은 양자화 구간(quantization step)을 조절하는 것과 동일하거나 상응하는 효과를 갖기 때문에 이하에서 용어 "스케일 파라메터"를 "양자화 구간 결정 파라메터"라고 지칭한다. 양자화 구간 결정 파라메터는 시간적 계층에 따라 최적의 양자화 구간을 결정하

도록 미리 학습되어 결정될 수 있다. 입력 영상과 참조 영상 간의 거리가 먼 상위 시간적 계층일수록, 즉 시간적 계층 값이 감소할수록 참조 영상으로 사용되는 경우가 많기 때문에 하위 시간적 계층의 영상에 비해 높은 화질 수준으로 복원되도록 할 수 있다. 이를 위해 양자화 구간 결정 파라메터는 상위 시간적 계층일수록 더 작은 양자화 구간을 사용하도록 학습되고, 하위 시간적 계층일수록 더 큰 양자화 구간을 사용하도록 학습될 수 있다.

[0068] 입력 영상 또는 양방향 움직임 벡터에 대한 잠재표현 $y \in \mathbb{R}^{C \times H \times W}$ (여기서, R: 실수, C: 특징맵으로 표현된 잠재표현의 채널(channel) 수, H, W: 잠재표현의 공간적 높이와 너비)이 있을 때, 양자화 구간 결정 파라메터 세트($\boldsymbol{Q}_{temp}$)와 입력 영상의 시간적 계층 (l_t)에 대한 양자화 구간 결정 파라메터($\boldsymbol{q}_{l_t}$)는 수학식 1과 같이 정의할 수 있다. 여기서, L은 전체 시간적 계층의 수를 나타낸다.

수학식 1

[0069] $$\boldsymbol{Q}_{temp} \in \mathbb{R}^{L \times C}, \qquad \boldsymbol{q}_{l_t} = \boldsymbol{Q}_{temp}[l_t] \in \mathbb{R}^{C}, \qquad l_t = 0, 1, \ldots, L$$

[0071] 도 3a는 GOP(Group Of Pictures) 크기가 12인 양방향 예측 구조에서 네 개의 시간적 계층(l_t=0,1,2,3)에 따라 서로 다른 양자화 구간 결정 파라메터를 사용하는 예시를 나타낸다. 설명의 편의를 위해 최상위 시간적 계층(l_t=0)에는 정지 영상 부호화 방법을 적용할 영상, 즉 I-프레임만 포함되는 것을 가정하나 이에 제한되는 것은 아니다. 도 3a에서 I는 I-프레임, B는 B-프레임을 나타낸다.

[0072] 다시 도 2b를 참조하면, 그 다음 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}_t^s$)을 입력으로 이에 대한 양자화를 수행하며, 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}_t^s$)을 출력할 수 있다(240). 이때, 양자화는 균등 양자화, 즉 반올림 연산을 사용할 수 있으나 이에 한정되지 않는다.

[0073] 그 다음, 양자화 및 스케일링된 움직임 벡터의 잠재표현($\hat{g}_t^s$)을 입력으로, 잠재표현($\hat{g}_t^s$)의 엔트로피 부호화를 위해 잠재표현($\hat{g}_t^s$)의 확률 분포를 추정할 수 있다(250). 엔트로피 모델을 이용할 수 있으며, 엔트로피 모델은 잠재표현이 라플라시안 분포(Laplacian distribution) 또는 가우시안 분포(Gaussian distribution) 등의 분포를 따른다는 가정하에 분포의 평균(μ_t^g)과 표준편차(σ_t^g)를 출력할 수 있다.

[0074] 그 다음, 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}_t^s$)에 대하여 엔트로피 부호화 및 복호화를 수행할 수 있다(260). 양자화 및 스케일링된 움직임 벡터의 잠재표현($\hat{g}_t^s$)과 단계(250)에서 추정된 평균(μ_t^g)과 표준편차(σ_t^g)를 입력으로 잠재표현($\hat{g}_t^s$)에 대한 엔트로피 부호화를 진행하여 비트스트림(b_m)을 출력할 수 있다. 또한, 엔트로피 부호화된 비트스트림(b_m)을 입력으로 엔트로피 부호화시 사용한 평균(μ_t^g)과 표준편차(σ_t^g)를 이용해 엔트로피 복호화를 진행하여 잠재표현($\hat{g}_t^s$)을 복원할 수 있다. 엔트로피 부호화 과정은 일반적으로 무손실 부호화이므로 복원된 잠재표현($\hat{g}_t^s$)은 엔트로피 부호화기의 입력($\hat{g}_t^s$)과 동일하다. 엔트로피 부호화 및 복호화는 일반적인 산술 부호화(arithmetic encoding) 및 산술 복호화(arithmetic decoding) 등을 이용할 수 있으나 이에 제한되지 않는다.

[0075] 그 다음, 입력 영상의 시간적 계층(l_t)과, 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}_t^s$)을 입력으로, 잠재표현을 기존의 스케일로 되돌리기 위한 리스케일링(rescaling)을 수행하여 리스케일링된 양방향 움직임

벡터의 잠재표현($\hat{g}_t$)을 출력할 수 있다(270).

[0076] 도 3b는 l_t 번째 시간적 계층에 대한 잠재표현 스케일링 및 리스케일링을 설명하기 위한 도면이다.

[0077] 수학식 1 및 도 3b를 참조하면, 잠재표현(y_t)을 구성하는 C개의 채널맵을 양자화 구간 결정 파라메터($\boldsymbol{q}_{l_t}$)의 C 개의 파라메터로 나누어 스케일링할 수 있다. 이는 수학식 2와 같이 표현될 수 있다.

수학식 2

[0078] $$y_t^s[c][h][w] = y_t^s[c][h][w] \div \boldsymbol{q}_{l_t}[c]$$

[0079] $$c = 0, 1, \ldots, C-1, h = 0, 1, \ldots, H-1, w = 0, 1, \ldots, W-1$$

[0080] 여기서, C는 특징맵으로 표현된 잠재표현의 채널 수, H, W는 각각 잠재표현의 공간적 높이와 너비를 나타낸다.

[0081] 스케일링된 잠재표현(y_t^s)에 대하여 양자화, 엔트로피 부호화 및 복호화를 수행하고, 스케일링에서와 동일한 양자화 구간 결정 파라메터를 양자화 및 스케일링된 잠재표현($\hat{y}_t^s$)에 곱하여 리스케일링할 수 있다. 결과적으로 양자화 및 리스케일링된 잠재표현($\hat{y}_t$)이 출력된다. 리스케일링은 수식 3과 같이 표현할 수 있다.

수학식 3

[0082] $$\hat{y}_t[c][h][w] = \hat{y}_t^s[c][h][w] \times \boldsymbol{q}_{l_t}[c]$$

[0083] $$c = 0, 1, \ldots, C-1,\ h = 0, 1, \ldots, H-1,\ w = 0, 1, \ldots, W-1$$

[0084] 동일한 신경망 모델로 서로 다른 시간적 계층의 영상에 대해 양자화 구간을 다르게 사용할 수 있으므로 시간적 계층별로 서로 다른 화질 수준으로 부호화하는 효과를 가질 수 있다. 또한, 적은 수의 신경망 파라메터만을 메모리에 저장하여 사용하므로 메모리 효율을 향상시킬 수 있다.

[0085] 다시 도 2b를 참조하면, 그 다음 리스케일링된 양방향 벡터의 잠재표현($\hat{g}_t$)을 입력으로 양방향 움직임 벡터들($\hat{m}_f, \hat{m}_b$)을 복원하여 출력할 수 있다(280). 이때, 합성곱 신경망을 이용할 수 있으나 이에 한정되지 않는다.

[0086] 그 다음, 복원된 양방향 움직임 벡터($\hat{m}_f, \hat{m}_b$)와 참조 영상($\hat{x}_f, \hat{x}_b$)을 입력으로, 입력 영상을 부호화하기 위한 양방향 시간적 문맥(C_t)을 생성하여 출력할 수 있다(290). 이때, 움직임 보상(motion compensation) 기법과 합성곱 신경망을 이용할 수 있다. 움직임 보상은 일반적인 쌍선형 와핑(bilinear warping)을 사용할 수 있으나 이에 제한되지 않는다.

[0087] 도 4는 일 실시예에 따른 양방향 시간적 문맥 평활화 단계(400)의 흐름도이다. 도 5는 시간적 계층에 따른 양방향 시간적 문맥 평활화를 설명하기 위한 도면이다.

[0088] 도 4는 양방향 시간적 문맥 평활화 단계(400)는 시간적 계층(l_t)에 따른 평활화 수준(δ_t) 결정 단계(410)와, 평활화 필터($\boldsymbol{F}_t$) 생성 단계(420), 평활화 단계(430)를 포함할 수 있다.

[0089] 평활화 필터는 가우시안 필터(Gaussian filter) 또는 가우시안 필터 기반의 쌍방 필터(Bilateral filter)와 같이 사전 정의된 필터이거나, 합성곱 신경망과 같이 학습 가능한 필터일 수 있다. 다만, 이에 제한되지 않는다.

학습 가능한 필터를 평활화 필터로 사용하는 경우, 여러 개의 학습 가능한 필터를 시간적 계층(l_t)에 따라 서로 달리 학습시킨 후 시간적 계층(l_t)에 따라 그 중 하나를 선택하여 사용할 수 있다. 평활화 수준(s_t)은 평활화 필터가 가우시안 필터인 경우 표준편차가 될 수 있으며, 학습 가능한 필터인 경우 시간적 계층에 따라 서로 달리 학습된 필터들의 인덱스가 될 수 있다.

[0090] 평활화 필터로 사전 정의된 평활화 필터를 사용하는 경우를 예시로 하여 양방향 시간적 문맥 평활화 단계(400)를 설명한다.

[0091] 먼저, 비디오 부호화 장치는 입력 영상의 시간적 계층(l_t)을 입력으로, 양방향 시간적 문맥(C_t)을 평활화 하기 위한 평활화 수준(smoothing level) (s_t)을 결정하여 출력할 수 있다(410).

[0092] 예를 들어, 가우시안 필터(Gaussian filter)를 이용할 수 있으며, 가우시안 필터를 생성하기 위한 표준편차를 평활화 수준으로 출력할 수 있다. 이때 표준편차는 시간적 계층에 따라 정해진 상수 값을 시간적 계층에 따라 정해두고 그대로 사용할 수도 있고, 시간적 계층에 따라 서로 다른 파라메터를 학습하여 사용할 수도 있다. 상위 시간적 계층일수록 높은 평활화 수준(큰 표준편차)를 출력하고, 하위 시간적 계층일수록 낮은 평활화 수준(작은 표준편차)을 출력할 수 있다. 또한, 동일한 시간적 계층에서도 공간적 위치에 따라 양방향 시간적 문맥에 대한 평활화 수준을 서로 다르게 결정할 수도 있다. 이때, 평활화를 위한 공간적 위치에 대한 정보는 단계(260)에서 복호화시 출력된 정보를 활용할 수 있다. 단계(260)에서 복호화할 때 공간적 위치에 대한 평활화 정보를 함께 출력할 수 있다.

[0093] 그 다음, 단계(410)에서 결정된 평활화 수준(s_t)을 입력으로 사전 정의된 필터의 파라메터를 결정함으로써 실제 평활화에 적용될 평활화 필터(F_t)를 생성할 수 있다(420). 사전 정의된 필터는 예컨대, 가우시안 필터 또는 가우시안 커널을 이용한 쌍방 필터(Bilateral filter) 등을 포함할 수 있으나 이에 제한되지 않는다.

[0094] 그 다음, 단계(420)에서 생성된 평활화 필터를 입력으로 양방향 시간적 문맥에 대한 평활화를 진행한 후 평활화된 양방향 시간적 문맥($\tilde{C}_t$)을 출력할 수 있다(430).

[0095] 도 5는 시간적 계층에 따라 서로 다른 평활화 수준을 이용하여 평활화된 양방향 시간적 문맥의 예시를 나타낸 것이다. 도 5의 ①은 시간적 계층(l_t)이 증가할수록, 즉 하위 시간적 계층으로 갈수록 더 낮은 평활화 수준이 적용된 결과를 보여준다. 또한, 도 5의 ②는 동일한 시간적 계층에서도 공간적 위치에 따라 다른 평활화 수준을 사용한 결과를 보여준다. 여기서, C_t와 $\tilde{C}_t$ 는 시각화의 편의를 위해 모든 채널에 대한 평균으로 계산된 것을 나타낸 것이다.

[0096] 일반적으로 입력 영상과 참조 영상 간의 거리가 먼 상위 시간적 계층일수록 부정확한 시간적 문맥이 생성될 확률이 높다. 본 실시예에서 시간적 문맥을 시간적 계층에 따라 서로 다르게 평활화함으로써 부정확한 시간적 문맥에 대한 규제가 가능해져 일반 기술 대비 높은 부호화 성능을 달성할 수 있다.

[0097] 도 6a 및 도 6b는 일 실시예에 따른 입력 영상의 부호화 및/또는 복호화 단계(600)의 흐름도이다.

[0098] 도 6a를 참조하면, 비디오 부호화 장치는 입력 영상(x_t)과 평활화된 양방향 시간적 문맥($\tilde{C}_t$)을 입력으로, 입력 영상(x_t)을 잠재표현(y_t)으로 변환하여 출력할 수 있다(610). 합성곱 신경망을 이용할 수 있으며 단방향 참조 영상으로부터 생성한 시간적 문맥을 사용하는 대신 평활화된 양방향 시간적 문맥을 사용할 수 있다.

[0099] 그 다음, 입력 영상의 시간적 계층(l_t)과 입력 영상의 잠재표현(y_t)을 입력으로 입력 영상의 잠재표현(y_t)에 대해 시간적 계층에 따른 잠재표현 스케일링을 수행하고, 스케일링된 입력 영상의 잠재표현(y_t^s)을 출력할 수 있다(620). 단계(230)에서 설명한 바와 같이 입력 영상의 잠재표현(y_t)을 시간적 계층에 따라 정의된 양자화 구간 결정 파라메터로 나누어 스케일링할 수 있다.

[0100] 그 다음, 스케일링된 입력 영상의 잠재표현(y_t^s)의 양자화를 수행하여 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}_t^s$)을 출력할 수 있다(630). 균등 양자화, 즉 통상적인 반올림 연산을 사용할 수 있으나 이에 제한되지 않는다.

[0101] 그 다음, 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}_t^s$)을 입력으로, 잠재표현($\hat{y}_t^s$)의 엔트로피 부호화를 위해 잠재표현($\hat{y}_t^s$)의 확률 분포를 추정할 수 있다(640). 이때, 엔트로피 모델을 이용할 수 있고, 엔트로피 모델은 잠재표현이 라플라시안 분포(Laplacian distribution) 또는 가우시안 분포(Gaussian distribution) 등의 분포를 따른다는 가정하에 분포의 평균(μ_t^y)과 표준편차(σ_t^y)를 출력할 수 있다.

[0102] 그 다음, 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}_t^s$)과, 평균(μ_t^y)과 표준편차(σ_t^y)를 입력으로 잠재표현($\hat{y}_t^s$)에 대한 엔트로피 부호화를 수행하여 비트스트림(b_x)을 출력할 수 있다(650). 엔트로피 부호화는 산술 부호화일 수 있다.

[0103] 도 6b를 참조하면, 엔트로피 부호화(650) 결과를 기초로 복호화를 수행하는 과정이 수행될 수 있다.

[0104] 엔트로피 부호화된 비트스트림(b_x)과, 엔트로피 부호화시 사용한 평균(μ_t^y)과 표준편차(σ_t^y)를 입력으로 엔트로피 복호화를 수행하여 잠재표현($\hat{y}_t^s$)을 복원할 수 있다(660). 엔트로피 부호화 과정(650)은 일반적으로 무손실 부호화이므로 단계(660)에서 복원된 잠재표현($\hat{y}_t^s$)은 엔트로피 부호화의 입력인 잠재표현($\hat{y}_t^s$)과 동일할 수 있다. 엔트로피 복호화를 위해 산술 복호화일 수 있으나 이에 제한되지 않는다.

[0105] 그 다음, 입력 영상의 시간적 계층(l_t)과 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}_t^s$)을 입력으로, 스케일링된 입력 영상의 잠재표현을 기존의 스케일로 되돌리기 위한 리스케일링을 수행하여, 리스케일링된 입력 영상의 잠재표현($\hat{y}_t$)을 출력할 수 있다(670). 리스케일링은 전술한 단계(270)에서 설명한 바와 동일한 방식으로 수행할 수 있다.

[0106] 마지막으로, 리스케일링된 입력 영상의 잠재표현($\hat{y}_t$)과 평활화된 양방향 시간적 문맥($\tilde{C}_t$)을 입력으로, 입력 영상($\hat{x}_t$)을 복원하여 출력할 수 있다(680). 이때, 합성곱 신경망을 이용할 수 있으나 단방향 참조 영상으로부터 생성한 시간적 문맥을 사용하는 대신 평활화된 양방향 시간적 문맥을 사용할 수 있다. 복원된 영상의 이후의 입력 영상에 대한 참조 영상으로 활용될 수 있다.

[0107] 도 7은 일 실시예에 따른 전자장치의 블록도이다.

[0108] 전자장치(700)는 전술한 비디오 부호화 방법 및/또는 복호화 방법의 다양한 실시예들을 수행하는 신경망 기반 비디오 부호화 장치 및/또는 복호화 장치를 포함할 수 있다. 전자장치는 예를 들어 에지 기기(edge device)와 같이 컴퓨팅 자원이 제한된 환경에서 비디오 부호화와 같은 응용을 필요로 하는 장치, 휴대폰, TV, 모니터, IoT(Internet of Things) 기기 등 영상 송/수신을 수행하는 다양한 장치, 클라우드 컴퓨팅 장치 등을 포함할 수 있다.

[0109] 도 7을 참조하면, 전자장치(700)는 메모리(710) 및 프로세서(720)를 포함한다.

[0110] 메모리(710)는 비디오 부호화 등의 영상 처리를 위한 다양한 명령어들 및 영상 처리와 관련된 데이터(신경망, 가중치, 시간적 계층에 따른 양자화 구간 결정 파라메터, 참조 영상, 복원 영상, 부호화를 통해 생성된 비트스트림 등)를 저장할 수 있다. 메모리(710)는 DRAM(Dynamic Random Access Memory), SRAM(Static Random Access Memory) 등의 RAM(Random Access Memory), ROM(Read-Only Memory), 플래시 메모리(Flash Memory), 캐시 메모리(Cache Memory), 및 가상 메모리(Virtual Memory) 등을 포함할 수 있다. 다만 이에 제한되지는 않는다.

[0111] 프로세서(720)는 메모리(710)에 저장된 명령어들을 실행하여, 메모리(710)에 저장된 다양한 데이터들을 참조로 다양한 비디오 처리 과정을 수행할 수 있다. 명령어들은 전술한 비디오 부호화 방법의 다양한 실시예들을 수행

하도록 구성된 하나 이상의 명령어들을 포함할 수 있다. 예를 들어, 입력 영상, 양방향 참조 영상 및 상기 입력 영상의 시간적 계층을 입력으로 상기 입력 영상에 대한 양방향 시간적 문맥을 생성하고, 상기 양방향 시간적 문맥과 상기 시간적 계층을 입력으로 상기 양방향 시간적 문맥을 평활화하고, 평활화된 양방향 시간적 문맥과 입력 영상을 입력으로 입력 영상을 부호화 및 복호화 하는 것을 포함할 수 있다.

[0112] 도 8a 내지 도 8i는 일 실시예에 따른 비디오 부호화 및/또는 복호화 과정의 전체 동작을 설명하기 위한 도면이다. 도 9a는 시간적 문맥 융합 신경망의 일 예이다. 도 9b는 확률분포를 추정하는 엔트로피 모델의 일 예이다.

[0113] 도 8a를 참조하여 전자장치(700)에서 수행되는 비디오 부호화의 전반적인 동작을 설명한다. 전자장치(700)는 비디오 부호화 장치(800) 및/또는 비디오 복호화 장치(800a)를 포함할 수 있다. 도시된 바와 같이, 비디오 부호화 장치(800)는 비디오 복호화 장치(800a)는 양방향 예측 및 시간적 계층 정보를 활용하기 위한 구성들을 공유할 수 있다. 비디오 복호화 장치(800a)는 비디오 부호화 장치(800)와는 별도의 하드웨어칩 및/또는 소프트웨어로 분리 구성될 수 있다.

[0114] 비디오 부호화 장치(800)는 입력 영상(x_t)과 양방향 참조 영상($\hat{x}_f$, $\hat{x}_b$)을 단일 움직임 벡터 추정 신경망(Motion Vector Estimation Network)(811)에 입력하고, 순방향 참조 영상($\hat{x}_f$)과 입력 영상(x_t) 간의 움직임 벡터(m_f)와, 역방향 참조 영상($\hat{x}_b$)과 입력 영상(x_t) 간의 움직임 벡터(m_b)를 추정할 수 있다. 이때 움직임 벡터 추정 신경망은 다양한 광학 흐름 추정 신경망을 이용할 수 있다.

[0115] 그 다음, 움직임 벡터 추정 신경망(Motion Vector Estimation Network)(811)에서 출력된 두 움직임 벡터(m_f, m_b)를 하나의 양방향 움직임 벡터의 잠재표현(g_t)으로 변환하기 위해 두 움직임 벡터에 대해 채널 간 연결(concatenation)을 수행하고, 그 결과를 움직임 벡터 부호화 신경망(MV Enc: Motion Vector Encoder)(812)에 입력할 수 있다. 움직임 벡터 부호화 신경망은 합성곱 신경망을 기반으로 할 수 있으며, 양방향 움직임 벡터를 입력으로 하기 때문에 입력 합성곱 층(convolution layer)의 채널 수가 두 배로 구성될 수 있다.

[0116] 그 다음, 입력 영상의 시간적 계층(l_t)에 따라 양자화 구간 파라메터 셋(Q^g_{temp})에서 양자화 구간 파라메터($q^g_{l_t}$)를 선택하고, 움직임 벡터 부호화 신경망(812)에서 출력된 양방향 움직임 벡터의 잠재표현(g_t)을 그 선택된 양자화 구간 파라메터로 나누어(813) 시간적 계층에 따른 양방향 움직임 벡터의 잠재표현 스케일링을 수행할 수 있다.

[0117] 그 다음, 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}^s_t$)을 양자화하고(814), 엔트로피 모델(815)을 이용하여 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}^s_t$)의 확률 분포를 추정할 수 있다. 엔트로피 모델은 예를 들어, 초사전정보(hyperprior)와 자기회귀컨텍스트(autoregressive context)를 함께 활용한 엔트로피 모델 등을 사용할 수 있으나 이에 제한되지 않는다.

[0118] 엔트로피 모델(815)에서 추정된 확률 분포, 예컨대 평균(μ^g_t)과 표준편차(σ^g_t)를 이용하여 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}^s_t$)에 대한 부호화(8161) 및 복호화(8162)가 진행된다. 부호화 결과 양방향 움직임 벡터의 잠재표현에 대한 비트스트림(b_m)이 출력되고, 비트스트림(b_m)과 확률 분포를 이용하여 복호화가 진행된다.

[0119] 그 다음, 입력 영상의 시간적 계층(l_t)을 이용하여 양자화 구간 파라메터 셋(Q^g_{temp})에서 양자화 구간 파라메터($q^g_{l_t}$)를 선택하고, 선택된 양자화 구간 파라메터를 양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현에 곱하여(817) 시간적 계층에 따른 리스케일링을 수행할 수 있다.

[0120] 그 다음, 리스케일링된 양방향 움직임 벡터의 잠재표현($\hat{g}_t$)을 움직임 벡터 복원 신경망(MV Decoder: Motion Vector Decoder)(818)에 입력하고, 순방향 움직임 벡터($\hat{m}_f$)와 역방향 움직임 벡터($\hat{m}_b$)를 복원할 수 있다. 이

때 움직임 벡터 복원 신경망은 신경망은 합성곱 신경망을 기반으로 하며, 양방향 움직임 벡터를 모두 복원해야 하기 때문에 출력 합성곱 층(convolution layer)의 채널 수를 두 배로 구성할 수 있다. 또한, 움직임 벡터 복원 신경망은 공간적 위치에 따라 서로 다른 평활화를 진행하기 위해 채널 수를 증가하여 공간적 위치와 관련된 평활화 정보를 출력하도록 할 수 있다.

[0121] 그 다음, 복원된 양방향 움직임 벡터($\hat{m}_f, \hat{m}_b$)와 양방향 참조 영상($\hat{x}_f, \hat{x}_b$)를 입력으로 양방향 시간적 문맥(C_t)를 출력한다(819).

[0122] 먼저 양방향 참조 영상을 각각 특징 추출 신경망(Feature Extractor)(8191)에 입력하여 순방향 참조 특징맵($\check{x}_f$)과 역방향 참조 특징맵($\check{x}_b$)을 출력할 수 있다. 이때 특징 추출 신경망(8191)은 합성곱 신경망일 수 있다. 이후, 순방향 움직임 벡터($\hat{m}_f$)를 기반으로 순방향 참조 특징맵($\check{x}_f$)에 대해 통상적인 쌍선형 와핑(bilinear warping)(8192)을 수행하여 와핑된 순방향 참조 특징맵($\ddot{x}_f$)을 출력할 수 있다. 마찬가지로 역방향 움직임 벡터($\hat{m}_b$)와 역방향 참조 특징맵($\check{x}_b$)에 대하여 와핑된 역방향 참조 특징맵($\ddot{x}_b$)를 출력할 수 있다. 와핑된 순방향 참조 특징맵과 와핑된 역방향 참조 특징맵은 각각 후처리를 위한 신경망(Context Refinement)(8193)을 거쳐 순방향 시간적 문맥(C_f)과 역방향 시간적 문맥(C_b)으로 출력된다. 이때 후처리를 위한 신경망은 합성곱 신경망일 수 있다. 마지막으로 순방향 시간적 문맥(C_f)과 역방향 시간적 문맥(C_b)을 시간적 문맥 융합 신경망(Context Fusion Network)(8194)에 입력하여 양방향 시간적 문맥(C_t)를 출력할 수 있다.

[0123] 도 9a를 참조하면, 시간적 문맥 융합 신경망(8194)은 채널 간 연결(concatenation)(911)과 합성곱 층(912,914), 및 ReLU(예: LeakyReLU)(913)와 같은 비선형 층으로 구성할 수 있다. 합성곱 층(912,914) Conv(k, C_{in}, C_{out}, s)에서 k는 커널의 가로, 세로 크기, C_{in}은 입력 채널 수, C_{out}은 출력 채널 수, s는 합성곱의 스트라이드(stride)를 각각 나타낸다. 도 9a에서 N은 순방향 시간적 문맥과 양방향 시간적 문맥의 채널 수를 의미한다.

[0124] 다시 도 8a를 참조하면, 입력 영상의 시간적 계층(l_t)을 입력으로 평활화 하는 과정이 수행될 수 있다. 입력 영상의 시간적 계층(l_t)을 입력으로 평활화 필터를 생성하기 위한 평활화 수준(δ_t)을 출력할 수 있다(821).

[0125] 평활화 방법으로 가우시안 평활화(Gaussian smoothing)나 가우시안 커널에 기반한 쌍방 필터(bilateral filter)를 사용할 수 있으며, 이 경우 평활화 필터를 생성하기 위한 표준편차를 출력할 수 있다. 이때 상위 시간적 계층일수록 더 큰 평활화 수준을 사용하게 하기 위해서 예를 들어, 시간적 계층(l_t)에 따라 $s_1 = 10$, $s_2 = 5$, $s_3 = 2.5$ 와 같이 단조 감소하는 평활화 수준을 사용할 수 있다. 가우시안 평활화를 통해 공간적인 위치 차이를 고려할 수 있고, 쌍방 필터를 이용하여 공간적인 위치 차이뿐 아니라 평활화 대상 입력의 값 차이도 함께 고려하도록 할 수 있다.

[0126] 그 다음, 양방향 시간적 문맥 평활화에 사용할 평활화 필터를 생성할 수 있다(822). 가우시안 평활화를 진행한다고 가정할 때, 시간적 계층에 따른 평활화 수준(δ_t), 즉 표준편차를 입력으로 평활화 필터를 생성할 수 있다. 평활화 수준을 δ_t라 하고, 가우시안 커널의 크기를 $\mathrm{K} \times \mathrm{K}$라 하면, K 값은 예를 들어 수학식 4와 같이 계산할 수 있다.

수학식 4

[0127] $$\mathrm{K} = 2 \times \lceil 3 \times s_{l_t} \rceil + 1$$

[0128] 여기서, $\lceil \cdot \rceil$는 일반적인 올림 연산을 의미한다.

[0129] K 값이 주어질 때, 주어진 평활화 수준(s_t)에 따른 가우시안 커널(F_t)은 수식 5와 같이 구할 수 있다. 이 수식에서 (k_x, k_y) 는 커널의 공간적 위치를 나타낸다.

수학식 5

[0130]
$$F_t[k_x][k_y] = \frac{\exp\left(-\frac{(k_x)^2 + (k_y)^2}{2(s_{l_t})^2}\right)}{\sum_{\forall k_x}\sum_{\forall k_y}\exp\left(-\frac{(k_x)^2 + (k_y)^2}{2(s_{l_t})^2}\right)}$$

[0131]
$$k_x, k_y \in \left\{-\frac{K-1}{2}, -\frac{K-3}{2}, \ldots, 0, \ldots, \frac{K-3}{2}, \frac{K-1}{2}\right\}$$

[0132] 그 다음, 양방향 시간적 문맥(c_t)과 평활화 필터, 예를 들어 가우시안 커널을 입력으로 하여 평활화된 양방향 시간적 문맥($\tilde{c}_t$)를 출력한다(823). 이때 커널을 스트라이드(stride) 1로 하여 양방향 시간적 문맥(C_t)의 모든 공간적 위치마다 곱해 줄 수 있다.

[0133] 그 다음, 평활화된 양방향 시간적 문맥($\tilde{C}_t$)과 입력 영상(x_t)을 입력으로, 채널 간 연결(concatenation) 및 시간적 문맥 기반 부호화 신경망(Enc: contextual encoder)(831)을 통해 입력 영상의 잠재표현(y_t)을 출력할 수 있다.

[0134] 그 다음, 입력 영상의 시간적 계층(l_t)을 입력으로, 양자화 구간 파라메터 셋(Q^y_{temp})에서 양자화 구간 파라메터($q^y_{l_t}$)를 선택할 수 있다. 입력 영상의 잠재표현(y_t)을 선택된 양자화 구간 파라메터로 나누어(832), 시간적 계층에 따른 양방향 움직임 벡터의 잠재표현 스케일링을 수행할 수 있다.

[0135] 그 다음, 스케일링된 양방향 움직임 벡터의 잠재표현의 양자화를 수행하고(833), 엔트로피 모델(834)를 이용하여 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}^s_t$)의 확률 분포, 즉 확률 분포의 평균(μ^y_t)과 표준편차(σ^y_t)를 추정할 수 있다.

[0136] 도 9b를 참조하면, 엔트로피 모델(834)은 일반적인 초사전정보(hyperprior) 모델(921), 자기회귀컨텍스트(autoregressive context) 모델(923)와, 시간적 사전 정보 추출 신경망(Temporal prior encoder)(924) 및 사전정보 융합(Prior fusion) 신경망(925)으로 구성될 수 있다. 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}^s_t$)은 초사전정보 모델(921)과 자기회귀컨텍스트 모델(923)에 입력되어 각각 초사전정보와 자기회귀컨텍스트가 출력된다. 또한, 시간적 사전 정보 추출 신경망(Temporal prior encoder)(924)은 평활화된 양방향 시간적 문맥($\tilde{c}_t$)을 입력으로 시간적 사전 정보를 추출할 수 있다. 사전정보 융합 신경망(925)은 시간적 사전 정보와, 초사전정보, 자기회귀컨텍스트를 입력으로, 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}^s_t$)에 대한 평균(μ^y_t)과 표준편차(σ^y_t)를 출력할 수 있다.

[0137] 다시 도 8a를 참조하면, 엔트로피 모델(834)에서 추정된 평균(μ^y_t)과 표준편차(σ^y_t)를 이용하여 양자화 및 스케일링된 입력 영상의 잠재표현($\hat{y}^s_t$)에 대한 엔트로피 부호화(8351)가 진행되고 결과로서 입력 영상에 대한 비트스

트림(b_x)이 출력된다.

[0138] 이후, 출력된 비트스트림(b_x)과 평균(μ_t^y)과 표준편차(σ_t^y)를 이용하여 복호화(8352)를 수행하고, 입력 영상의 시간적 계층(l_t)에 따른 양자화 구간 파라메터($q_{l_t}^y$)를 양자화 및 스케일링된 입력 영상의 잠재표현에 곱하여(836) 시간적 계층에 따른 리스케일링을 수행한 다음, 시간적 문맥 기반 복원 신경망(Dec: contextual decoder)(837)에 평활화된 양방향 시간적 문맥($\tilde{C}_t$)와 입력 영상(x_t)를 입력하여 입력 영상에 대한 복원 영상($\hat{x}_t$)을 출력할 수 있다. 복원 영상($\hat{x}_t$)은 메모리에 저장되어 이후 입력 영상의 부호화를 위한 참조 영상으로 활용될 수 있다.

[0139] 상술한 바와 같이 시간적 계층에 따른 평활화가 양방향 시간적 문맥(C_t)에 대해 수행될 수도 있지만(821,822,823), 그 밖의 다양한 평활화 위치에서도 유사하게 수행될 수 있다.

[0140] 도 8a에서 ①, ②, ③, ④, ⑤, ⑥, ⑦, ⑧, ⑨는 다양한 평활화 위치를 나타내고 있다. 각 위치에서 평활화를 수행할 경우의 평활화 대상은 아래 표 1과 같다. 평활화 대상의 개수가 두 개인 경우 비고 란에 표시한 바와 같이 각각에 대해 평활화를 수행할 수도 있고, 두 대상을 채널 간 결합(concatenation)하여 하나로 만든 후 평활화를 수행할 수도 있다.

[0141] 도 8b 내지 도 8i는 평활화 위치가 ② ~ ⑨인 경우의 평활화를 도시한 것이다. 도시된 바와 같이 시간적 계층(l_t)과 평활화 대상이 평활화 수행 모듈(SM)에 입력되어 평활화가 수행된다. 예를 들어, 평활화 위치가 도 8a의 ⑧인 경우, 평활화 대상은 도 8a의 움직임 벡터 추정 신경망(Motion Vector Estimation Network, 811)의 출력인 양방향 움직임 벡터(m_f, m_b)가 되며, 평활화 수행 모듈(SM)을 이용하여 m_f, m_b 각각에 대해 평활화를 수행하여 평활화된 양방향 움직임 벡터 $\tilde{m}_f, \tilde{m}_b$를 획득할 수 있다.

표 1

[0142]

평활화 위치	평활화 대상	비고
도 8a의 ①	8194의 출력 C_t	C_t에 대해 평활화
도 8a의 ②	8193의 출력 C_f, C_b	C_f, C_b 각각에 대해 평활화
도 8a의 ③	8192의 출력 $\ddot{x}_f$, $\ddot{x}_b$	$\ddot{x}_f$, $\ddot{x}_b$ 각각에 대해 평활화
도 8a의 ④	8191의 출력 $\check{x}_f$, $\check{x}_b$	$\check{x}_f$, $\check{x}_b$ 각각에 대해 평활화
도 8a의 ⑤	8191의 입력 $\hat{x}_f$, $\hat{x}_b$	$\hat{x}_f$, $\hat{x}_b$ 각각에 대해 평활화
도 8a의 ⑥	811의 입력 $\hat{x}_f$, $\hat{x}_b$	$\hat{x}_f$, $\hat{x}_b$ 각각에 대해 평활화
도 8a의 ⑦	811의 입력 x_t	x_t에 대해 평활화
도 8a의 ⑧	811의 출력 m_f, m_b	m_f, m_b 각각에 대해 평활화
도 8a의 ⑨	818의 출력 $\hat{m}_f$, $\hat{m}_b$	$\hat{m}_f$, $\hat{m}_b$ 각각에 대해 평활화

[0144] 다시 도 8a를 참조하면, 신경망 기반 비디오 복호화 장치(800a)는 비디오 부호화 장치(800a)에서 생성된 양방향 움직임 벡터의 잠재표현에 대한 비트스트림(b_m), 및 양방향 참조 영상을 이용하여 움직임 벡터 복원 및 양방향 시간적 문맥을 생성할 수 있고, 입력 영상에 대한 비트스트림(b_x)과 양방향 시간적 문맥을 이용하여 입력 영상을 복원할 수 있다. 이때, 시간적 계층을 이용하여 잠재표현의 리스케일링 및/또는 평활화를 수행할 수 있다. 중복되는 구성에 대하여 비디오 부호화 장치에서 자세히 설명하였으므로 이하 생략한다.

[0145] 도 10은 다른 실시예에 따른 전자장치의 블록도이다.

[0146] 도 10을 참조하면, 전자장치(1000)는 메모리(100), 프로세서(1020), 촬영장치(1030), 통신장치(1040), 및 출력

장치(1050)를 포함할 수 있다. 메모리(100) 및 프로세서(1020)에 대해서는 앞에서 상술하였으므로 생략한다.

[0147] 촬영장치(1030)는 정지영상 및 동영상 등을 촬영하는 카메라 등의 장치를 포함하며 촬영된 영상을 메모리(1030)에 저장하고 프로세서(1020)에 전달할 수 있다. 촬영장치(1030)는 하나 이상의 렌즈를 포함하는 렌즈 어셈블리, 이미지센서들, 이미지 시그널 프로세서들, 및/또는 플래시들을 포함할 수 있다. 카메라모듈에 포함된 렌즈 어셈블리는 이미지 촬영의 대상인 피사체로부터 방출되는 빛을 수집할 수 있다.

[0148] 통신장치(1040)는 다양한 통신 기술을 이용하여 비디오 부호화 장치(1000)와 네트워크 환경 내에 있는 다른 전자장치나 서버 또는 센서장치 사이의 직접(유선) 통신 채널 및/또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신장치(1040)은 촬영장치(1030)에 의해 촬영된 영상, 및/또는 프로세서(1020)에 의해 처리된 데이터(예: 비디오 부호화 영상 등)를 다른 전자장치에 전송할 수 있다. 또한, 통신장치(1040)는 클라우드 또는 다른 전자장치로부터 처리할 영상을 수신할 수 있고, 수신된 영상을 메모리(1020)에 저장하고, 프로세서(1020)가 영상을 처리할 수 있도록 프로세서(1020)에 전달할 수 있다.

[0149] 출력장치(1050)는 촬영장치(1030)에 의해 촬영된 영상, 및/또는 프로세서(1020)에 의해 처리된 데이터를 시각적/비시각적인 방식으로 출력할 수 있다. 출력장치(1050)는 음향 출력 장치, 표시 장치(예: 디스플레이), 오디오 모듈 및/또는 햅틱 모듈을 포함할 수 있다. 프로세서(1020)에서 처리된 영상 등을 디스플레이에 표시하여 영상 사용자 경험을 개선할 수 있다.

[0150] 또한, 전자장치(1000)는 다양한 데이터를 감지하는 센서장치(예: 가속도센서, 자이로스코프, 자기장센서, 근접센서, 조도 센서, 지문 센서 등), 사용될 명령 및/또는 데이터를 사용자 등으로부터 수신하는 입력장치(예: 마이크, 마우스, 키보드, 및/또는 디지털 펜(스타일러스 펜 등) 등)등을 더 포함할 수 있다.

[0151] 한편, 본 실시 예들은 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터가 읽을 수 있는 코드로 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다.

[0152] 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현하는 것을 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 실시예들을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술 분야의 프로그래머들에 의하여 용이하게 추론될 수 있다.

[0153] 본 개시가 속하는 기술분야의 통상의 지식을 가진 자는 개시된 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

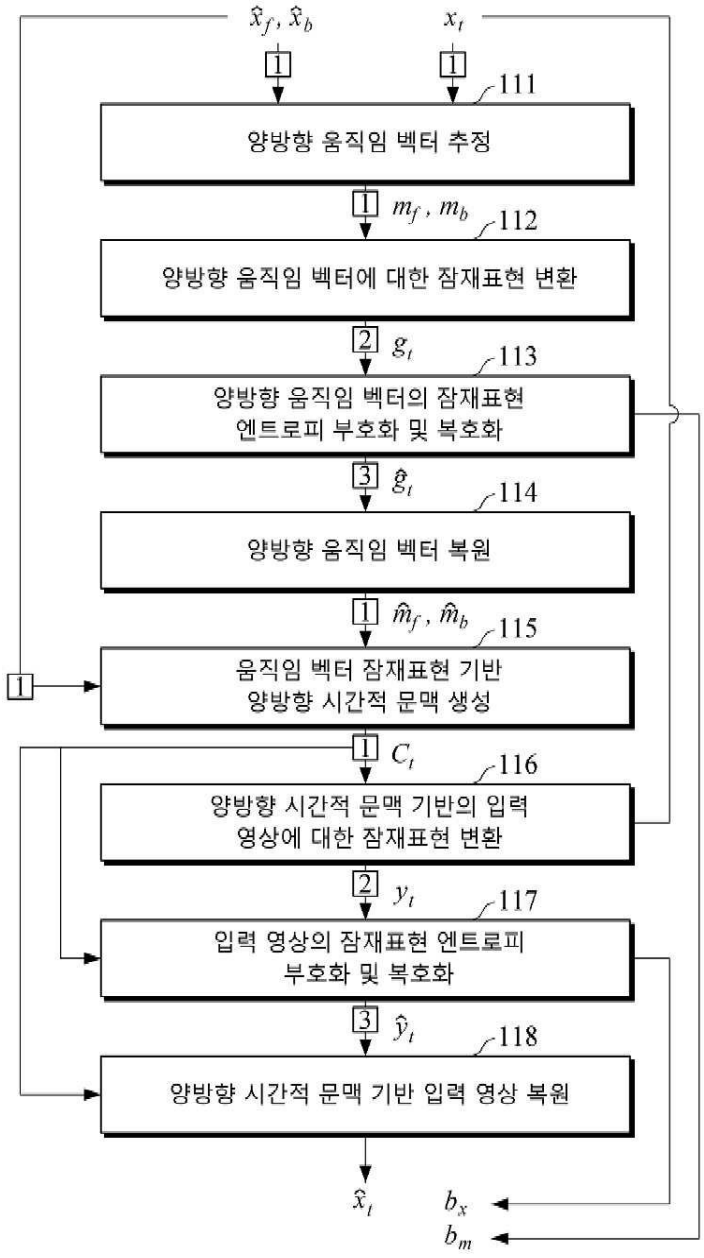
[0154] 700,1000: 전자장치 710,1010: 메모리

720, 1020: 프로세서 1030: 촬영장치

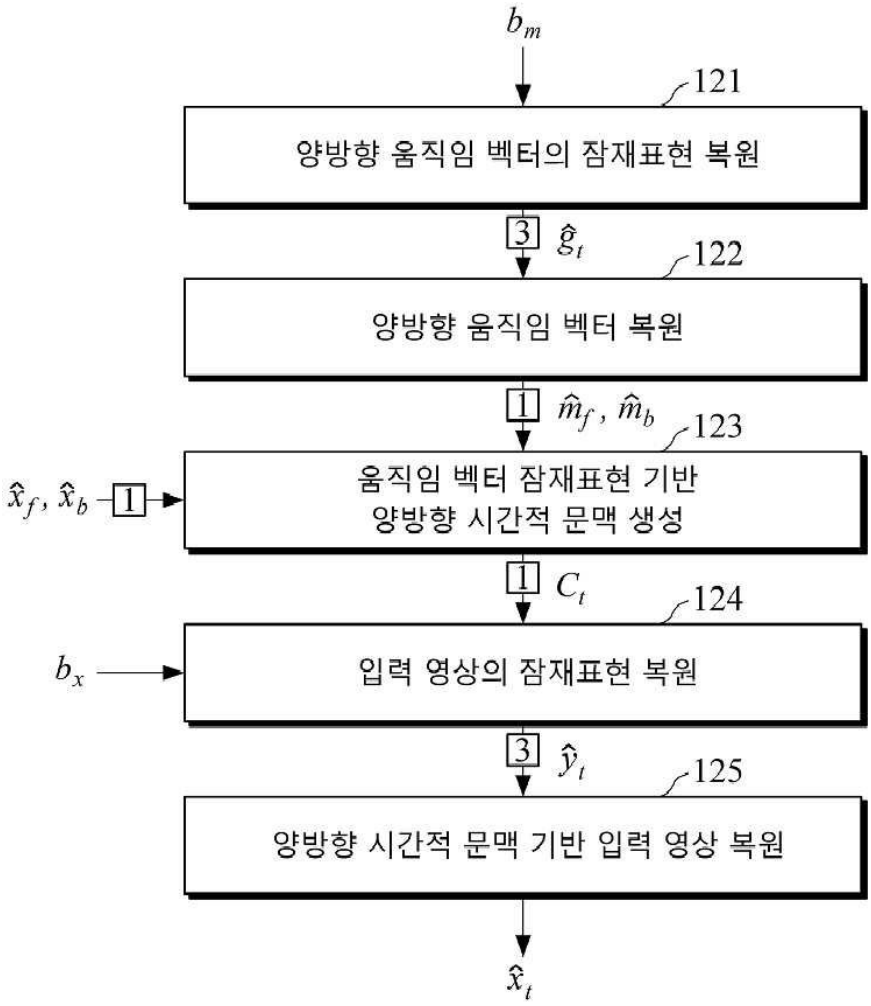
1040: 통신장치 1050: 출력장치

도면

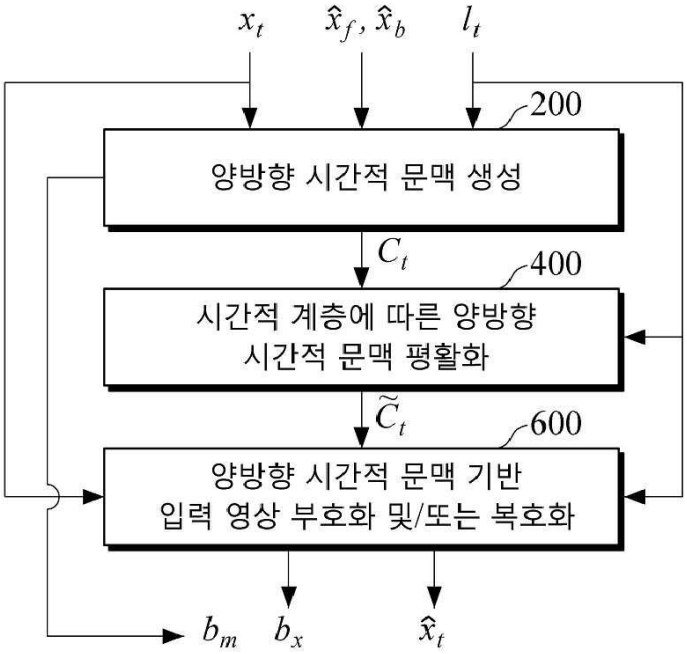
도면1a


$\hat{x}_f, \hat{x}_b$
x_t
111
양방향 움직임 벡터 추정
m_f, m_b
112
양방향 움직임 벡터에 대한 잠재표현 변환
g_t
113
양방향 움직임 벡터의 잠재표현
엔트로피 부호화 및 복호화
$\hat{g}_t$
114
양방향 움직임 벡터 복원
$\hat{m}_f, \hat{m}_b$
115
움직임 벡터 잠재표현 기반
양방향 시간적 문맥 생성
C_t
116
양방향 시간적 문맥 기반의 입력
영상에 대한 잠재표현 변환
y_t
117
입력 영상의 잠재표현 엔트로피
부호화 및 복호화
$\hat{y}_t$
118
양방향 시간적 문맥 기반 입력 영상 복원
$\hat{x}_t$
b_x
b_m

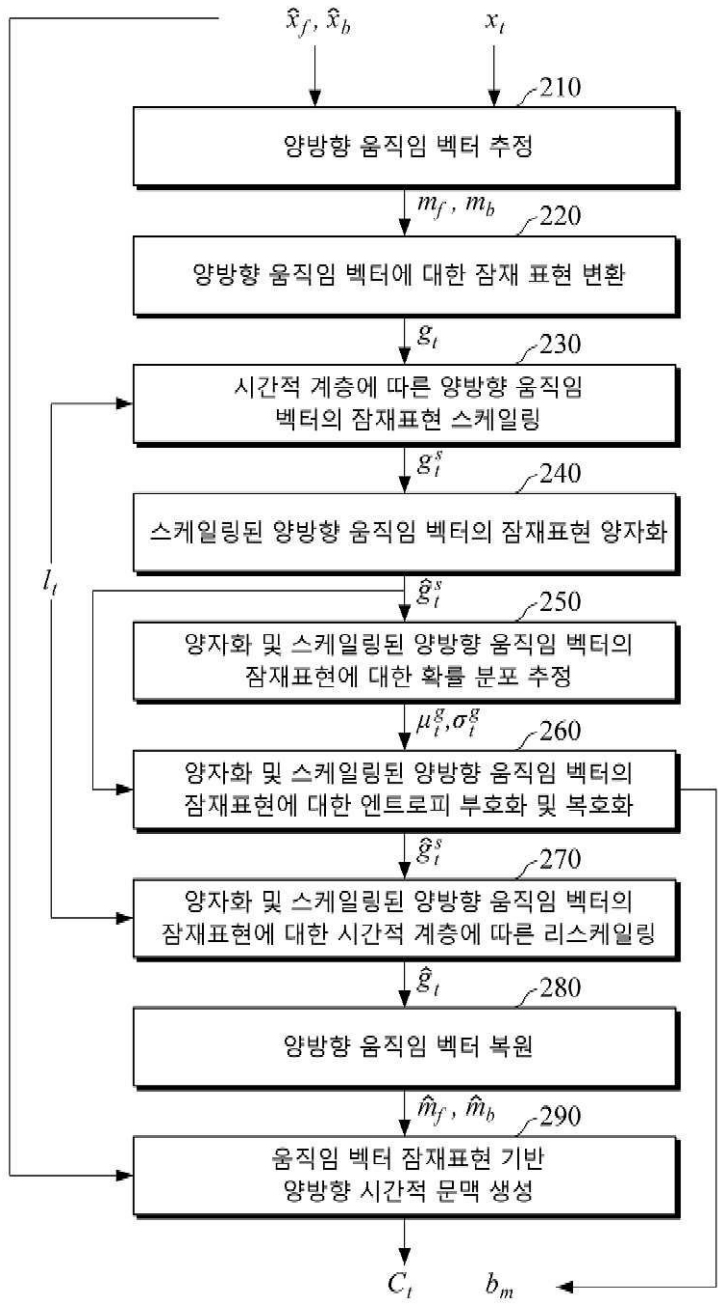
도면1b


b_m
121
양방향 움직임 벡터의 잠재표현 복원
3 $\hat{g}_t$
122
양방향 움직임 벡터 복원
1 $\hat{m}_f$, $\hat{m}_b$
123
움직임 벡터 잠재표현 기반
양방향 시간적 문맥 생성
$\hat{x}_f$, $\hat{x}_b$ 1
1 C_t
124
b_x
입력 영상의 잠재표현 복원
3 $\hat{y}_t$
125
양방향 시간적 문맥 기반 입력 영상 복원
$\hat{x}_t$

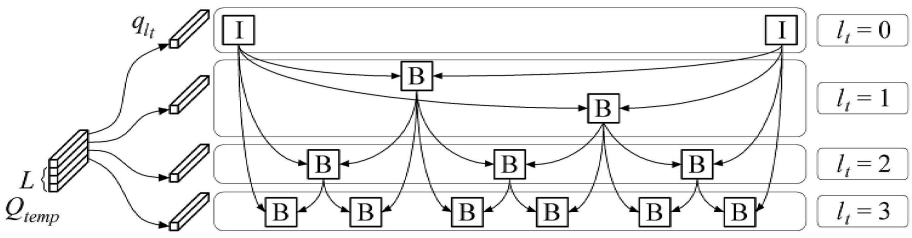
도면2a


x_t
$\hat{x}_f$, $\hat{x}_b$
l_t
200
양방향 시간적 문맥 생성
C_t
400
시간적 계층에 따른 양방향
시간적 문맥 평활화
$\tilde{C}_t$
600
양방향 시간적 문맥 기반
입력 영상 부호화 및/또는 복호화
b_m
b_x
$\hat{x}_t$

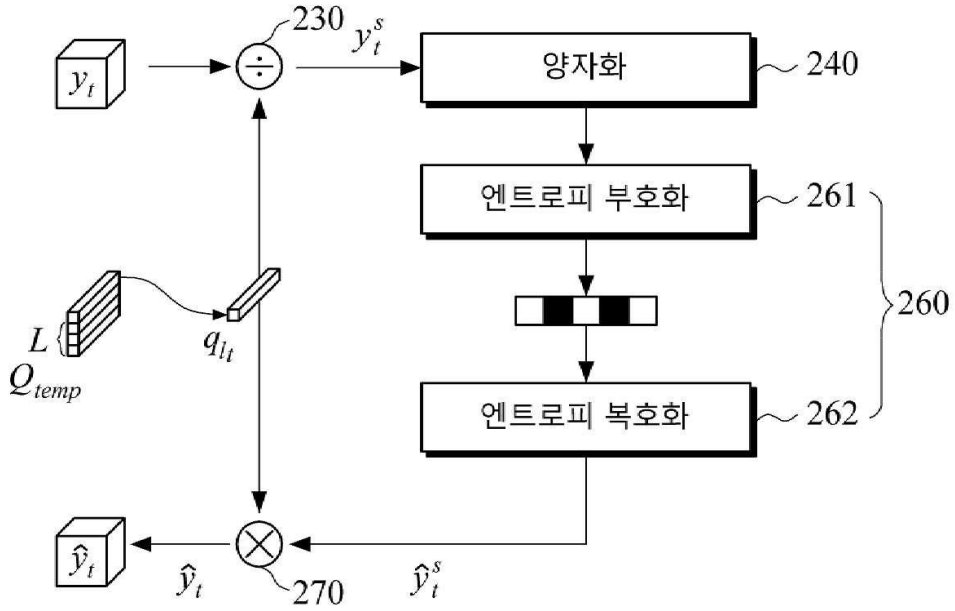
도면2b


$\hat{x}_f, \hat{x}_b$
x_t
210
양방향 움직임 벡터 추정
m_f, m_b
220
양방향 움직임 벡터에 대한 잠재 표현 변환
g_t
230
시간적 계층에 따른 양방향 움직임 벡터의 잠재표현 스케일링
g_t^s
240
스케일링된 양방향 움직임 벡터의 잠재표현 양자화
l_t
$\hat{g}_t^s$
250
양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현에 대한 확률 분포 추정
μ_t^g, σ_t^g
260
양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현에 대한 엔트로피 부호화 및 복호화
$\hat{g}_t^s$
270
양자화 및 스케일링된 양방향 움직임 벡터의 잠재표현에 대한 시간적 계층에 따른 리스케일링
$\hat{g}_t$
280
양방향 움직임 벡터 복원
$\hat{m}_f, \hat{m}_b$
290
움직임 벡터 잠재표현 기반 양방향 시간적 문맥 생성
C_t
b_m

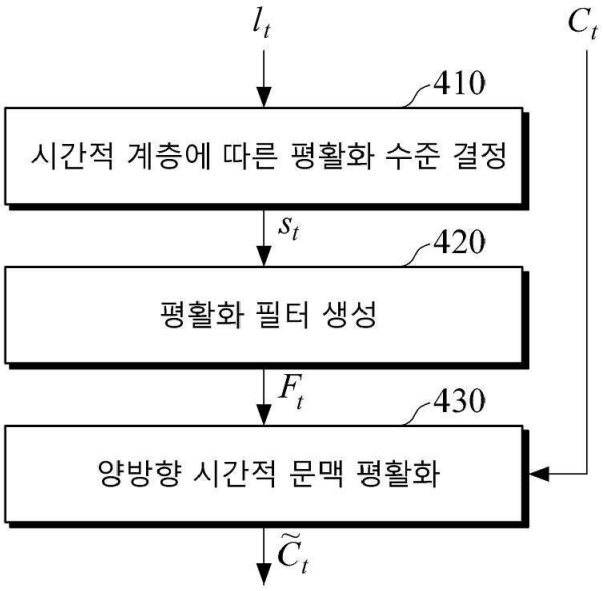
도면3a


q_{l_t}
I
I
$l_t = 0$
B
B
$l_t = 1$
B
B
B
B
$l_t = 2$
B
B
B
B
B
B
$l_t = 3$
L
Q_{temp}

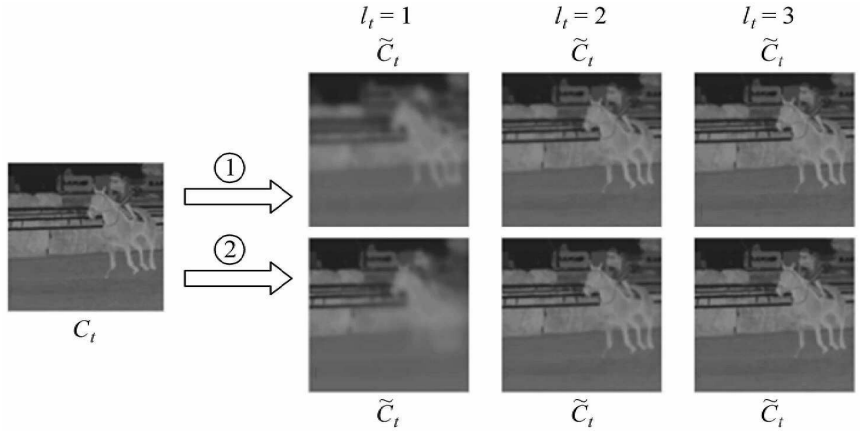
도면3b


230
y_t^s
양자화
240
엔트로피 부호화
261
260
엔트로피 복호화
262
y_t
L
Q_{temp}
q_{l_t}
$\hat{y}_t$
$\hat{y}_t$
270
$\hat{y}_t^s$

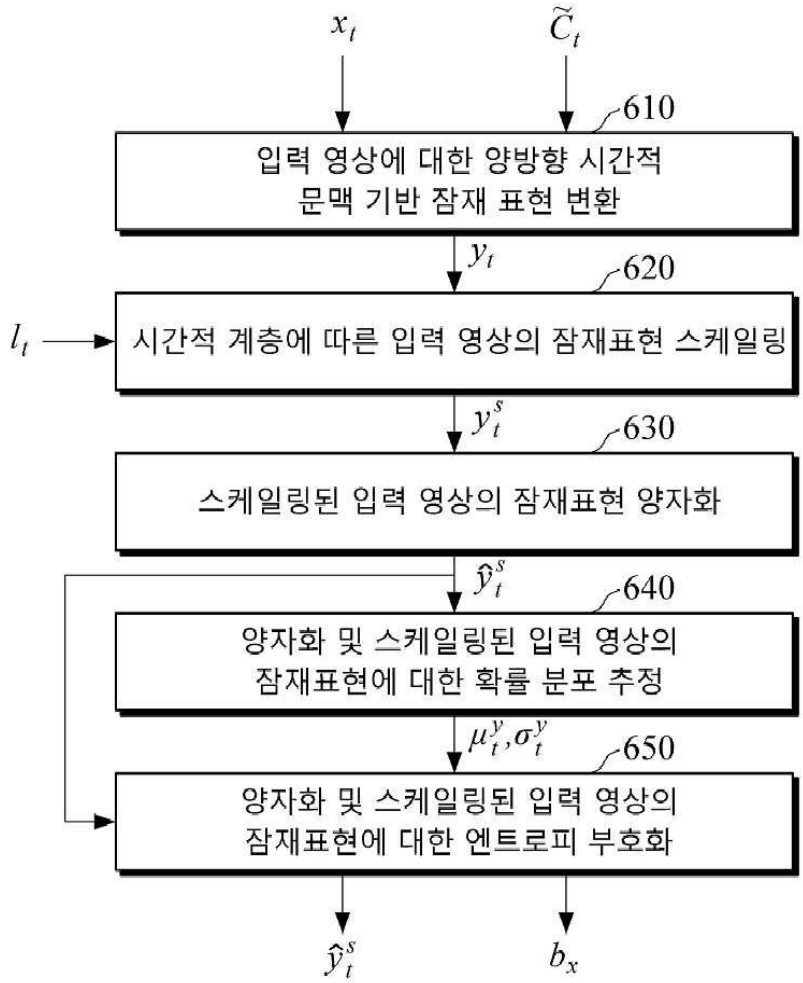
도면4


l_t
C_t
410
시간적 계층에 따른 평활화 수준 결정
s_t
420
평활화 필터 생성
F_t
430
양방향 시간적 문맥 평활화
$\tilde{C}_t$

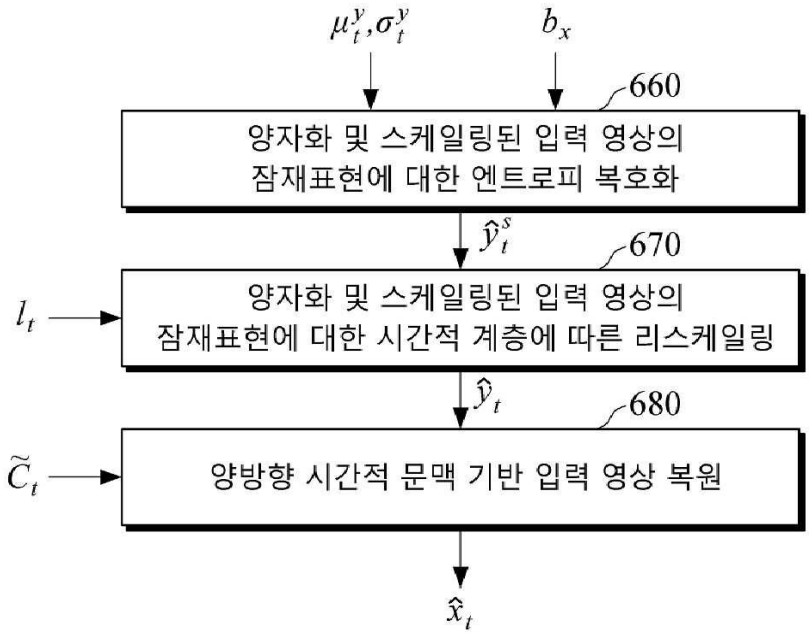
도면5


$l_t = 1$
$l_t = 2$
$l_t = 3$
$\tilde{C}_t$
①
②
C_t

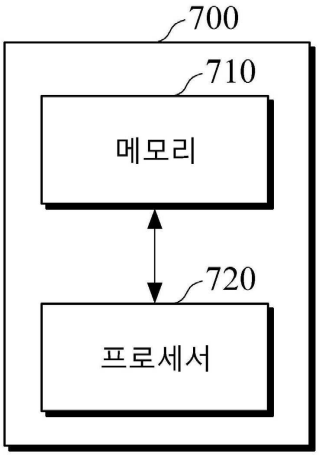
도면6a


x_t
$\tilde{C}_t$
610
입력 영상에 대한 양방향 시간적 문맥 기반 잠재 표현 변환
y_t
620
l_t
시간적 계층에 따른 입력 영상의 잠재표현 스케일링
y_t^s
630
스케일링된 입력 영상의 잠재표현 양자화
$\hat{y}_t^s$
640
양자화 및 스케일링된 입력 영상의 잠재표현에 대한 확률 분포 추정
μ_t^y, σ_t^y
650
양자화 및 스케일링된 입력 영상의 잠재표현에 대한 엔트로피 부호화
$\hat{y}_t^s$
b_x

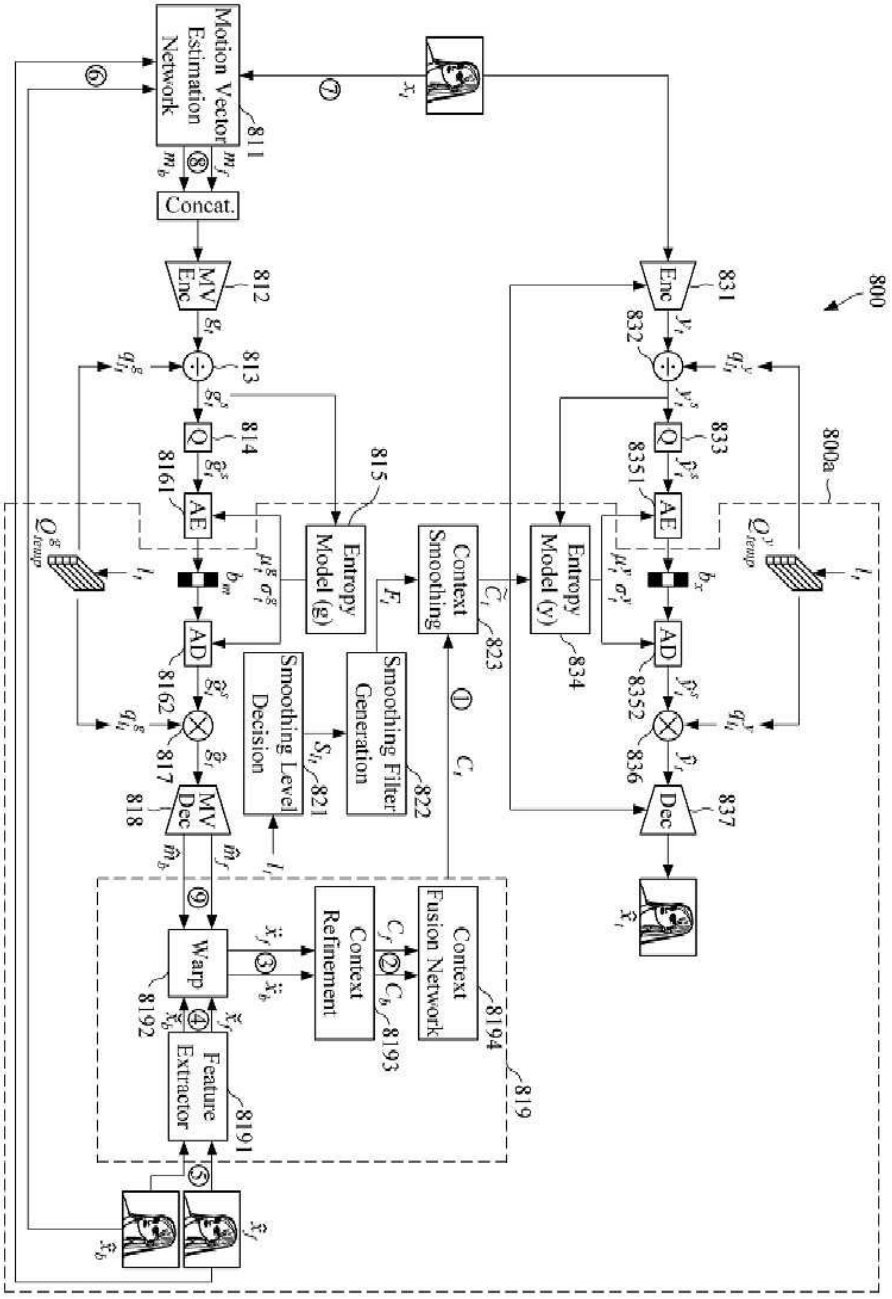
도면6b


μ_t^y, σ_t^y
b_x
660
양자화 및 스케일링된 입력 영상의
잠재표현에 대한 엔트로피 복호화
$\hat{y}_t^s$
670
양자화 및 스케일링된 입력 영상의
잠재표현에 대한 시간적 계층에 따른 리스케일링
l_t
$\hat{y}_t$
680
양방향 시간적 문맥 기반 입력 영상 복원
$\widetilde{C}_t$
$\hat{x}_t$

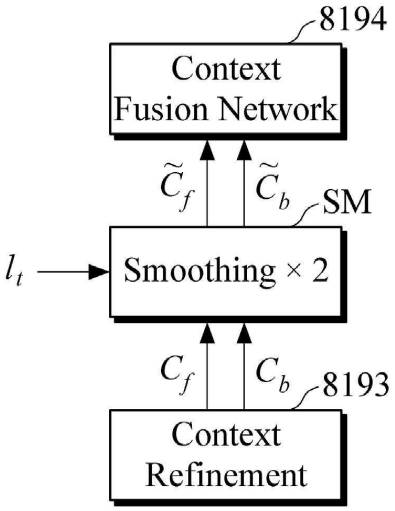
도면7


700
710
메모리
720
프로세서

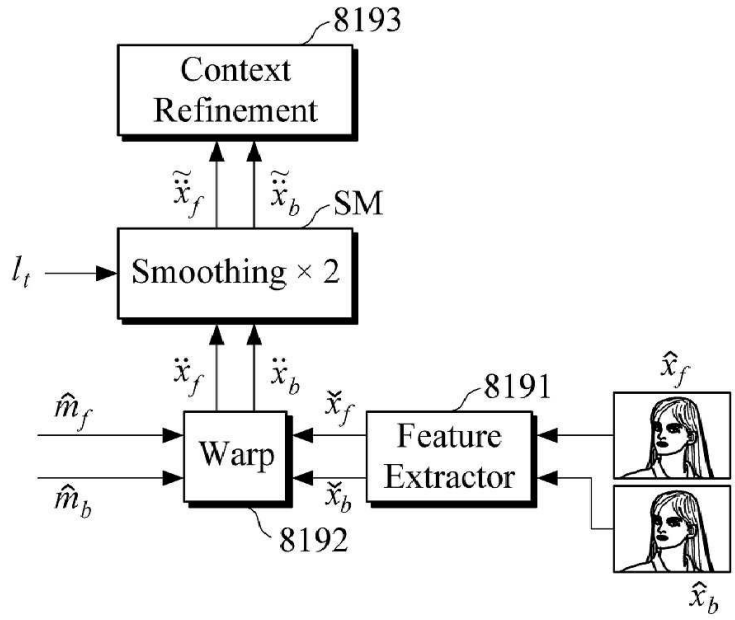
도면8a


800
800a
Motion Vector Estimation Network
811
Concat.
MV Enc
812
MV Dec
818
Enc
831
Dec
837
Q
813
814
833
AE
8161
8351
AD
8162
8352
Entropy Model (g)
815
Entropy Model (y)
834
Context Smoothing
823
Smoothing Level Decision
821
Smoothing Filter Generation
822
Warp
8192
Feature Extractor
8191
Context Refinement
8193
Context Fusion Network
8194
819

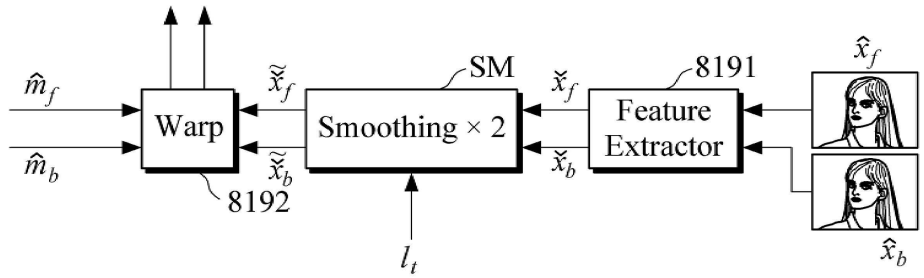
도면8b



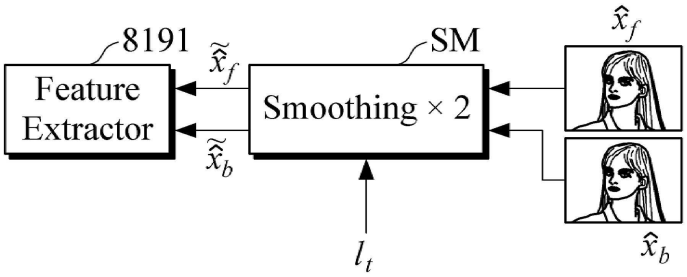
도면8c



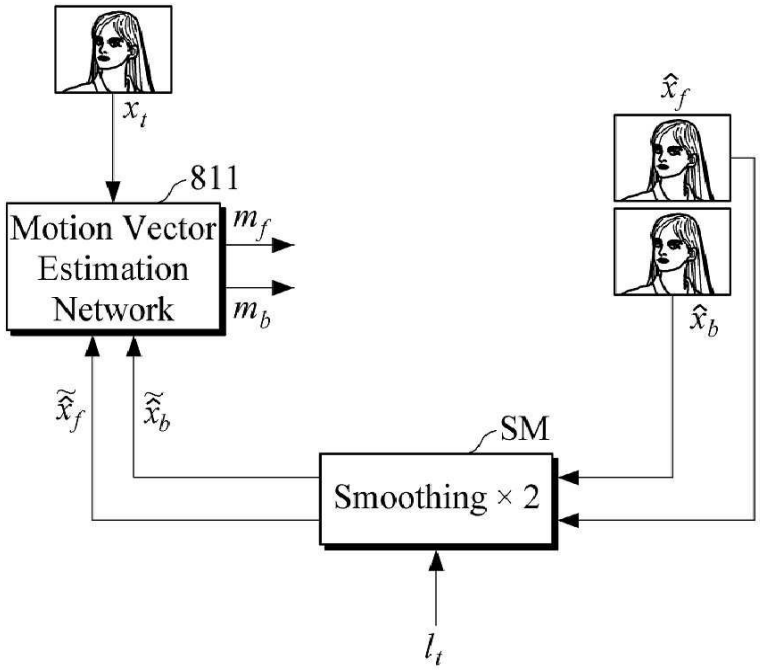
도면8d



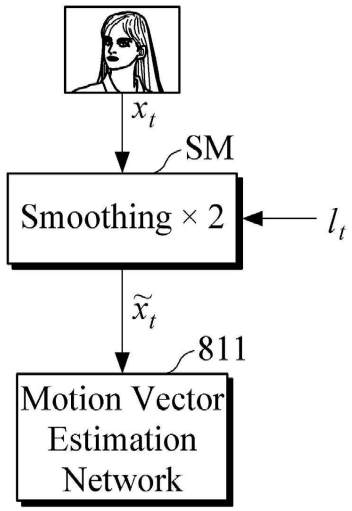
도면8e



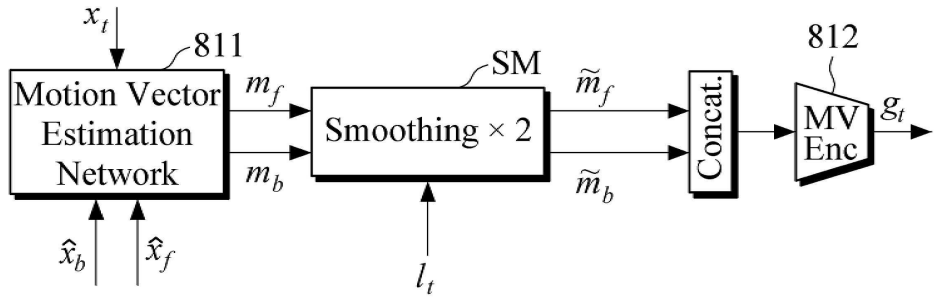
도면8f



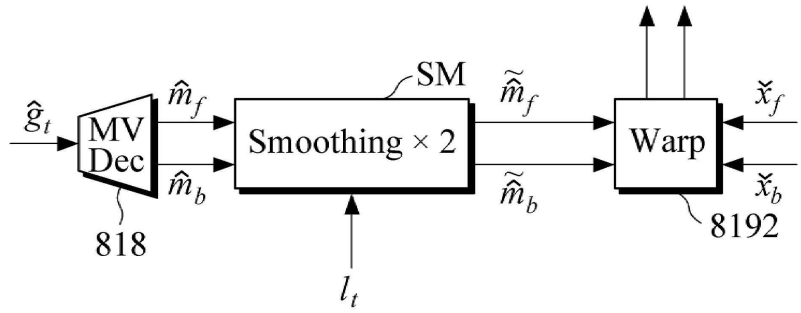
도면8g



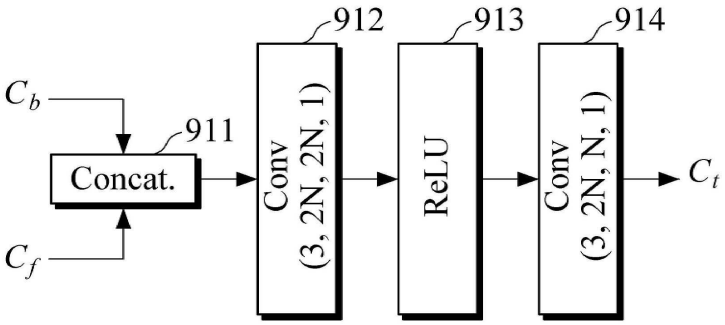
도면8h



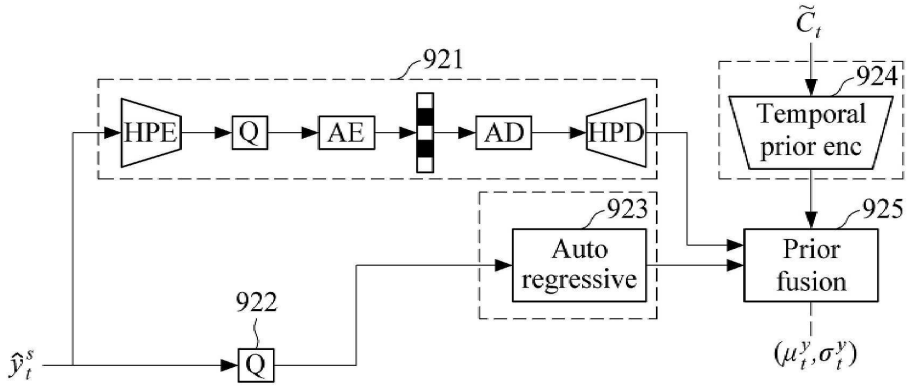
도면8i



도면9a



도면9b



도면10

