



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0110594
(43) 공개일자 2024년07월15일

- (51) 국제특허분류(Int. Cl.)
H04N 19/30 (2014.01) G06N 3/0464 (2023.01)
G06N 3/08 (2023.01) H04N 19/184 (2014.01)
H04N 19/20 (2014.01) H04N 19/70 (2014.01)
- (52) CPC특허분류
H04N 19/30 (2015.01)
G06N 3/0464 (2023.01)
- (21) 출원번호 10-2024-7018367
- (22) 출원일자(국제) 2022년11월01일
심사청구일자 없음
- (85) 번역문제출일자 2024년05월31일
- (86) 국제출원번호 PCT/KR2022/016900
- (87) 국제공개번호 WO 2023/075564
국제공개일자 2023년05월04일
- (30) 우선권주장
1020210148089 2021년11월01일 대한민국(KR)

- (71) 출원인
엘지전자 주식회사
서울특별시 영등포구 여의대로 128 (여의도동)
- (72) 발명자
장형문
서울특별시 서초구 양재대로11길 19 LG전자 특허센터
- (74) 대리인
성병기

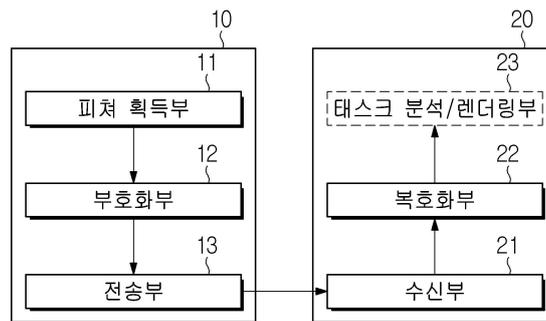
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 피쳐 부호화/복호화 방법, 장치 및 비트스트림을 저장한 기록 매체

(57) 요약

피쳐 부호화/복호화 방법, 장치 및 상기 피쳐 부호화 방법에 의해 생성된 컴퓨터 판독가능한 기록 매체가 제공된다. 본 개시에 따른 피쳐 복호화 방법은, CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 비트스트림으로부터 획득하는 단계, 상기 제1 정보에 기초하여, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 상기 비트스트림으로부터 획득하는 단계, 상기 제3 정보에 기초하여 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 상기 비트스트림으로부터 획득하는 단계 및 상기 제4 정보에 기초하여 상기 채널 레이어를 복원하는 단계를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/08 (2023.01)

H04N 19/184 (2015.01)

H04N 19/20 (2015.01)

H04N 19/70 (2015.01)

명세서

청구범위

청구항 1

피쳐 복호화 장치에 의해 수행되는 피쳐 복호화 방법에 있어서,

CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 비트스트림으로부터 획득하는 단계;

상기 제1 정보에 기초하여, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 상기 비트스트림으로부터 획득하는 단계;

상기 제3 정보에 기초하여 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 상기 비트스트림으로부터 획득하는 단계; 및

상기 제4 정보에 기초하여 상기 채널 레이어를 복원하는 단계;를 포함하는, 피쳐 복호화 방법.

청구항 2

제1 항에 있어서,

상기 제1 정보는 상기 최대 인터널 레이어의 수에서 1을 감한 값을 나타내는, 피쳐 복호화 방법.

청구항 3

제1 항에 있어서,

상기 제3 정보는 상기 채널 레이어의 수에서 1을 감한 값을 나타내는, 피쳐 복호화 방법.

청구항 4

제1 항에 있어서,

상기 제1 정보 및 상기 제3 정보에 기초하여 상기 인터널 레이어 간의 종속성에 관한 제5 정보를 상기 비트스트림으로부터 획득하는 단계;를 더 포함하는, 피쳐 복호화 방법.

청구항 5

제4 항에 있어서,

상기 제5 정보가 제1 값이면, 상기 인터널 레이어 간의 종속성이 있음을 나타내는, 피쳐 복호화 방법.

청구항 6

제5 항에 있어서,

상기 제1 값은 1인, 피쳐 복호화 방법.

청구항 7

제4 항에 있어서,

상기 제5 정보가 제2 값이면, 상기 인터널 레이어 간의 종속성이 없음을 나타내는, 피쳐 복호화 방법.

청구항 8

제7 항에 있어서,

상기 제2 값은 0인, 피쳐 복호화 방법.

청구항 9

제4 항에 있어서,

상기 제5 정보는 인터널 레이어 인덱스가 서로 다른 상기 인터널 레이어에 대하여만 획득되는, 피쳐 복호화 방법.

청구항 10

제4 항에 있어서,

상기 제5 정보가 상기 비트스트림에 존재하지 않는 경우, 상기 제5 정보는 제1 값으로 유도되는, 피쳐 복호화 방법.

청구항 11

10 항에 있어서,

기 제1 값은 1인, 피쳐 복호화 방법.

청구항 12

피쳐 부호화 장치에 의해 수행되는 피쳐 부호화 방법에 있어서,

CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계;

인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계; 및

상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계;를 포함하되,

상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화되는, 피쳐 부호화 방법.

청구항 13

피쳐 부호화 방법에 의해 생성된 비트스트림을 저장하는 컴퓨터 판독가능한 기록 매체에 있어서, 상기 피쳐 부호화 방법은,

CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계;

인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계; 및

상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계;를 포함하되,

상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화되는, 매체.

청구항 14

피쳐 부호화 방법에 의해 생성된 비트스트림을 전송하는 방법에 있어서, 상기 피쳐 부호화 방법은,

CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계;

인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계; 및

상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계;를 포함하되,

상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화되는, 비트스트림을 전송하는 방법.

발명의 설명

기술 분야

본 개시는 피쳐 부호화/복호화 방법 및 장치에 관한 것으로서, 보다 상세하게는, 인터널 레이어 및 채널 레이어

[0001]

에 대한 피쳐 부호화/복호화 방법, 장치 및 본 개시의 피쳐 부호화 방법/장치에 의해 생성된 비트스트림을 저장한 기록 매체에 관한 것이다.

배경 기술

[0002] 머신러닝 기술의 발전과 함께 영상 처리 기반의 인공지능 서비스에 대한 수요가 증가하고 있다. 인공지능 서비스에서 요구되는 방대한 양의 영상 데이터를 한정된 리소스 내에서 효과적으로 처리하기 위해서는, 머신 태스크 수행에 최적화된 영상 압축 기술이 필수적이다. 하지만, 기존의 영상 압축 기술은 휴먼 비전을 위한 고해상도, 고품질의 영상 처리를 목표로 발전해 왔는 바, 인공지능 서비스에는 부적합하다는 문제가 있다. 이에 따라, 인공지능 서비스에 적합한 머신 지향의 새로운 영상 압축 기술에 대한 연구 개발이 활발히 진행되고 있다.

발명의 내용

해결하려는 과제

- [0003] 본 개시는 부호화/복호화 효율이 향상된 피쳐 부호화/복호화 방법 및 장치를 제공하는 것을 목적으로 한다.
- [0004] 또한, 본 개시는, 네트워크에 종속적인 피쳐 압축을 지원하도록 시스템 레벨에서 필요한 정보를 비트스트림으로 전송하는 것을 목적으로 한다.
- [0005] 또한, 본 개시는 네트워크(network) 텐서(tensor)를 부/복호화 하는 과정에서 여러 태스크를 고려하여 비트스트림을 전송하는 것을 목적으로 한다.
- [0006] 또한, 본 개시는 머신 러닝 기반 영상 분석을 위해 영상 전송이 필요한 응용 분야에 적용되는 것을 목적으로 한다.
- [0007] 또한, 본 개시는 서로 다른 포맷으로 구성된 복수의 피쳐에 대한 압축을 지원하도록 시스템 레벨에서 필요한 정보를 비트스트림을 통해 전송하는 방법을 제안함으로써 다중 태스크를 구성하는 피쳐 구조에 대한 부호화/복호화를 지원할 수 있도록 자유도를 부여하는 것을 목적으로 한다.
- [0008] 또한, 본 개시는 본 개시에 따른 피쳐 부호화 방법 또는 장치에 의해 생성된 비트스트림을 전송하는 방법 또는 장치를 제공하는 것을 목적으로 한다.
- [0009] 또한, 본 개시는 본 개시에 따른 피쳐 부호화 방법 또는 장치에 의해 생성된 비트스트림을 저장한 기록 매체를 제공하는 것을 목적으로 한다.
- [0010] 또한, 본 개시는 본 개시에 따른 피쳐 복호화 장치에 의해 수신되고 복호화되어 피쳐의 복원에 이용되는 비트스트림을 저장한 기록 매체를 제공하는 것을 목적으로 한다.
- [0011] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0012] 본 개시의 일 양상에 따른 피쳐 복호화 장치에 의해 수행되는 피쳐 복호화 방법은, CFS(coded feature sequence) 내 허용된 최대 인터널(internal) 레이어의 수에 관한 제1 정보를 비트스트림으로부터 획득하는 단계, 상기 제1 정보에 기초하여, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 상기 비트스트림으로부터 획득하는 단계, 상기 제3 정보에 기초하여 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 상기 비트스트림으로부터 획득하는 단계 및 상기 제4 정보에 기초하여 상기 채널 레이어를 복원하는 단계를 포함할 수 있다.
- [0013] 본 개시의 일 실시예에 따르면, 상기 제1 정보는 상기 최대 인터널 레이어의 수에서 1을 감한 값을 나타낼 수 있다.
- [0014] 본 개시의 일 실시예에 따르면, 상기 제3 정보는 상기 채널 레이어의 수에서
- [0015] 1을 감한 값을 나타낼 수 있다.
- [0016] 본 개시의 일 실시예에 따르면, 상기 제1 정보 및 상기 제3 정보에 기초하여 상기 인터널 레이어 간의 종속성에

관한 제5 정보를 상기 비트스트림으로부터 획득하는 단계를 더 포함할 수 있다.

- [0017] 본 개시의 일 실시예에 따르면, 상기 제5 정보가 제1 값이면, 상기 인터널 레이어 간의 종속성이 있음을 나타낼 수 있다.
- [0018] 본 개시의 일 실시예에 따르면, 상기 제1 값은 1일 수 있다.
- [0019] 본 개시의 일 실시예에 따르면, 상기 제5 정보가 제2 값이면, 상기 인터널 레이어 간의 종속성이 없음을 나타낼 수 있다.
- [0020] 본 개시의 일 실시예에 따르면, 상기 제2 값은 0일 수 있다.
- [0021] 본 개시의 일 실시예에 따르면, 상기 제5 정보는 인터널 레이어 인덱스가 서로 다른 상기 인터널 레이어에 대하여만 획득될 수 있다.
- [0022] 본 개시의 일 실시예에 따르면, 상기 제5 정보가 상기 비트스트림에 존재하지 않는 경우, 상기 제5 정보는 제1 값으로 유도될 수 있다.
- [0023] 본 개시의 일 실시예에 따르면, 상기 제1 값은 1일 수 있다.
- [0024] 본 개시의 일 실시예에 따른 피쳐 부호화 장치에 의해 수행되는 피쳐 부호화 방법은 CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계 및 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계를 포함하되, 상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화될 수 있다.
- [0025] 본 개시의 일 실시예에 따른 피쳐 부호화 방법에 의해 생성된 비트스트림을 저장하는 컴퓨터 판독가능한 기록 매체에 있어서, 상기 피쳐 부호화 방법은, CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계 및 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계를 포함하되, 상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화될 수 있다.
- [0026] 본 개시의 일 실시예에 따른 피쳐 부호화 방법에 의해 생성된 비트스트림을 전송하는 방법에 있어서, 상기 피쳐 부호화 방법은, CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 관한 제1 정보를 결정하는 단계, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 상기 인터널 레이어의 채널 레이어의 수에 관한 제3 정보를 결정하는 단계 및 상기 채널 레이어 각각에 대한 식별자에 관한 제4 정보를 결정하는 단계를 포함하되, 상기 제1 정보, 상기 제2 정보, 상기 제3 정보 및 상기 제4 정보는 비트스트림으로 부호화될 수 있다.
- [0027] 본 개시의 또 다른 양상에 따른 기록 매체는, 본 개시의 피쳐 부호화 방법 또는 피쳐 부호화 장치에 의해 생성된 비트스트림을 저장할 수 있다.
- [0028] 본 개시의 또 다른 양상에 따른 비트스트림 전송 방법은, 본 개시의 피쳐 부호화 방법 또는 피쳐 부호화 장치에 의해 생성된 비트스트림을 피쳐 복호화 장치로 전송할 수 있다.
- [0029] 본 개시에 대하여 위에서 간략하게 요약된 특징들은 후술하는 본 개시의 상세한 설명의 예시적인 양상일 뿐이며, 본 개시의 범위를 제한하는 것은 아니다.

발명의 효과

- [0030] 본 개시에 따르면, 부호화/복호화 효율이 향상된 피쳐 정보 부호화/복호화 방법 및 장치가 제공될 수 있다.
- [0031] 또한, 본 개시에 따르면, 인터널 레이어의 수 및 채널 레이어의 수에 대한 정보를 시스템 레벨에서 비트스트림으로 전송/획득함으로써 자유도를 지원할 수 있다.
- [0032] 또한, 본 개시에 따르면, 인터널 레이어 간의 종속성을 나타내어 효율적인 피쳐 정보 부호화/복호화를 수행할 수 있다.
- [0033] 또한, 본 개시에 따르면, 머신 태스크가 어떤 레이어를 수행할 것인지에 대한 여부를 알 수 없는 경우도 지원할 수 있다.

- [0034] 또한, 본 개시에 따르면, 복원된 피쳐의 정확성이 향상됨으로써 피쳐에 대한 압축 효율을 향상시킬 수 있다.
- [0035] 또한, 본 개시에 따르면, 동일 시간 인스턴스에서 하나 이상의 피쳐 레이어 구조를 복호화 할 수 있으며, 각 피쳐 레이어 구조에 대해 개별적인 계층적 레이어를 가지고 상호 참조하는 구조를 지원할 수 있다.
- [0036] 또한, 본 개시에 따르면, 머신 태스크의 피쳐를 압축하는 시나리오와 다수 태스크 간의 피쳐를 공유하는 구조를 지원할 수 있다.
- [0037] 또한, 본 개시에 따르면, 하나 이상의 피쳐 레이어를 부호화하거나, 다수의 머신 태스크가 공유하는 피쳐를 부호화하고자 할 때 각 피쳐간 계층적 구조를 지원하고 이를 복호화 하기 위하여 필요한 정보를 비트스트림으로 전송할 수 있다.
- [0038] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0039] 도 1은 본 개시의 실시예들이 적용될 수 있는 VCM 시스템을 개략적으로 설명하기 위한 도면이다.
- 도 2는 본 개시의 실시예들이 적용될 수 있는 VCM 파이프라인 구조를 개략적으로 설명하기 위한 도면이다.
- 도 3은 본 개시의 실시예들이 적용될 수 있는 영상/비디오 인코더를 개략적으로 설명하기 위한 도면이다.
- 도 4는 본 개시의 실시예들이 적용될 수 있는 영상/비디오 디코더를 개략적으로 설명하기 위한 도면이다.
- 도 5는 본 개시의 실시예들이 적용될 수 있는 피쳐/피쳐맵 인코딩 절차를 개략적으로 설명하기 위한 흐름도이다.
- 도 6은 본 개시의 실시예들이 적용될 수 있는 피쳐/피쳐맵 디코딩 절차를 개략적으로 설명하기 위한 흐름도이다.
- 도 7은 본 개시의 실시예들이 적용될 수 있는 피쳐 추출 및 복원 방법의 일 예를 설명하기 위한 도면이다.
- 도 8은 피쳐 추출 네트워크를 이용한 피쳐 추출 방법의 일 예를 설명하기 위한 도면이다.
- 도 9는 추출된 피쳐의 일 예를 설명하기 위한 도면이다.
- 도 10은 추출된 피쳐 셋의 크기에 대한 일 예를 설명하기 위한 도면이다.
- 도 11은 피쳐 추출 및 복원 방법의 일 예를 설명하기 위한 도면이다.
- 도 12는 본 개시의 실시예들이 적용될 수 있는 소스 디바이스 및 수신 디바이스의 일 예를 설명하기 위한 도면이다.
- 도 13 내지 도 20은 본 개시의 실시예들에 적용될 수 있는 네트워크 입출력 단위의 일 예를 설명하기 위한 도면이다.
- 도 21 내지 도 25는 본 개시의 실시예들에 대한 멀티 피쳐 레이어의 일 예를 설명하기 위한 도면이다.
- 도 26은 본 개시의 일 실시예에 따른 네트워크 구조의 일 예를 설명하기 위한 도면이다.
- 도 27은 본 개시의 일 실시예에 따른 피쳐 복호화 장치가 수행하는 방법을 설명하기 위한 도면이다.
- 도 28은 본 개시의 일 실시예에 따른 피쳐 부호화 장치가 수행하는 방법을 설명하기 위한 도면이다.
- 도 29는 본 개시의 실시예들이 적용될 수 있는 콘텐츠 스트리밍 시스템의 일 예를 설명하기 위한 도면이다.
- 도 30은 본 개시의 실시예들이 적용될 수 있는 콘텐츠 스트리밍 시스템의 다른 예를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0040] 이하에서는 첨부한 도면을 참고로 하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나, 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다.

- [0041] 본 개시의 실시예를 설명함에 있어서 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그에 대한 상세한 설명은 생략한다. 그리고, 도면에서 본 개시에 대한 설명과 관계없는 부분은 생략하였으며, 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0042] 본 개시에 있어서, 어떤 구성요소가 다른 구성요소와 "연결", "결합" 또는 "접속"되어 있다고 할 때, 이는 직접적인 연결관계뿐만 아니라, 그 중간에 또 다른 구성요소가 존재하는 간접적인 연결관계도 포함할 수 있다. 또한 어떤 구성요소가 다른 구성요소를 "포함한다" 또는 "가진다"고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 배제하는 것이 아니라 또 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0043] 본 개시에 있어서, 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용되며, 특별히 언급되지 않는 한 구성요소들간의 순서 또는 중요도 등을 한정하지 않는다. 따라서, 본 개시의 범위 내에서 일 실시예에서의 제1 구성요소는 다른 실시예에서 제2 구성요소라고 칭할 수도 있고, 마찬가지로 일 실시예에서의 제2 구성요소를 다른 실시예에서 제1 구성요소라고 칭할 수도 있다.
- [0044] 본 개시에 있어서, 서로 구별되는 구성요소들은 각각의 특징을 명확하게 설명하기 위함이며, 구성요소들이 반드시 분리되는 것을 의미하지는 않는다. 즉, 복수의 구성요소가 통합되어 하나의 하드웨어 또는 소프트웨어 단위로 이루어질 수도 있고, 하나의 구성요소가 분산되어 복수의 하드웨어 또는 소프트웨어 단위로 이루어질 수도 있다. 따라서, 별도로 언급하지 않더라도 이와 같이 통합된 또는 분산된 실시예도 본 개시의 범위에 포함된다.
- [0045] 본 개시에 있어서, 다양한 실시예에서 설명하는 구성요소들이 반드시 필수적인 구성요소들을 의미하는 것은 아니며, 일부는 선택적인 구성요소일 수 있다. 따라서, 일 실시예에서 설명하는 구성요소들의 부분집합으로 구성되는 실시예도 본 개시의 범위에 포함된다. 또한, 다양한 실시예에서 설명하는 구성요소들에 추가적으로 다른 구성요소를 포함하는 실시예도 본 개시의 범위에 포함된다.
- [0046] 본 개시는 영상의 부호화 및 복호화에 관한 것으로서, 본 개시에서 사용되는 용어는, 본 개시에서 새롭게 정의되지 않는 한 본 개시가 속한 기술 분야에서 통용되는 통상의 의미를 가질 수 있다.
- [0047] 본 개시는 VVC(Versatile Video Coding) 표준 및/또는 VCM(Video Coding for Machines) 표준에 개시되는 방법에 적용될 수 있다. 또한, 본 개시는 EVC (essential video coding) 표준, AV1 (AOMedia Video 1) 표준, AVS2 (2nd generation of audio video coding standard) 또는 차세대 비디오/영상 코딩 표준(e.g., H.267 or H.268 등)에 개시되는 방법에 적용될 수 있다.
- [0048] 본 개시는 비디오/영상 코딩에 관한 다양한 실시예들을 제시하며, 다른 언급이 없는 한 상기 실시예들은 서로 조합되어 수행될 수도 있다. 본 개시에서 "비디오(video)"는 시간의 흐름에 따른 일련의 영상(image)들의 집합을 의미할 수 있다. "영상(image)"은 AI(artificial intelligence)에 의해 생성된 정보일 수 있다. AI가 일련의 태스크를 수행하는 과정에서 사용하는 입력 정보, 정보 처리 과정 중에 발생하는 정보와 출력하는 정보가 영상(image)으로 쓰일 수 있다. "픽처(picture)"는 일반적으로 특정 시간대의 하나의 영상을 나타내는 단위를 의미하며, 슬라이스(slice)/타일(tile)은 부호화에 있어서 픽처의 일부를 구성하는 부호화 단위이다. 하나의 픽처는 하나 이상의 슬라이스/타일로 구성될 수 있다. 또한, 슬라이스/타일은 하나 이상의 CTU(coding tree unit)를 포함할 수 있다. 상기 CTU는 하나 이상의 CU로 분할될 수 있다. 타일은 픽처 내의 특정 타일 행(Tile Row) 및 특정 타일 열(Tile Column) 내에 존재하는 사각 영역으로, 복수의 CTU로 구성될 수 있다. 타일 열은 CTU들의 사각 영역으로 정의될 수 있으며, 픽처의 높이와 동일한 높이를 가지고, 픽처 파라미터 셋과 같은 비트스트림 부분으로부터 시그널링 되는 선택스 요소에 의하여 명세되는 너비를 가질 수 있다. 타일 행은 CTU들의 사각 영역으로 정의될 수 있으며, 픽처의 너비와 동일한 너비를 가지고, 픽처 파라미터 셋과 같은 비트스트림 부분으로부터 시그널링 되는 선택스 요소에 의하여 명세되는 높이를 가질 수 있다. 타일 스캔은 픽처를 분할하는 CTU들의 소정의 연속된 순서 지정 방법이다. 여기서, CTU들은 타일 내에서 CTU 래스터 스캔(raster scan)에 따라 연속적으로 순서를 부여받을 수 있고, 픽처내의 타일들은 픽처의 타일들의 래스터 스캔 순서에 따라 연속적으로 순서를 부여받을 수 있다. 슬라이스는 정수개의 완전한 타일들을 포함하거나, 하나의 픽처의 하나의 타일 내의 연속하는 정수개의 완전한 CTU 행을 포함할 수 있다. 슬라이스는 하나의 싱글 NAL 유닛에 독립적으로 포함될 수 있다. 하나의 픽처는 하나 이상의 타일 그룹으로 구성될 수 있다. 하나의 타일 그룹은 하나 이상의 타일들을 포함할 수 있다. 브릭은 픽처 내 타일 이내의 CTU행들의 사각 영역을 나타낼 수 있다. 하나의 타일은 하나 이상의 브릭(Brick)을 포함할 수 있다. 브릭은 타일 내 CTU 행들의 사각 영역을 나타낼 수 있다. 하나의 타일은 복수의 브릭으로 분할될 수 있으며, 각각의 브릭은 타일에 속한 하나 이상의 CTU행을 포함할 수 있다. 복수의 브릭으로 분할되지 않는 타일 또한 브릭으로 취급될 수 있다.

- [0049] 본 개시에서 "픽셀(pixel)" 또는 "펠(peI)"은 하나의 픽처(또는 영상)를 구성하는 최소의 단위를 의미할 수 있다. 또한, 픽셀에 대응하는 용어로서 "샘플(sample)"이 사용될 수 있다. 샘플은 일반적으로 픽셀 또는 픽셀의 값을 나타낼 수 있으며, 루마(luma) 성분의 픽셀/픽셀값만을 나타낼 수도 있고, 크로마(chroma) 성분의 픽셀/픽셀 값을 나타낼 수도 있다.
- [0050] 일 실시예에서, 특히 VCM에 적용되는 경우, 픽셀/픽셀값은 다른 특성 및 의미를 갖는 성분들의 집합으로 구성된 픽처가 있을 때 각 성분들의 독립적 정보 혹은 조합, 합성, 분석을 통해 생성된 성분의 픽셀/픽셀값을 나타낼 수도 있다. 예를 들어 RGB 입력에서 R의 픽셀/픽셀값만을 나타낼 수도 있고 G의 픽셀/픽셀값만을 나타낼 수도 있고 B의 픽셀/픽셀값만을 나타낼 수도 있다. 예를 들어 R, G, B 성분을 이용해 합성된 루마(Luma) 성분의 픽셀/픽셀값만을 나타낼 수도 있다. 예를 들어 R, G, B 성분을 성분에서 분석을 통해 추출한 영상, 정보의 픽셀/픽셀 값을 나타낼 수도 있다.
- [0051] 본 개시에서 "유닛(unit)"은 영상 처리의 기본 단위를 나타낼 수 있다. 유닛은 픽처의 특정 영역 및 해당 영역에 관련된 정보 중 적어도 하나를 포함할 수 있다. 하나의 유닛은 하나의 루마 블록 및 두개의 크로마(e.g., Cb, Cr) 블록을 포함할 수 있다. 유닛은 경우에 따라서 "샘플 어레이", "블록(block)" 또는 "영역(area)" 등의 용어와 혼용하여 사용될 수 있다. 일반적인 경우, MxN 블록은 M개의 열과 N개의 행으로 이루어진 샘플들(또는 샘플 어레이) 또는 변환 계수(transform coefficient)들의 집합(또는 어레이)을 포함할 수 있다. 일 실시예에서, 특히 VCM에 적용되는 경우, 유닛은 특정 태스크를 수행하기 위한 정보를 담고 있는 기본 단위를 나타낼 수 있다.
- [0052] 본 개시에서 "현재 블록"은 "현재 코딩 블록", "현재 코딩 유닛", "부호화 대상 블록", "복호화 대상 블록" 또는 "처리 대상 블록" 중 하나를 의미할 수 있다. 예측이 수행되는 경우, "현재 블록"은 "현재 예측 블록" 또는 "예측 대상 블록"을 의미할 수 있다. 변환(역변환)/양자화(역양자화)가 수행되는 경우, "현재 블록"은 "현재 변환 블록" 또는 "변환 대상 블록"을 의미할 수 있다. 필터링이 수행되는 경우, "현재 블록"은 "필터링 대상 블록"을 의미할 수 있다.
- [0053] 또한, 본 개시에서 "현재 블록"은 크로마 블록이라는 명시적인 기재가 없는 한 "현재 블록의 루마 블록"을 의미할 수 있다. "현재 블록의 크로마 블록"은 명시적으로 "크로마 블록" 또는 "현재 크로마 블록"과 같이 크로마 블록이라는 명시적인 기재를 포함하여 표현될 수 있다.
- [0054] 본 개시에서 "/"와 ", "는 "및/또는"으로 해석될 수 있다. 예를 들어, "A/B"와 "A, B"는 "A 및/또는 B"로 해석될 수 있다. 또한, "A/B/C"와 "A, B, C"는 "A, B 및/또는 C 중 적어도 하나"를 의미할 수 있다.
- [0055] 본 개시에서 "또는"은 "및/또는"으로 해석될 수 있다. 예를 들어, "A 또는 B"는, 1) "A" 만을 의미하거나 2) "B" 만을 의미하거나, 3) "A 및 B"를 의미할 수 있다. 또는, 본 개시에서 "또는"은 "추가적으로 또는 대체적으로(Additionally or alternatively)"를 의미할 수 있다.
- [0056] 본 개시는 VCM(Video/image coding for machines)에 관한 것이다.
- [0057] VCM은 머신 비전을 목적으로 하여 소스 영상/비디오의 일부 또는 소스 영상/비디오로부터 획득된 정보를 부호화/복호화하는 압축 기술을 일컫는다. VCM에서 부호화/복호화 대상은 피쳐(feature)로 지칭될 수 있다. 피쳐는 태스크 목적, 요구사항, 주변 환경 등에 기반하여 소스 영상/비디오로부터 추출된 정보를 의미할 수 있다. 피쳐는 소스 영상/비디오와는 상이한 정보 형태를 가질 수 있으며, 이에 따라 피쳐의 압축 방법 및 표현 형식 또한 비디오 소스와는 상이할 수 있다.
- [0058] VCM은 다양한 응용 분야에 적용될 수 있다. 예를 들어, 물체나 사람을 인식하고 추적하는 감시 시스템(Surveillance system)에 있어서, VCM은 객체 인식 정보를 저장하거나 전송하기 위해 이용될 수 있다. 또한, 지능형 운송(Intelligent Transportation) 또는 스마트 트래픽 시스템(Smart Traffic system)에 있어서, VCM은 GPS로부터 수집한 차량의 위치 정보, 라이다(LIDAR), 레이더(Radar) 등으로부터 수집한 센싱 정보 및 각종 차량 제어 정보를 다른 차량이나 인프라 스트럭처(infrastructure)로 전송하기 위해 이용될 수 있다. 또한, 스마트 시티(Smart city) 분야에 있어서, VCM은 상호 연결된 센서 노드 또는 장치의 개별 태스크 수행을 위해 이용될 수 있다.
- [0059] 본 개시는 피쳐/피쳐맵 코딩에 관한 다양한 실시예들을 제공한다. 다른 특별한 언급이 없는 한, 본 개시의 실시예들은 각각 개별적으로 구현될 수도 있고, 또는 2 이상의 조합으로 구현될 수도 있다.
- [0060] **VCM 시스템 개요**

- [0061] 도 1은 본 개시의 실시예들이 적용될 수 있는 VCM 시스템을 개략적으로 나타낸 도면이다.
- [0062] 도 1을 참조하면, VCM 시스템은 부호화 장치(10) 및 복호화 장치(20)를 포함할 수 있다.
- [0063] 부호화 장치(10)는 소스 영상/비디오로부터 추출된 피쳐/피쳐맵을 압축/부호화하여 비트스트림을 생성하고, 생성된 비트스트림을 저장매체 또는 네트워크를 통해 복호화 장치(20)로 전송할 수 있다. 부호화 장치(10)는 피쳐 부호화 장치로 지칭될 수도 있다. VCM 시스템에서, 피쳐/피쳐맵은 신경망의 각 은닉층(hidden layer)에서 생성될 수 있다. 생성된 피쳐맵의 크기 및 채널 수는 신경망의 종류나 은닉층의 위치에 따라 달라질 수 있다. 본 개시에서, 피쳐맵은 피쳐셋으로 지칭될 수 있으며, 피쳐 또는 피쳐맵은 '피쳐 정보'로 지칭될 수 있다.
- [0064] 부호화 장치(10)는 피쳐 획득부(11), 부호화부(12) 및 전송부(13)를 포함할 수 있다.
- [0065] 피쳐 획득부(11)는 소스 영상/비디오에 대한 피쳐/피쳐맵을 획득할 수 있다. 실시예에 따라, 피쳐 획득부(11)는 외부 장치, 예컨대 피쳐 추출 네트워크로부터 피쳐/피쳐맵을 획득할 수 있다. 이 경우, 피쳐 획득부(11)는 피쳐 수신 인터페이스 기능을 수행하게 된다. 또는, 피쳐 획득부(11)는 소스 영상/비디오를 입력으로 하여 신경망(e.g., CNN, DNN 등)을 실행함으로써 피쳐/피쳐맵을 획득할 수도 있다. 이 경우, 피쳐 획득부(11)는 피쳐 추출 네트워크 기능을 수행하게 된다.
- [0066] 실시예에 따라, 부호화 장치(10)는 소스 영상/비디오를 획득하기 위한 소스 영상 생성부(미도시)를 더 포함할 수 있다. 소스 영상 생성부는 이미지 센서, 카메라 모듈 등으로 구현될 수 있으며, 영상/비디오의 캡처, 합성 또는 생성 과정 등을 통해 소스 영상/비디오를 획득할 수 있다. 이 경우, 생성된 소스 영상/비디오는 피쳐 추출 네트워크로 전달되어, 피쳐/피쳐맵을 추출하기 위한 입력 데이터로 이용될 수 있다.
- [0067] 부호화부(12)는 피쳐 획득부(11)에 의해 획득된 피쳐/피쳐맵을 부호화할 수 있다. 부호화부(12)는 부호화 효율을 높이기 위해 예측, 변환, 양자화 등 일련의 절차를 수행할 수 있다. 부호화된 데이터(부호화된 피쳐/피쳐맵 정보)는 비트스트림 형태로 출력될 수 있다. 부호화된 피쳐/피쳐맵 정보를 포함하는 비트스트림은 VCM 비트스트림으로 지칭될 수 있다.
- [0068] 전송부(13)는 비트스트림 형태로 출력된 피쳐/피쳐맵 정보 또는 데이터를 획득할 수 있으며, 이를 파일 또는 스트리밍 형태로 디지털 저장매체 또는 네트워크를 통해 복호화 장치(20) 또는 다른 외부 객체로 전달할 수 있다. 여기서, 디지털 저장 매체는 USB, SD, CD, DVD, 블루레이, HDD, SSD 등 다양한 저장 매체들을 포함할 수 있다. 전송부(13)는 소정의 파일 포맷을 갖는 미디어 파일을 생성하기 위한 엘리먼트들 또는 방송/통신 네트워크를 통한 데이터 전송을 위한 엘리먼트들을 포함할 수 있다. 전송부(13)는 부호화부(12)와는 별개의 전송 장치로 구비될 수 있으며, 이 경우 전송 장치는 비트스트림 형태로 출력된 피쳐/피쳐맵 정보 또는 데이터를 획득하는 적어도 하나의 프로세서와 이를 파일 또는 스트리밍 형태로 전달하는 전송부를 포함할 수 있다.
- [0069] 복호화 장치(20)는 부호화 장치(10)로부터 피쳐/피쳐맵 정보를 획득하고, 획득된 정보에 기반하여 피쳐/피쳐맵을 복원할 수 있다.
- [0070] 복호화 장치(20)는 수신부(21) 및 복호화부(22)를 포함할 수 있다.
- [0071] 수신부(21)는 부호화 장치(10)로부터 비트스트림을 수신하고, 수신된 비트스트림으로부터 피쳐/피쳐맵 정보를 획득하여 복호화부(22)로 전달할 수 있다.
- [0072] 복호화부(22)는 획득된 피쳐/피쳐맵 정보에 기반하여 피쳐/피쳐맵을 복호화할 수 있다. 복호화부(22)는 복호화 효율을 높이기 위해 부호화부(12)의 동작에 대응하는 역양자화, 역변환, 예측 등 일련의 절차를 수행할 수 있다.
- [0073] 실시예에 따라, 복호화 장치(20)는 태스크 분석/렌더링부(23)를 더 포함할 수 있다.
- [0074] 태스크 분석/렌더링부(23)는 복호화된 피쳐/피쳐맵에 기반하여 태스크 분석을 수행할 수 있다. 또한, 태스크 분석/렌더링부(23)는 복호화된 피쳐/피쳐맵을 태스크 수행에 적합한 형태로 렌더링(rendering)할 수 있다. 태스크 분석 결과 및 렌더링된 피쳐/피쳐맵에 기반하여 다양한 머신(지향형) 태스크가 수행될 수 있다.
- [0075] 이상, VCM 시스템은 사용자 및/또는 머신의 요청, 태스크 목적 및 주변 환경에 따라 소스 영상/비디오로부터 추출된 피쳐를 부호화/복호화하고, 복호화된 피쳐에 기반하여 다양한 머신(지향형) 태스크들을 수행할 수 있다. VCM 시스템은 비디오/영상 코딩 시스템을 확장/재설계함으로써 구현될 수도 있으며, VCM 표준에서 정의되는 다양한 부호화/복호화 방법들을 수행할 수 있다.

[0076] **VCM 파이프라인**

[0077] 도 2는 본 개시의 실시예들이 적용될 수 있는 VCM 파이프라인 구조를 개략적으로 나타낸 도면이다.

[0078] 도 2를 참조하면, VCM 파이프라인(200)은 영상/비디오의 부호화/복호화를 위한 제1 파이프라인(210) 및 피쳐/피쳐맵의 부호화/복호화를 위한 제2 파이프라인(220)을 포함할 수 있다. 본 개시에서, 제1 파이프라인(210)은 비디오 코덱 파이프라인으로 지칭될 수 있고, 제2 파이프라인(220)은 피쳐 코덱 파이프라인으로 지칭될 수 있다.

[0079] 제1 파이프라인(210)은 입력된 영상/비디오를 부호화하는 제1 스테이지(211) 및 부호화된 영상/비디오를 복호화하여 복원된 영상/비디오를 생성하는 제2 스테이지(212)를 포함할 수 있다. 복원된 영상/비디오는 사람의 시청용, 즉 휴먼 비전을 위해 이용될 수 있다.

[0080] 제2 파이프라인(220)은 입력된 영상/비디오로부터 피쳐/피쳐맵을 추출하는 제3 스테이지(221), 추출된 피쳐/피쳐맵을 부호화하는 제4 스테이지(222), 및 부호화된 피쳐/피쳐맵을 복호화하여 복원된 피쳐/피쳐맵을 생성하는 제5 스테이지(223)를 포함할 수 있다. 복원된 피쳐/피쳐맵은 머신 (비전) 태스크를 위해 이용될 수 있다. 여기서, 머신 (비전) 태스크란 머신에 의해 영상/비디오가 소비되는 형태의 태스크를 의미할 수 있다. 머신 (비전) 태스크는, 예컨대, 감시(Surveillance), 지능형 교통(Intelligent Transportation), 스마트 시티(Smart City), 지능형 산업(Intelligent Industry), 지능형 콘텐츠(Intelligent Content) 등과 같은 서비스 시나리오에 적용될 수 있다. 실시예에 따라, 복원된 피쳐/피쳐맵은 휴먼 비전을 위해 이용될 수도 있다.

[0081] 실시예에 따라, 제4 스테이지(222)에서 부호화된 피쳐/피쳐맵은 제1 스테이지(221)로 전달되어 영상/비디오를 부호화하는 데 이용될 수 있다. 이 경우, 부호화된 피쳐/피쳐맵에 기반하여 부가 비트스트림이 생성될 수 있으며, 생성된 부가 비트스트림은 제2 스테이지(222)로 전달되어 영상/비디오를 복호화하는 데 이용될 수 있다.

[0082] 실시예에 따라, 제5 스테이지(223)에서 복호화된 피쳐/피쳐맵은 제2 스테이지(222)로 전달되어 영상/비디오를 복호화하는 데 이용될 수 있다.

[0083] 도 2에서는 VCM 파이프라인(200)이 제1 파이프라인(210) 및 제2 파이프라인(220)을 포함하는 경우를 도시하나, 이는 예시적인 것일 뿐 본 개시의 실시예들이 이에 제한되는 것은 아니다. 예를 들어, VCM 파이프라인(200)은 제2 파이프라인(220)만을 포함할 수도 있고, 또는 제2 파이프라인(220)이 복수의 피쳐 코덱 파이프라인들로 확장될 수도 있다.

[0084] 한편, 제1 파이프라인(210)에서, 제1 스테이지(211)는 영상/비디오 인코더에 의해 수행되고, 제2 스테이지(212)는 영상/비디오 디코더에 의해 수행될 수 있다. 또한, 제2 파이프라인(220)에서, 제3 스테이지(221)는 VCM 인코더(또는, 피쳐/피쳐맵 인코더)에 의해 수행되고, 제4 스테이지(222)는 VCM 디코더(또는, 피쳐/피쳐맵 디코더)에 의해 수행될 수 있다. 이하, 인코더/디코더 구조를 상세히 설명한다.

[0085] **인코더(Encoder)**

[0086] 도 3은 본 개시의 실시예들이 적용될 수 있는 영상/비디오 인코더를 개략적으로 나타낸 도면이다.

[0087] 도 3을 참조하면, 영상/비디오 인코더(300)는 영상 분할부(image partitioner, 310), 예측부(predictor, 320), 레지듀얼 처리부(residual processor, 330), 엔트로피 인코딩부(entropy encoder, 340), 가산부(adder, 350), 필터링부(filter, 360), 및 메모리(memory, 370)를 포함할 수 있다. 예측부(320)는 인터 예측부(321) 및 인트라 예측부(322)를 포함할 수 있다. 레지듀얼 처리부(330)는 변환부(transformer, 332), 양자화부(quantizer, 333), 역양자화부(dequantizer, 334) 및 역변환부(inverse transformer, 335)를 포함할 수 있다. 레지듀얼 처리부(330)는 감산부(subtractor, 331)를 더 포함할 수 있다. 가산부(350)는 복원부(reconstructor) 또는 복원 블록 생성부(reconstructged block generator)로 지칭될 수 있다. 상술한 영상 분할부(310), 예측부(320), 레지듀얼 처리부(330), 엔트로피 인코딩부(340), 가산부(350) 및 필터링부(360)는 실시예에 따라 하나 이상의 하드웨어 컴포넌트(예컨대, 인코더 칩셋 또는 프로세서)에 의하여 구성될 수 있다. 또한, 메모리(370)는 DPB(decoded picture buffer)를 포함할 수 있고, 디지털 저장 매체에 의해 구성될 수도 있다. 상술한 하드웨어 컴포넌트는 메모리(370)를 내/외부 컴포넌트로 더 포함할 수도 있다.

[0088] 영상 분할부(310)는 영상/비디오 인코더(300)에 입력된 입력 영상(또는, 픽처, 프레임)을 하나 이상의 프로세싱 유닛(processing unit)으로 분할할 수 있다. 일 예로, 프로세싱 유닛은 코딩 유닛(coding unit, CU)으로 지칭될 수도 있다. 코딩 유닛은 코딩 트리 유닛(coding tree unit, CTU) 또는 최대 코딩 유닛(largest coding unit, LCU)으로부터 QTBT(Quad-tree binary-tree ternary-tree) 구조에 따라 재귀적으로(recursively) 분할될 수 있다. 예를 들어, 하나의 코딩 유닛은 쿼드 트리 구조, 바이너리 트리 구조, 및/또는 터너리 구조를 기반

으로 하위(deeper) 맵스의 복수의 코딩 유닛들로 분할될 수 있다. 이 경우, 예를 들어 쿼드 트리 구조가 먼저 적용되고 바이너리 트리 구조 및/또는 터너리 구조가 나중에 적용될 수 있다. 또는, 바이너리 트리 구조가 먼저 적용될 수도 있다. 더 이상 분할되지 않는 최종 코딩 유닛을 기반으로 본 개시에 따른 영상/비디오 코딩 절차가 수행될 수 있다. 이 경우, 영상 특성에 따른 코딩 효율 등을 기반으로 최대 코딩 유닛이 바로 최종 코딩 유닛으로 사용될 수도 있고, 또는 필요에 따라 코딩 유닛은 재귀적으로(recursively) 보다 하위 맵스의 코딩 유닛들로 분할되어 최적의 사이즈의 코딩 유닛이 최종 코딩 유닛으로 사용될 수도 있다. 여기서, 코딩 절차라 함은 후술하는 예측, 변환, 및 복원 등의 절차를 포함할 수 있다. 다른 예로, 프로세싱 유닛은 예측 유닛(PU: Prediction Unit) 또는 변환 유닛(TU: Transform Unit)을 더 포함할 수 있다. 이 경우, 예측 유닛 및 변환 유닛은 각각 상술한 최종 코딩 유닛으로부터 분할 또는 파티셔닝될 수 있다. 예측 유닛은 샘플 예측의 단위일 수 있고, 변환 유닛은 변환 계수를 유도하는 단위 및/또는 변환 계수로부터 레지듀얼 신호(residual signal)를 유도하는 단위일 수 있다.

[0089] 유닛은 경우에 따라 블록(block) 또는 영역(area) 등의 용어와 혼용하여 사용될 수 있다. 일반적인 경우, MxN 블록은 M개의 열과 N개의 행으로 이루어진 샘플들 또는 변환 계수들(transform coefficients)의 집합을 나타낼 수 있다. 샘플은 일반적으로 픽셀 또는 픽셀의 값을 나타낼 수 있으며, 휘도(luma) 성분의 픽셀/픽셀값만을 나타낼 수도 있고, 채도(chroma) 성분의 픽셀/픽셀 값을 나타낼 수도 있다. 샘플은 픽셀(pixel) 또는 펠(pel)에 대응하는 용어로서 사용될 수 있다.

[0090] 영상/비디오 인코더(300)는 입력 영상 신호(원본 블록, 원본 샘플 어레이)에서 인터 예측부(321) 또는 인트라 예측부(322)로부터 출력된 예측 신호(예측된 블록, 예측 샘플 어레이)를 감산하여 레지듀얼 신호(residual signal, 잔여 블록, 잔여 샘플 어레이)를 생성할 수 있고, 생성된 레지듀얼 신호는 변환부(332)로 전송된다. 이 경우, 도시된 바와 같이 영상/비디오 인코더(300) 내에서 입력 영상 신호(원본 블록, 원본 샘플 어레이)에서 예측 신호(예측 블록, 예측 샘플 어레이)를 감산하는 유닛은 감산부(331)라고 지칭될 수 있다. 예측부는 처리 대상 블록(이하, 현재 블록이라 함)에 대한 예측을 수행하고, 현재 블록에 대한 예측 샘플들을 포함하는 예측된 블록(predicted block)을 생성할 수 있다. 예측부는 현재 블록 또는 CU 단위로 인트라 예측이 적용되는지 또는 인터 예측이 적용되는지를 결정할 수 있다. 예측부는 예측 모드 정보 등 예측에 관한 다양한 정보를 생성하여 엔트로피 인코딩부(340)로 전달할 수 있다. 예측에 관한 정보는 엔트로피 인코딩부(340)에서 인코딩되어 비트스트림 형태로 출력될 수 있다.

[0091] 인트라 예측부(322)는 현재 픽처 내의 샘플들을 참조하여 현재 블록을 예측할 수 있다. 이 때, 참조되는 샘플들은 예측 모드에 따라 현재 블록의 주변(neighbor)에 위치할 수도 있고, 또는 떨어져서 위치할 수도 있다. 인트라 예측에서 예측 모드들은 복수의 비방향성 모드와 복수의 방향성 모드를 포함할 수 있다. 비방향성 모드는, 예를 들어 DC 모드 및 플래너 모드(Planar 모드)를 포함할 수 있다. 방향성 모드는 예측 방향의 세밀한 정도에 따라, 예를 들어 33개의 방향성 예측 모드 또는 65개의 방향성 예측 모드를 포함할 수 있다. 다만 이는 예시로서, 설정에 따라 그 이상 또는 그 이하의 개수의 방향성 예측 모드들이 사용될 수 있다. 인트라 예측부(322)는 주변 블록에 적용된 예측 모드를 이용하여, 현재 블록에 적용되는 예측 모드를 결정할 수도 있다.

[0092] 인터 예측부(321)는 참조 픽처 상에서 움직임 벡터에 의해 특정되는 참조 블록(참조 샘플 어레이)을 기반으로, 현재 블록에 대한 예측된 블록을 유도할 수 있다. 이 때, 인터 예측 모드에서 전송되는 움직임 정보의 양을 줄이기 위해 주변 블록과 현재 블록 간의 움직임 정보의 상관성에 기초하여 움직임 정보를 블록, 서브블록 또는 샘플 단위로 예측할 수 있다. 움직임 정보는 움직임 벡터 및 참조 픽처 인덱스를 포함할 수 있다. 움직임 정보는 인터 예측 방향(L0 예측, L1 예측, Bi 예측 등) 정보를 더 포함할 수 있다. 인터 예측의 경우에, 주변 블록은 현재 픽처 내에 존재하는 공간적 주변 블록(spatial neighboring block)과 참조 픽처에 존재하는 시간적 주변 블록(temporal neighboring block)을 포함할 수 있다. 참조 블록을 포함하는 참조 픽처와 시간적 주변 블록을 포함하는 참조 픽처는 동일할 수도 있고, 다를 수도 있다. 시간적 주변 블록은 동일 위치 참조 블록(collocated reference block), 동일 위치 CU(colCU) 등으로 지칭될 수 있으며, 시간적 주변 블록을 포함하는 참조 픽처는 동일 위치 픽처(collocated picture, colPic)로 지칭될 수도 있다. 예를 들어, 인터 예측부(321)는 주변 블록들을 기반으로 움직임 정보 후보 리스트를 구성하고, 현재 블록의 움직임 벡터 및/또는 참조 픽처 인덱스를 도출하기 위하여 어떤 후보가 사용되는지를 지시하는 정보를 생성할 수 있다. 다양한 예측 모드를 기반으로 인터 예측이 수행될 수 있으며, 예를 들어 스킵 모드와 머지 모드의 경우에, 인터 예측부(321)는 주변 블록의 움직임 정보를 현재 블록의 움직임 정보로 이용할 수 있다. 스킵 모드의 경우, 머지 모드와 달리 레지듀얼 신호가 전송되지 않을 수 있다. 움직임 정보 예측(motion vector prediction, MVP) 모드의 경우, 주변 블록의 움직임 벡터를 움직임 벡터 예측자(motion vector predictor)로 이용하고, 움직임 벡터 차분(motion vector

difference)을 시그널링함으로써 현재 블록의 움직임 벡터를 지시할 수 있다.

[0093] 예측부(320)는 다양한 예측 방법을 기반으로 예측 신호를 생성할 수 있다. 예를 들어, 예측부는 하나의 블록에 대한 예측을 위하여 인트라 예측 또는 인터 예측을 적용할 수 있을 뿐 아니라, 인트라 예측과 인터 예측을 동시에 적용할 수도 있다. 이를 combined inter and intra prediction (CIIP)라고 부를 수 있다. 또한, 예측부는 블록에 대한 예측을 위하여 인트라 블록 카피(intra block copy, IBC) 예측 모드에 기반할 수도 있고 또는 팔레트 모드(palette mode)에 기반할 수도 있다. IBC 예측 모드 또는 팔레트 모드는 예를 들어 SCC(screen content coding) 등과 같이 게임 등의 콘텐츠 영상/동영상 코딩을 위하여 사용될 수 있다. IBC는 기본적으로 현재 픽처 내에서 예측을 수행하나 현재 픽처 내에서 참조 블록을 도출하는 점에서 인터 예측과 유사하게 수행될 수 있다. 즉, IBC는 본 개시에서 설명되는 인터 예측 기법들 중 적어도 하나를 이용할 수 있다. 팔레트 모드는 인트라 코딩 또는 인트라 예측의 일 예로 볼 수 있다. 팔레트 모드가 적용되는 경우 팔레트 테이블 및 팔레트 인덱스에 관한 정보를 기반으로 픽처 내 샘플 값을 시그널링할 수 있다.

[0094] 예측부(320)에 의해 생성된 예측 신호는 복원 신호를 생성하기 위해 이용되거나 레지듀얼 신호를 생성하기 위해 이용될 수 있다. 변환부(332)는 레지듀얼 신호에 변환 기법을 적용하여 변환 계수들(transform coefficients)을 생성할 수 있다. 예를 들어, 변환 기법은 DCT(Discrete Cosine Transform), DST(Discrete Sine Transform), KLT(Karhunen-Loeve Transform), GBT(Graph-Based Transform), 또는 CNT(Conditionally Non-linear Transform) 중 적어도 하나를 포함할 수 있다. 여기서, GBT는 픽셀 간의 관계 정보를 그래프로 표현한다고 할 때 이 그래프로부터 얻어진 변환을 의미한다. CNT는 이전에 복원된 모든 픽셀들(all previously reconstructed pixels)을 이용하여 예측 신호를 생성하고 그에 기초하여 획득되는 변환을 의미한다. 또한, 변환 과정은 정사각형의 동일한 크기를 갖는 픽셀 블록에 적용될 수도 있고, 정사각형이 아닌 가변 크기의 블록에도 적용될 수 있다.

[0095] 양자화부(333)는 변환 계수들을 양자화하여 엔트로피 인코딩부(340)로 전송하고, 엔트로피 인코딩부(340)는 양자화된 신호(양자화된 변환 계수들에 관한 정보)를 인코딩하여 비트스트림으로 출력할 수 있다. 양자화된 변환 계수들에 관한 정보는 레지듀얼 정보라고 불릴 수 있다. 양자화부(333)는 계수 스캔 순서(scan order)를 기반으로 블록 형태의 양자화된 변환 계수들을 1차원 벡터 형태로 재정렬할 수 있고, 1차원 벡터 형태의 양자화된 변환 계수들을 기반으로 양자화된 변환 계수들에 관한 정보를 생성할 수도 있다. 엔트로피 인코딩부(340)는, 예를 들어 지수 곱셈(exponential Golomb), CAVLC(context-adaptive variable length coding), CABAC(context-adaptive binary arithmetic coding) 등과 같은 다양한 인코딩 방법을 수행할 수 있다. 엔트로피 인코딩부(340)는 양자화된 변환 계수들 외 영상/비디오 복원에 필요한 정보들(e.g., 선택 요소들(syntax elements)의 값 등)을 함께 또는 별도로 인코딩할 수도 있다. 인코딩된 정보(e.g., 인코딩된 영상/비디오 정보)는 비트스트림 형태로 NAL(network abstraction layer) 유닛 단위로 전송 또는 저장될 수 있다. 영상/비디오 정보는 어댑테이션 파라미터 세트(APS), 픽처 파라미터 세트(PPS), 시퀀스 파라미터 세트(SPS) 또는 비디오 파라미터 세트(VPS) 등 다양한 파라미터 세트에 관한 정보를 더 포함할 수 있다. 또한, 영상/비디오 정보는 일반 제한 정보(general constraint information)를 더 포함할 수 있다. 또한, 영상/비디오 정보는 부호화된 정보의 생성 방법 및 사용 방법, 목적 등을 더 포함할 수 있다. 본 개시에서 영상/비디오 인코더에서 영상/비디오 디코더로 전달/시그널링되는 정보 및/또는 선택 요소들은 영상/비디오 정보에 포함될 수 있다. 영상/비디오 정보는 상술한 인코딩 절차를 통하여 인코딩되어 비트스트림에 포함될 수 있다. 비트스트림은 네트워크를 통하여 전송될 수 있고, 또는 디지털 저장매체에 저장될 수 있다. 여기서, 네트워크는 방송망 및/또는 통신망 등을 포함할 수 있고, 디지털 저장매체는 USB, SD, CD, DVD, 블루레이, HDD, SSD 등 다양한 저장매체를 포함할 수 있다. 엔트로피 인코딩부(340)로부터 출력된 신호를 전송하는 전송부(미도시) 및/또는 저장하는 저장부(미도시)가 영상/비디오 인코더(300)의 내/외부 엘리먼트로서 구성될 수 있고, 또는 전송부는 엔트로피 인코딩부(340)에 포함될 수도 있다.

[0096] 양자화부(333)로부터 출력된 양자화된 변환 계수들은 예측 신호를 생성하기 위해 이용될 수 있다. 예를 들어, 양자화된 변환 계수들에 역양자화부(334) 및 역변환부(335)를 통해 역양자화 및 역변환을 적용함으로써 레지듀얼 신호(레지듀얼 블록 or 레지듀얼 샘플들)를 복원할 수 있다. 가산부(350)는 복원된 레지듀얼 신호를 인터 예측부(321) 또는 인트라 예측부(322)로부터 출력된 예측 신호에 더함으로써 복원(reconstructed) 신호(복원 픽처, 복원 블록, 복원 샘플 어레이)가 생성될 수 있다. 스킵 모드가 적용된 경우와 같이 처리 대상 블록에 대한 레지듀얼이 없는 경우, 예측된 블록이 복원 블록으로 사용될 수 있다. 가산부(350)는 복원부 또는 복원 블록 생성부라고 불릴 수 있다. 생성된 복원 신호는 현재 픽처 내 다음 처리 대상 블록의 인트라 예측을 위하여 사용될 수 있고, 후술하는 바와 같이 필터링을 거쳐서 다음 픽처의 인터 예측을 위하여 사용될 수도 있다.

- [0097] 한편 픽처 인코딩 및/또는 복원 과정에서 LMCS (luma mapping with chroma scaling)가 적용될 수도 있다.
- [0098] 필터링부(360)는 복원 신호에 필터링을 적용하여 주관적/객관적 화질을 향상시킬 수 있다. 예를 들어, 필터링부(360)은 복원 픽처에 다양한 필터링 방법을 적용하여 수정된(modified) 복원 픽처를 생성할 수 있고, 수정된 복원 픽처를 메모리(370), 구체적으로 메모리(370)의 DPB에 저장할 수 있다. 다양한 필터링 방법은, 예를 들어 디블록킹 필터링, 샘플 적응적 오프셋(sample adaptive offset), 적응적 루프 필터(adaptive loop filter), 양방향 필터(bilateral filter) 등을 포함할 수 있다. 필터링부(360)은 필터링에 관한 다양한 정보를 생성하여 엔트로피 인코딩부(340)로 전달할 수 있다. 필터링 관한 정보는 엔트로피 인코딩부(340)에서 인코딩되어 비트스트림 형태로 출력될 수 있다.
- [0099] 메모리(370)에 전송된 수정된 복원 픽처는 인터 예측부(321)에서 참조 픽처로 사용될 수 있다. 이를 통해, 인코더 단 및 디코더 단에서의 예측 미스매치를 피할 수 있고, 부호화 효율도 향상시킬 수 있다.
- [0100] 메모리(370)의 DPB는 수정된 복원 픽처를 인터 예측부(321)에서의 참조 픽처로 사용하기 위해 저장할 수 있다. 메모리(370)는 현재 픽처 내 움직임 정보가 도출된(또는, 인코딩된) 블록의 움직임 정보 및/또는 이미 복원된 픽처 내 블록들의 움직임 정보를 저장할 수 있다. 저장된 움직임 정보는 공간적 주변 블록의 움직임 정보 또는 시간적 주변 블록의 움직임 정보로 활용하기 위하여 인터 예측부(321)에 전달될 수 있다. 메모리(370)는 현재 픽처 내 복원된 블록들의 복원 샘플들을 저장할 수 있고, 저장된 복원 샘플들을 인트라 예측부(322)에 전달할 수 있다.
- [0101] 한편, VCM 인코더(또는, 피쳐/피쳐맵 인코더)는, 피쳐/피쳐맵을 인코딩하기 위하여 예측, 변환, 양자화 등 일련의 절차를 수행한다는 점에서, 기본적으로 도 3을 참조하여 설명한 영상/비디오 인코더(300)와 동일/유사한 구조를 가질 수 있다. 다만, VCM 인코더는 피쳐/피쳐맵을 부호화 대상으로 한다는 점에서 영상/비디오 인코더(300)와는 차이가 있으며, 이에 따라 각 유닛(또는, 구성요소)의 명칭(e.g., 영상 분할부(310) 등)과 그 구체적 동작 내용에 있어서 영상/비디오 인코더(300)와는 상이할 수 있다. VCM 인코더의 구체적 동작 내용에 대해서는 상세히 후술하기로 한다.
- [0102] **디코더(Decoder)**
- [0103] 도 4는 본 개시의 실시예들이 적용될 수 있는 영상/비디오 디코더를 개략적으로 나타낸 도면이다.
- [0104] 도 4를 참조하면, 영상/비디오 디코더(400)는 엔트로피 디코딩부(entropy decoder, 410), 레지듀얼 처리부(residual processor, 420), 예측부(predictor, 430), 가산부(adder, 440), 필터링부(filter, 450) 및 메모리(memoery, 460)를 포함할 수 있다. 예측부(430)는 인터 예측부(431) 및 인트라 예측부(432)를 포함할 수 있다. 레지듀얼 처리부(420)는 역양자화부(dequantizer, 421) 및 역변환부(inverse transformer, 422)를 포함할 수 있다. 상술한 엔트로피 디코딩부(410), 레지듀얼 처리부(420), 예측부(430), 가산부(440) 및 필터링부(450)는 실시예에 따라 하나의 하드웨어 컴포넌트(예컨대, 디코더 칩셋 또는 프로세서)에 의하여 구성될 수 있다. 또한 메모리(460)는 DPB(decoded picture buffer)를 포함할 수 있고, 디지털 저장 매체에 의하여 구성될 수도 있다. 하드웨어 컴포넌트는 메모리(460)을 내/외부 컴포넌트로 더 포함할 수도 있다.
- [0105] 비디오/영상 정보를 포함하는 비트스트림이 입력되면, 영상/비디오 디코더(400)는 도 3의 영상/비디오 인코더(300)에서 영상/비디오 정보가 처리된 프로세스에 대응하여 영상/비디오를 복원할 수 있다. 예를 들어, 영상/비디오 디코더(400)는 비트스트림으로부터 획득한 블록 분할 관련 정보를 기반으로 유닛들/블록들을 도출할 수 있다. 영상/비디오 디코더(400)는 영상/비디오 인코더에서 적용된 프로세싱 유닛을 이용하여 디코딩을 수행할 수 있다. 따라서, 디코딩의 프로세싱 유닛은, 예컨대 코딩 유닛일 수 있고, 코딩 유닛은 코딩 트리 유닛 또는 최대 코딩 유닛으로부터 쿼드 트리 구조, 바이너리 트리 구조 및/또는 터너리 트리 구조를 따라서 분할될 수 있다. 코딩 유닛으로부터 하나 이상의 변환 유닛이 도출될 수 있다. 그리고, 영상/비디오 디코더(400)를 통해 디코딩 및 출력된 복원 영상 신호는 재생 장치를 통해 재생될 수 있다.
- [0106] 영상/비디오 디코더(400)는 도 3의 인코더로부터 출력된 신호를 비트스트림 형태로 수신할 수 있고, 수신된 신호는 엔트로피 디코딩부(410)를 통해 디코딩될 수 있다. 예를 들어, 엔트로피 디코딩부(410)는 비트스트림을 파싱하여 영상 복원(또는, 픽처 복원)에 필요한 정보(e.g., 영상/비디오 정보)를 도출할 수 있다. 영상/비디오 정보는 어댑테이션 파라미터 세트(APS), 픽처 파라미터 세트(PPS), 시퀀스 파라미터 세트(SPS) 또는 비디오 파라미터 세트(VPS) 등 다양한 파라미터 세트에 관한 정보를 더 포함할 수 있다. 또한, 영상/비디오 정보는 일반 제한 정보(general constraint information)를 더 포함할 수 있다. 또한, 영상/비디오 정보는 복호화된 정보의 생성 방법 및 사용 방법, 목적 등을 포함할 수 있다. 영상/비디오 디코더(400)는 파라미터 세트에 관한 정보

및/또는 일반 제한 정보를 더 기반으로 픽처를 디코딩할 수 있다. 시그널링/수신되는 정보 및/또는 선택스 요소들은 디코딩 절차를 통하여 디코딩되어 비트스트림으로부터 획득될 수 있다. 예를 들어, 엔트로피 디코딩부(410)는 지수 곱셈 부호화, CAVLC 또는 CABAC 등의 코딩 방법을 기초로 비트스트림 내 정보를 디코딩하고, 영상 복원에 필요한 선택스 요소들의 값들, 레지듀얼에 관한 변환 계수의 양자화된 값들을 출력할 수 있다. 보다 상세하게, CABAC 엔트로피 디코딩 방법은, 비트스트림에서 각 구문 요소에 해당하는 빈(bin)을 수신하고, 디코딩 대상 구문 요소 정보와 주변 및 디코딩 대상 블록의 디코딩 정보 혹은 이전 단계에서 디코딩된 심볼/빈의 정보를 이용하여 문맥(context) 모델을 결정하고, 결정된 문맥 모델에 따라 빈의 발생 확률을 예측하여 빈의 산술 디코딩(arithmetic decoding)를 수행하여 각 선택스 요소의 값에 해당하는 심볼을 생성할 수 있다. 이 때, CABAC 엔트로피 디코딩 방법은 문맥 모델 결정 후 다음 심볼/빈의 문맥 모델을 위해 디코딩된 심볼/빈의 정보를 이용하여 문맥 모델을 업데이트할 수 있다. 엔트로피 디코딩부(410)에서 디코딩된 정보 중 예측에 관한 정보는 예측부(인터 예측부(432) 및 인트라 예측부(431))로 제공되고, 엔트로피 디코딩부(410)에서 엔트로피 디코딩이 수행된 레지듀얼 값, 즉 양자화된 변환 계수들 및 관련 파라미터 정보는 레지듀얼 처리부(420)로 입력될 수 있다. 레지듀얼 처리부(420)는 레지듀얼 신호(레지듀얼 블록, 레지듀얼 샘플들, 레지듀얼 샘플 어레이)를 도출할 수 있다. 또한, 엔트로피 디코딩부(410)에서 디코딩된 정보 중 필터링에 관한 정보는 필터링부(450)로 제공될 수 있다. 한편, 영상/비디오 인코더로부터 출력된 신호를 수신하는 수신부(미도시)가 영상/비디오 디코더(400)의 내/외부 엘리먼트로서 더 구성될 수 있고, 또는 수신부는 엔트로피 디코딩부(410)의 구성요소일 수도 있다. 한편, 본 개시에 따른 영상/비디오 디코더는 영상/비디오 디코딩 장치라고 불릴 수 있고, 영상/비디오 디코더는 정보 디코더(영상/비디오 정보 디코더) 및 샘플 디코더(영상/비디오 샘플 디코더)로 구분될 수도 있다. 이 경우, 정보 디코더는 엔트로피 디코딩부(410)를 포함할 수 있고, 샘플 디코더는 역양자화부(321), 역변환부(322), 가산부(440), 필터링부(450), 메모리(460), 인터 예측부(432) 및 인트라 예측부(431) 중 적어도 하나를 포함할 수 있다.

[0107] 역양자화부(421)에서는 양자화된 변환 계수들을 역양자화하여 변환 계수들을 출력할 수 있다. 역양자화부(421)는 양자화된 변환 계수들을 2차원의 블록 형태로 재정렬할 수 있다. 이 경우, 재정렬은 영상/비디오 인코더에서 수행된 계수 스캔 순서를 기반으로 재정렬을 수행할 수 있다. 역양자화부(321)는 양자화 파라미터(e.g., 양자화 스텝 사이즈 정보)를 이용하여 양자화된 변환 계수들에 대한 역양자화를 수행하고, 변환 계수들(transform coefficients)을 획득할 수 있다.

[0108] 역변환부(422)에서는 변환 계수들을 역변환하여 레지듀얼 신호(레지듀얼 블록, 레지듀얼 샘플 어레이)를 획득하게 된다.

[0109] 예측부(430)는 현재 블록에 대한 예측을 수행하고, 현재 블록에 대한 예측 샘플들을 포함하는 예측된 블록(predicted block)을 생성할 수 있다. 예측부는 엔트로피 디코딩부(410)로부터 출력된 예측에 관한 정보를 기반으로 현재 블록에 인트라 예측이 적용되는지 또는 인터 예측이 적용되는지 결정할 수 있고, 구체적인 인트라/인터 예측 모드를 결정할 수 있다.

[0110] 예측부(420)는 다양한 예측 방법을 기반으로 예측 신호를 생성할 수 있다. 예를 들어, 예측부는 하나의 블록에 대한 예측을 위하여 인트라 예측 또는 인터 예측을 적용할 수 있을 뿐 아니라, 인트라 예측과 인터 예측을 동시에 적용할 수 있다. 이를 combined inter and intra prediction (CIIP)라고 부를 수 있다. 또한, 예측부는 블록에 대한 예측을 위하여 인트라 블록 카피(intra block copy, IBC) 예측 모드에 기반할 수도 있고 또는 팔레트 모드(palette mode)에 기반할 수도 있다. IBC 예측 모드 또는 팔레트 모드는 예를 들어 SCC(screen content coding) 등과 같이 게임 등의 콘텐츠 영상/동영상 코딩을 위하여 사용될 수 있다. IBC는 기본적으로 현재 픽처 내에서 예측을 수행하나 현재 픽처 내에서 참조 블록을 도출하는 점에서 인터 예측과 유사하게 수행될 수 있다. 즉, IBC는 본 문서에서 설명되는 인터 예측 기법들 중 적어도 하나를 이용할 수 있다. 팔레트 모드는 인트라 코딩 또는 인트라 예측의 일 예로 볼 수 있다. 팔레트 모드가 적용되는 경우 팔레트 테이블 및 팔레트 인덱스에 관한 정보가 영상/비디오 정보에 포함되어 시그널링될 수 있다.

[0111] 인트라 예측부(431)는 현재 픽처 내의 샘플들을 참조하여 현재 블록을 예측할 수 있다. 참조되는 샘플들은 예측 모드에 따라 현재 블록의 주변(neighbor)에 위치할 수 있고, 또는 떨어져서 위치할 수도 있다. 인트라 예측에서 예측 모드들은 복수의 비방향성 모드와 복수의 방향성 모드를 포함할 수 있다. 인트라 예측부(431)는 주변 블록에 적용된 예측 모드를 이용하여, 현재 블록에 적용되는 예측 모드를 결정할 수도 있다.

[0112] 인터 예측부(432)는 참조 픽처 상에서 움직임 벡터에 의해 특정되는 참조 블록(참조 샘플 어레이)을 기반으로, 현재 블록에 대한 예측된 블록을 유도할 수 있다. 이 때, 인터 예측 모드에서 전송되는 움직임 정보의 양을 줄

이기 위해 주변 블록과 현재 블록 간의 움직임 정보의 상관성에 기초하여 움직임 정보를 블록, 서브블록 또는 샘플 단위로 예측할 수 있다. 움직임 정보는 움직임 벡터 및 참조 픽처 인덱스를 포함할 수 있다. 움직임 정보는 인터 예측 방향(L0 예측, L1 예측, Bi 예측 등) 정보를 더 포함할 수 있다. 인터 예측의 경우에, 주변 블록은 현재 픽처 내에 존재하는 공간적 주변 블록(spatial neighboring block)과 참조 픽처에 존재하는 시간적 주변 블록(temporal neighboring block)을 포함할 수 있다. 예를 들어, 인터 예측부(432)는 주변 블록들을 기반으로 움직임 정보 후보 리스트를 구성하고, 수신한 후보 선택 정보를 기반으로 현재 블록의 움직임 벡터 및/또는 참조 픽처 인덱스를 도출할 수 있다. 다양한 예측 모드를 기반으로 인터 예측이 수행될 수 있으며, 예측에 관한 정보는 현재 블록에 대한 인터 예측의 모드를 지시하는 정보를 포함할 수 있다.

[0113] 가산부(440)는 획득된 레지듀얼 신호를 예측부(인터 예측부(432) 및/또는 인트라 예측부(431) 포함)로부터 출력된 예측 신호(예측된 블록, 예측 샘플 어레이)에 더함으로써 복원 신호(복원 픽처, 복원 블록, 복원 샘플 어레이)를 생성할 수 있다. 스킵 모드가 적용된 경우와 같이 처리 대상 블록에 대한 레지듀얼이 없는 경우, 예측된 블록이 복원 블록으로 사용될 수 있다.

[0114] 가산부(440)는 복원부 또는 복원 블록 생성부라고 불릴 수 있다. 생성된 복원 신호는 현재 픽처 내 다음 처리 대상 블록의 인트라 예측을 위하여 사용될 수 있고, 후술하는 바와 같이 필터링을 거쳐서 출력될 수도 있고 또는 다음 픽처의 인터 예측을 위하여 사용될 수도 있다.

[0115] 한편, 픽처 디코딩 과정에서 LMCS (luma mapping with chroma scaling)가 적용될 수도 있다.

[0116] 필터링부(450)는 복원 신호에 필터링을 적용하여 주관적/객관적 화질을 향상시킬 수 있다. 예를 들어, 필터링부(450)는 복원 픽처에 다양한 필터링 방법을 적용하여 수정된(modified) 복원 픽처를 생성할 수 있고, 수정된 복원 픽처를 메모리(460), 구체적으로 메모리(460)의 DPB에 전송할 수 있다. 다양한 필터링 방법은, 예컨대 디블록킹 필터링, 샘플 적응적 오프셋(sample adaptive offset), 적응적 루프 필터(adaptive loop filter), 양방향 필터(bilateral filter) 등을 포함할 수 있다.

[0117] 메모리(460)의 DPB에 저장된 (수정된) 복원 픽처는 인터 예측부(432)에서 참조 픽처로 사용될 수 있다. 메모리(460)는 현재 픽처 내 움직임 정보가 도출된(또는 디코딩된) 블록의 움직임 정보 및/또는 이미 복원된 픽처 내 블록들의 움직임 정보를 저장할 수 있다. 저장된 움직임 정보는 공간적 주변 블록의 움직임 정보 또는 시간적 주변 블록의 움직임 정보로 활용하기 위하여 인터 예측부(432)에 전달할 수 있다. 메모리(460)는 현재 픽처 내 복원된 블록들의 복원 샘플들을 저장할 수 있고, 인트라 예측부(431)에 전달할 수 있다.

[0118] 한편, VCM 디코더(또는, 피쳐/피쳐맵 디코더)는, 피쳐/피쳐맵을 디코딩하기 위하여 예측, 역변환, 역양자화 등 일련의 절차를 수행한다는 점에서, 기본적으로 도 4를 참조하여 상술한 영상/비디오 디코더(400)와 동일/유사한 구조를 가질 수 있다. 다만, VCM 디코더는 피쳐/피쳐맵을 복호화 대상으로 한다는 점에서 영상/비디오 디코더(400)와는 차이가 있으며, 이에 따라 각 유닛(또는, 구성요소)의 명칭(e.g., DPB 등)과 그 구체적 동작 내용에 있어서 영상/비디오 디코더(400)와는 상이할 수 있다. VCM 디코더의 동작은 VCM 인코더의 동작에 대응할 수 있으며, 그 구체적 동작 내용에 대해서는 상세히 후술하기로 한다.

[0119] **피쳐/피쳐맵 인코딩 절차**

[0120] 도 5는 본 개시의 실시예들이 적용될 수 있는 피쳐/피쳐맵 인코딩 절차를 개략적으로 나타낸 흐름도이다.

[0121] 도 5를 참조하면, 피쳐/피쳐맵 인코딩 절차는 예측 절차(S510), 레지듀얼 처리 절차(S520) 및 정보 인코딩 절차(S530)를 포함할 수 있다.

[0122] 예측 절차(S510)는 도 3을 참조하여 전술한 예측부(320)에 의해 수행될 수 있다.

[0123] 구체적으로, 인트라 예측부(322)는 현재 피쳐/피쳐맵 내의 피쳐 엘리먼트들을 참조하여 현재 블록(즉, 현재 부호화 대상이 되는 피쳐 엘리먼트들의 집합)을 예측할 수 있다. 인트라 예측은 피쳐/피쳐맵을 구성하는 피쳐 엘리먼트들의 공간적 유사성에 기반하여 수행될 수 있다. 예를 들어, 영상/비디오 내에서 동일한 관심 영역(Region of Interest, RoI)에 포함된 피쳐 엘리먼트들은 유사한 데이터 분포 특성을 갖는 것으로 추정될 수 있다. 따라서, 인트라 예측부(322)는 현재 블록을 포함하는 관심 영역 내에서 기복원된 피쳐 엘리먼트들을 참조하여 현재 블록을 예측할 수 있다. 이 때, 참조되는 피쳐 엘리먼트들은 예측 모드에 따라 현재 블록과 인접하여 위치할 수도 있고, 또는 현재 블록과 이격하여 위치할 수도 있다. 피쳐/피쳐맵 부호화를 위한 인트라 예측 모드들은 복수의 비방향성 예측 모드들 및 복수의 방향성 예측 모드들을 포함할 수 있다. 비방향성 예측 모드들은, 예컨대 영상/비디오 인코딩 절차의 DC 모드 및 플래너 모드에 대응하는 예측 모드들을 포함할 수 있다. 또한,

방향성 모드들은, 예컨대 영상/비디오 인코딩 절차의 33개의 방향성 모드들 또는 65개의 방향성 모드들에 대응하는 예측 모드들을 포함할 수 있다. 다만 이는 예시로서, 인트라 예측 모드들의 유형 및 개수는 실시예에 따라 다양하게 설정/변경될 수 있다.

[0124] 인트라 예측부(321)는 참조 피쳐/피쳐맵 상에서 움직임 정보에 의해 특정되는 참조 블록(즉, 참조되는 피쳐 엘리먼트들의 집합)을 기반으로, 현재 블록을 예측할 수 있다. 인트라 예측은 피쳐/피쳐맵을 구성하는 피쳐 엘리먼트들의 시간적 유사성에 기반하여 수행될 수 있다. 예를 들어, 시간적으로 연속된 피쳐들은 유사한 데이터 분포 특성을 가질 수 있다. 따라서, 인트라 예측부(321)는 현재 피쳐와 시간적으로 인접한 피쳐의 기록된 피쳐 엘리먼트들을 참조하여 현재 블록을 예측할 수 있다. 이 때, 참조되는 피쳐 엘리먼트들을 특정하기 위한 움직임 정보는 움직임 벡터 및 참조 피쳐/피쳐맵 인덱스를 포함할 수 있다. 움직임 정보는 인트라 예측 방향(e.g., L0 예측, L1 예측, Bi 예측 등)에 관한 정보를 더 포함할 수 있다. 인트라 예측의 경우에, 주변 블록은 현재 피쳐/피쳐맵 내에 존재하는 공간적 주변 블록(spatial neighboring block)과 참조 피쳐/피쳐맵 내에 존재하는 시간적 주변 블록(temporal neighboring block)을 포함할 수 있다. 참조 블록을 포함하는 참조 피쳐/피쳐맵과 시간적 주변 블록을 포함하는 참조 피쳐/피쳐맵은 동일할 수도 있고, 다를 수도 있다. 시간적 주변 블록은 동일 위치 참조 블록(collocated reference block) 등으로 지칭될 수 있으며, 시간적 주변 블록을 포함하는 참조 피쳐/피쳐맵은 동일 위치 피쳐/피쳐맵(collocated feature/feature map)으로 지칭될 수도 있다. 인트라 예측부(321)는 주변 블록들을 기반으로 움직임 정보 후보 리스트를 구성하고, 현재 블록의 움직임 벡터 및/또는 참조 피쳐/피쳐맵 인덱스를 도출하기 위하여 어떤 후보가 사용되는지를 지시하는 정보를 생성할 수 있다. 다양한 예측 모드들을 기반으로 인트라 예측이 수행될 수 있으며, 예를 들어 스킵 모드와 머지 모드의 경우에, 인트라 예측부(321)는 주변 블록의 움직임 정보를 현재 블록의 움직임 정보로 이용할 수 있다. 스킵 모드의 경우, 머지 모드와 달리 레지듀얼 신호가 전송되지 않을 수 있다. 움직임 정보 예측(motion vector prediction, MVP) 모드의 경우, 주변 블록의 움직임 벡터를 움직임 벡터 예측자(motion vector predictor)로 이용하고, 움직임 벡터 차분(motion vector difference)을 시그널링함으로써 현재 블록의 움직임 벡터를 지시할 수 있다. 예측부(320)는 상술한 인트라 예측 및 인트라 예측 이외에도 다양한 예측 방법에 기반하여 예측 신호를 생성할 수 있다.

[0125] 예측부(320)에 의해 생성된 예측 신호는 레지듀얼 신호(레지듀얼 블록, 레지듀얼 피쳐 엘리먼트들)를 생성하기 위해 이용될 수 있다(S520). 레지듀얼 처리 절차(S520)는 도 3을 참조하여 전송한 레지듀얼 처리부(330)에 의해 수행될 수 있다. 그리고, 레지듀얼 신호에 대한 변환 및/또는 양자화 절차를 통해 (양자화된) 변환 계수들이 생성될 수 있으며, 엔트로피 인코딩부(340)는 (양자화된) 변환 계수들에 관한 정보를 레지듀얼 정보로서 비트스트림 내에 인코딩할 수 있다(S530). 또한, 엔트로피 인코딩부(340)는 레지듀얼 정보 외에 피쳐/피쳐맵 복원에 필요한 정보, 예컨대 예측 정보(e.g., 예측 모드 정보, 움직임 정보 등)를 비트스트림 내에 인코딩할 수 있다.

[0126] 한편, 피쳐/피쳐맵 인코딩 절차는 피쳐/피쳐맵 복원을 위한 정보(e.g., 예측 정보, 레지듀얼 정보, 파티셔닝 정보 등)를 인코딩하여 비트스트림 형태로 출력하는 절차(S530)뿐만 아니라, 현재 피쳐/피쳐맵에 대한 복원 피쳐/피쳐맵을 생성하는 절차 및 복원 피쳐/피쳐맵에 대해 인루프 필터링을 적용하는 절차(optional)를 더 포함할 수 있다.

[0127] VCM 인코더는 역양자화 및 역변환을 통해 양자화된 변환 계수(들)로부터 (수정된) 레지듀얼 피쳐(들)을 도출할 수 있으며, 단계 S510의 출력인 예측 피쳐(들)과 (수정된) 레지듀얼 피쳐(들)을 기반으로 복원 피쳐/피쳐맵을 생성할 수 있다. 이렇게 생성된 복원 피쳐/피쳐맵은 VCM 디코더에서 생성된 복원 피쳐/피쳐맵과 동일할 수 있다. 복원 피쳐/피쳐맵에 대해 인루프 필터링 절차가 수행되는 경우, 복원 피쳐/피쳐맵에 대한 인루프 필터링 절차를 통해 수정된 복원 피쳐/피쳐맵이 생성될 수 있다. 수정된 복원 피쳐/피쳐맵은 복호 피쳐 버퍼(decoded feature buffer, DFB) 또는 메모리에 저장되어, 이후 피쳐/피쳐맵의 예측 절차에서 참조 피쳐/피쳐맵으로 이용될 수 있다. 또한, (인루프) 필터링 관련 정보(파라미터)가 인코딩되어 비트스트림 형태로 출력될 수 있다. 인루프 필터링 절차를 통해 피쳐/피쳐맵 코딩시 발생할 수 있는 노이즈를 제거할 수 있으며, 피쳐/피쳐맵 기반의 태스크 수행 성능을 향상시킬 수 있다. 또한, 인코더단 및 디코더단 모두에서 인루프 필터링 절차를 수행함으로써, 예측 결과의 동일성을 보장하고 피쳐/피쳐맵 코딩의 신뢰성을 향상시킬 수 있으며, 피쳐/피쳐맵 코딩을 위한 데이터 전송량을 감소시킬 수 있다.

[0128] **피쳐/피쳐맵 디코딩 절차**

[0129] 도 6은 본 개시의 실시예들이 적용될 수 있는 피쳐/피쳐맵 디코딩 절차를 개략적으로 나타낸 흐름도이다.

[0130] 도 6을 참조하면, 피쳐/피쳐맵 디코딩 절차는 영상/비디오 정보 획득 절차(S610), 피쳐/피쳐맵 복원 절차(S620-S640) 및 복원된 피쳐/피쳐맵에 대한 인루프 필터링 절차(S650)를 포함할 수 있다. 피쳐/피쳐맵 복원 절

차는 본 개시에서 설명된 인터/인트라 예측(S620) 및 레지듀얼 처리(S630), 양자화된 변환 계수에 대한 역양자화, 역변환) 과정을 통해 획득되는 예측 신호 및 레지듀얼 신호에 기반하여 수행될 수 있다. 복원 피쳐/피쳐맵에 대한 인루프 필터링 절차를 통해 수정된(modified) 복원 피쳐/피쳐맵이 생성될 수 있으며, 수정된 복원 피쳐/피쳐맵은 디코딩된 피쳐/피쳐맵으로서 출력될 수 있다. 디코딩된 피쳐/피쳐맵은 복호 피쳐 버퍼(DFB) 또는 메모리에 저장되어 이후 피쳐/피쳐맵의 디코딩시 인터 예측 절차에서 참조 피쳐/피쳐맵으로 사용될 수 있다. 경우에 따라, 상술한 인루프 필터링 절차는 생략될 수 있다. 이 경우, 복원 피쳐/피쳐맵이 디코딩된 피쳐/피쳐맵으로서 그대로 출력될 수 있고, 복호 피쳐 버퍼(DFB) 또는 메모리에 저장되어 이후 피쳐/피쳐맵의 디코딩시 인터 예측 절차에서 참조 피쳐/피쳐맵으로 사용될 수 있다.

[0131] **피쳐 추출 방법 및 데이터 분포 특성**

[0132] 본 개시의 실시예들은 심층 신경망의 은닉층에서 생성된 activation(feature) 맵을 압축하는 데 따른 정보량 감소를 보상하기 위한 필터링 과정 및 관련 비트스트림을 생성하는 방법을 제안한다.

[0133] 심층 신경망으로 입력된 입력 데이터는 여러 은닉 계층의 연산 과정을 거치며, 각 은닉 계층의 연산 결과는 사용 중인 심층 신경망의 종류와 해당 심층 신경망 내 은닉 계층의 위치에 따라 다양한 크기와 채널 수를 갖는 피쳐/피쳐맵으로 출력된다.

[0134] 도 7은 본 개시의 실시예들이 적용될 수 있는 피쳐 추출 및 복원 방법의 일 예를 나타내는 도면이다.

[0135] 도 7을 참조하면, 피쳐 추출 네트워크(710)는 심층 신경망의 중간 계층 activation(feature) 맵을 소스 영상/비디오(Image/Video)로부터 추출하여 추출된 피쳐맵을 출력할 수 있다. 피쳐 추출 네트워크(710)는 심층 신경망의 입력으로부터 연속되는 은닉 계층들의 집합일 수 있다.

[0136] 인코딩 장치(720)는 출력된 피쳐맵을 압축하여 비트스트림 형태로 출력할 수 있으며, 디코딩 장치(730)는 출력된 비트스트림으로부터 (압축된) 피쳐맵을 복원할 수 있다. 인코딩 장치(720)는 도 1의 부호화부(12)에 대응될 수 있으며, 디코딩 장치(720)는 도 1의 복호화부(22)에 대응될 수 있다. 테스트 네트워크(740)는 복원된 피쳐맵에 기반하여 테스트를 수행할 수 있다.

[0137] 도 8은 피쳐 추출 네트워크(710)를 이용한 피쳐 추출 방법의 일 예를 나타낸 도면이다.

[0138] 도 8을 참조하면, 피쳐 추출 네트워크(710)는 비디오 소스(Image/Video)를 입력받아 피쳐 추출 동작을 수행함으로써 비디오 소스의 피쳐셋(Feature set)을 출력할 수 있다. 피쳐셋은 비디오 소스로부터 추출된 복수의 피쳐들(C_0, C_1, \dots, C_n)을 포함할 수 있으며, 피쳐맵으로 표현될 수 있다. 각각의 피쳐(C_0, C_1, \dots, C_n)는 복수의 특징 엘리먼트들을 포함하며, 서로 다른 데이터 분포 특성을 가질 수 있다.

[0139] 도 8에서, W, H 및 C 는 각각 비디오 소스의 너비, 높이 및 채널 개수를 의미할 수 있다. 여기서, 비디오 소스의 채널 개수(C)는 비디오 소스의 영상 포맷에 기반하여 결정될 수 있다. 예를 들어, 비디오 소스가 RGB 영상 포맷을 갖는 경우, 비디오 소스의 채널 개수(C)는 3일 수 있다.

[0140] 또한, W', H' 및 C' 은 각각 피쳐셋의 너비, 높이 및 채널 개수를 의미할 수 있다. 피쳐셋의 채널 개수(C')는 비디오 소스로부터 추출된 피쳐들(C_0, C_1, \dots, C_n)의 총 개수($n+1$)와 같을 수 있다. 일 예에서, 피쳐셋의 채널 개수(C')는 비디오 소스의 채널 개수(C)보다 클 수 있다.

[0141] 피쳐셋의 속성(W', H', C')은 비디오 소스의 속성(W, H, C)에 따라 달라질 수 있다. 예를 들어, 비디오 소스의 채널 개수(C)가 증가함에 따라 피쳐셋의 채널 개수(C')도 함께 증가할 수 있다. 또한, 피쳐셋의 속성(W', H', C')은 피쳐 추출 네트워크(710)의 종류 및 속성에 따라 달라질 수 있다. 예를 들어, 피쳐 추출 네트워크(710)가 인공 신경망(e.g., CNN, DNN 등)으로 구현되는 경우, 각각의 피쳐(C_0, C_1, \dots, C_n)를 출력하는 레이어의 위치에 따라 피쳐셋의 속성(W', H', C') 또한 달라질 수 있다.

[0142] 비디오 소스 및 피쳐셋은 서로 다른 데이터 분포 특성을 가질 수 있다. 예를 들어, 비디오 소스는 일반적으로 한 개(Grayscale image) 채널 또는 세 개(RGB image) 채널들로 구성될 수 있다. 비디오 소스에 포함된 픽셀들은 모든 채널들에 대하여 동일한 정수값 범위를 가질 수 있으며, 음이 아닌 값을 가질 수 있다. 또한, 각각의 픽셀 값은 소정의 정수값 범위 내에서 고르게 분포할 수 있다. 이에 반해, 피쳐셋은 피쳐 추출 네트워크(710)의 유형(e.g., CNN, DNN 등) 및 레이어 위치에 따라 다양한 개수(e.g., 32, 64, 128, 256, 512 등)의 채널들로 구성될 수 있다. 피쳐셋에 포함된 피쳐 엘리먼트들은 각 채널별로 서로 다른 실수값 범위를 가질 수 있으며, 음의 값을 가질 수도 있다. 또한, 각각의 피쳐 엘리먼트값은 소정의 실수값 범위 내에서 특정 영역에 집중적으로 분포할

수 있다.

- [0143] 도 9는 추출된 피처의 일 예를 나타내는 도면이며, 도 10은 추출된 피처 셋의 크기에 대한 일 예를 나타내는 도면이다.
- [0144] 도 9의 좌측은 비디오 소스(Image/Video)를 나타내며, 도 9의 우측은 64개의 피처들을 갖는 피처맵을 나타낸다. 도 9를 참조하면, 하나의 비디오 소스에 대해 64개의 피처(채널)들이 추출된 것을 볼 수 있다. 또한, 각 피처의 너비와 높이의 크기 또한 비디오 소스의 너비 및 높이와 다름을 알 수 있다. 일반적으로 단독 피처의 크기(너비 x 높이)는 비디오 소스의 크기보다 작지만, 피처의 채널 수가 비디오 소스의 채널 수보다 많기 때문에 모든 피처들의 크기의 합은 비디오 소스의 크기보다 클 수 있다.
- [0145] 도 10의 (a)는 비디오 소스(Image/Video)의 크기를 나타내며, 도 10의 (b)는 채널의 수가 256인 피처맵을 나타낸다. 도 10를 참조하면, 개별 피처의 크기는 비디오 소스보다 작지만 채널 수의 증가로 인하여 총 크기(피처맵의 크기)는 비디오 소스보다 큰 것을 알 수 있다.
- [0146] 피처 또는 피처맵을 효율적으로 압축하기 위해 피처맵의 크기를 효율적으로 줄이는 방법이 적용될 수 있다. 피처맵 변환(또는, 피처 차원 변환)이 피처맵의 크기를 줄이는 방법으로 이용될 수 있으며, 피처 차원 변환을 통해 피처맵의 크기를 줄이는 방법에 대한 일 예가 도 11에 나타나 있다.
- [0147] 피처 변환부(1110)는 피처 추출 네트워크(710)로부터 출력된 피처맵(Extracted feature map)을 저차원의 피처맵(Low-dimensional feature map, 저차원 피처맵)으로 변환하고, 인코딩 장치(720)는 저차원 피처맵을 인코딩하여 비트스트림 형태로 출력할 수 있다. 디코딩 장치(730)는 비트스트림으로부터 저차원 피처맵을 디코딩하고, 피처 역변환부(1120, Inverse Feature Transform)는 저차원 피처맵을 역변환하여 고차원의 피처맵(High-dimensional feature map, 고차원 피처맵)으로 변환할 수 있다. 피처 역변환부(1120)의 기능은 저차원 피처맵을 고차원 피처맵으로 복원하여 복원된 피처맵(Reconstructed feature map)을 출력하는 것일 수 있다. 여기서, 복원된 피처맵의 차원은 피처 추출 네트워크(710)로 입력되기 전의 차원과 동일할 수 있다.
- [0148] 한편, VCM의 부호화 대상인 피처 맵은 복수 개의 채널로 구성될 수 있다. 종래의 이미지/비디오 압축 기술은 YUV/RGB 포맷에 따라 컴포넌트의 개수를 미리 인지하고 이를 기반으로 비트스트림을 구성할 수 있다. 그러나 VCM의 압축 대상인 네트워크의 피처는 다수의 네트워크를 지원하여야 하며, 피처는 네트워크에 종속적으로 그 개수나 전송하는 피처의 레이어 개수가 변화할 수 있다. 이러한 경우, 단일 부호화 시스템 안에서 하나 이상의 피처 레이어를 구성하는 피처 정보를 부호화 하여야 하며, 각 피처 레이어의 피처는 서로 다른 포맷으로 구성될 수 있기 때문에 부호화기에서는 동일 시간 인스턴스(time instance)에 전송된 복수의 피처 정보에 대한 포맷 정보를 전송받아야 부호화를 지원할 수 있다.
- [0149] 도 12는 상기에서 설명한 비디오/영상 코딩 시스템을 확장/재설계하여 사용자 혹은 머신(Machine)의 요청과 목적, 주변 환경에 따라 비디오 소스(VIDEO SOURCE)의 일부만을 사용하거나 비디오 소스에서 필요한 부분/정보를 획득하여 이용할 수 있는 소스 디바이스 및 수신 디바이스의 일 예를 나타낸 도면이다.
- [0150] VCM(Video coding for machines)은 사용자 또는/및 머신(Machine)의 요청, 목적, 주변 환경에 따라 비디오 소스의 일부 피처(Feature) 또는/와 비디오 소스에서 필요한 피처 정보(Feature)를 획득하여 부호화/복호화하는 것을 의미한다. VCM에서 부호화의 대상은 피처라고 지칭되며, 피처는 비디오 소스에서 사용자 또는/및 머신의 요청, 목적, 주변 환경에 따라 추출된 정보이자 시간의 흐름에 따른 일련의 정보들의 집합을 의미할 수 있다.
- [0151] 도 12를 참조하면, VCM 시스템은 제1 장치(소스 디바이스) 및 제2 장치(수신 디바이스)를 포함할 수 있다. 소스 디바이스는 인코딩된 피처 정보를 저장매체 또는 네트워크를 통하여 수신 디바이스로 전달할 수 있다. 제1 장치와 제2 장치의 사용 주체는 각각 사람 혹은/및 머신일 수 있다. 비디오 소스는 비디오/영상의 캡처, 합성 또는 생성 과정 등을 통하여 비디오/영상을 획득할 수 있다. 비디오 소스는 비디오/영상 캡처 디바이스 및/또는 비디오/영상 생성 디바이스를 포함할 수 있다.
- [0152] 피처 추출 네트워크(Feature Extraction Network)는 비디오 소스에서 피처를 획득할 수 있다. 피처 추출 네트워크는 비디오/영상 캡처 디바이스 및/또는 비디오/영상 생성 디바이스를 포함할 수 있다. 피처는 사용자 또는/및 머신의 요청, 목적, 주변 환경에 따라 비디오 소스에서 추출/가공된 정보 일 수 있으며 시간의 흐름에 따른 일련의 정보들의 집합을 의미할 수 있다.
- [0153] 피처 시그널링/인터페이스는 소스 디바이스에서 피처 압축과 인코딩 장치간의 시그널링 정보/인터페이스를 나타낼 수 있다. 또한 수신 디바이스에서 피처 시그널링/인터페이스는 디코딩 장치에서는 디코딩 장치와 태스크 분

석/렌더러(Task analysis/renderer) 간의 시그널링 정보/인터페이스를 의미한다. 피쳐 시그널링/인터페이스는 피쳐 추출 네트워크를 통해 획득한 피쳐의 종류, 획득 방법, 목적 등과 같은 피쳐에 대한 정보를 포함할 수 있다. 또한 피쳐 추출 네트워크가 필요로 하는 정보/동작을 요청하기 위한 정보를 포함할 수 있다.

[0154] 인코딩 장치는 피쳐의 압축 및 코딩 효율을 위하여 예측, 변환, 양자화 등 일련의 절차를 수행할 수 있다. 인코딩된 데이터(인코딩된 피쳐 정보)는 비트스트림(bitstream) 형태로 출력될 수 있다.

[0155] 송신/저장부(Transmission/storage)는 비트스트림 형태로 출력된 인코딩된 피쳐를 디지털 저장매체 또는 네트워크를 통하여 수신 디바이스의 수신부로 전달할 수 있다. 디지털 저장 매체는 USB, SD, CD, DVD, 블루레이, HDD, SSD 등 다양한 저장 매체를 포함할 수 있다. 전송부는 미리 정해진 파일 포맷을 통하여 미디어 파일을 생성하기 위한 엘리먼트를 포함할 수 있고, 방송/통신 네트워크를 통한 전송을 위한 엘리먼트를 포함할 수 있다. 수신부는 상기 비트스트림을 수신/추출하여 디코딩 장치로 전달할 수 있다.

[0156] 디코딩 장치는 인코딩 장치의 동작에 대응하는 역양자화, 역변환, 예측 등 일련의 절차를 수행하여 비디오/영상을 디코딩할 수 있다.

[0157] 디코딩된 피쳐는 렌더링될 수 있다. 그 뿐만 아니라 사용자 혹은 머신의 임무를 수행하는 데 이용될 수도 있다. 임무의 예로는 얼굴 인식, 행동 인식, 차선 인식 등과 같은 AI, 컴퓨터 비전 태스크(Computer Vision Task) 등이 포함될 수 있다.

[0158] 본 개시에서는 피쳐의 획득 및 코딩에 관한 다양한 실시예들을 제시하며, 다른 언급이 없는 한 상기 실시예들은 서로 조합되어 수행될 수도 있다. 이 문서에서 개시된 방법/실시예는 VCM(Video Coding for Machines)표준에 개시되는 방법에 적용될 수 있다.

[0159] 한편, VCM기술이 활용될 수 있는 분야 중 감시 시스템(Surveillance system)은 물체/사람을 인식하고 추적하는 등의 태스크를 수행할 수 있다. 감시 카메라(Surveillance camera)로부터 획득한 정보를 전송하거나 저장하는 목적으로 VCM을 사용할 수 있다.

[0160] 또한, 지능형 교통 시스템(Intelligent Transportation) 분야에 VCM을 활용할 수도 있다. 스마트 트래픽 시스템(Smart traffic system)을 구성하는 각 차량은 GPS로부터 수집한 위치정보와 카메라, 라이다(LIDAR), 레이더(radar), 울트라소닉 센서(ultrasonic sensor) 등 자율주행 및 안전에 필요한 센서정보, 각종 차량제어 정보(브레이크, 와이퍼, 헤드라이트등)를 수집하여 다른 차량이나 인프라스트럭처(Infrastructure)로 전송할 수 있다. 각종 센서로부터 수집된 영상 원본 또는 가공된 영상 신호를 VCM으로 압축하여 전송할 수 있으며 이때 데이터 전송은 5G 데이터 네트워크를 사용할 수 있다.

[0161] 뿐만 아니라, 스마트 시티 분야에도 VCM을 활용할 수 있다. 스마트 시티(Smart City)의 상호 연결된 노드 센서와 디바이스가 각각의 태스크를 수행하는데 필요한 정보를 VCM 기술을 사용하여 전송할 수 있다. 스마트 시티의 태스크의 예로는 교통 상황 모니터링, 밀집도 측정 및 예측, 교통 흐름 예측 및 리소스 할당등이 있다.

[0162] 한편, 사람의 비전만을 목적으로 하는 종래의 비디오 코딩과는 달리, VCM은 사람의 비전 뿐 아니라 인공지능/머신러닝의 발전에 따라 머신 태스크 수행을 목적으로 할 수 있으며, 이에 대한 시장의 수요가 증가하고 있다.

[0163] 본 개시를 설명함에 있어서, 채널 레이어는 상기에서 설명한 채널과 동일할 수 있으며, 인터널 레이어는 피쳐 레이어로 표현될 수 있다.

[0164] 이에 따라, 본 개시에서는 다양한 머신 태스크의 피쳐 정보를 효과적으로 비트스트림에 전송할 수 있는 프레임워크, 즉 구조를 제안하고, 디코더 시스템에 전송된 비트스트림으로부터 시스템에서 요구하는 정보만을 추출하여 복호화하고 출력할 수 있는 기술 및 이를 지원하기 위하여 비트스트림에서 전송되는 정보를 제안한다.

[0165] **네트워크 입출력 단위**

[0166] 이하, 도 13 내지 도 20을 참조하여 본 개시의 실시예에 대해 설명한다.

[0167] 근래 다양한 머신 태스크들이 다양한 신경망(neural network) 구조를 기반으로 사용되고 있다. 일 예로 객체 탐색/트래킹/세그멘테이션(Object Detection/Tracking/Segmentation) 등을 지원하는 Detectron2나, 객체 탐색을 지원하는 YOLO등이 있다. 상기 예로든 Detectron2나 YOLO는 서로 다른 신경망 구조를 가지고 있으며, 서로 다른 네트워크를 사용하기 때문에 네트워크에서 사용되는 피쳐의 텐서(tensor)의 크기나, 인터널 레이어(internal Layer)의 개수가 다를 수 있다. 보다 상세하게는, 도 13은 Detectron2의 네트워크 구조를 다이어그램으로 표현한 것이며, 도 14는 도 13의 네트워크 구조에서의 ResNet 레이어 부분을 상세 도식화한 것으로 각 레이어의 입

출력과 네트워크의 종류를 나타낸 것이다.

- [0168] 일 예로서, 도 13에 나타난 바와 같이 Detectron2의 네트워크 구조는 백본 네트워크와 영역 제안 네트워크(region proposal network)를 포함할 수 있으며, GeneralizedRCNN 구조일 수 있다.
- [0169] 도 15는 도 13의 ResNet 레이어에서 인터널 피쳐(internal feature)를 피라미드 형태로 연결(concatenate)하는 FPN 부분을 상세 도식화한 것으로서, 각 입출력의 크기의 예를 나타내기 위한 도면이다. 한편, 상기의 예시에서 설명한 Detectron2와는 달리 객체 탐색 머신 태스크 시스템(Object detection Machine Task System)인 YOLOv3는 도 16 내지 도 17과 같은 네트워크를 가질 수 있다. 도 16은 예시적인 YOLOv3의 네트워크 구조와 각 입출력의 크기를 나타낸 표이며, 도 17은 예시적인 YOLOv3의 네트워크를 도식화 한 것이다.
- [0170] 한편, 동일한 목적의 머신 태스크라 할지라도 세부 네트워크는 서로 다를 수 있으며, 이러한 네트워크의 차이는 인터널 피쳐 텐서의 크기의 차이를 야기할 수 있다. 도 18은, 네트워크의 입출력 단위가 되는 피쳐 텐서를 예를 들어 도식화한 것으로, 도 18을 참조하면, 하나의 레이어의 출력, 다음 레이어의 입력이 되는 피쳐 텐서는 네트워크의 크기에 따라 정해지면, 하나의 텐서를 구성하는 채널의 개수 역시 네트워크에 의하여 결정될 수 있다. 도 18에 의하면, 이러한 피쳐 텐서가 임의의 n 개 존재하면, 피쳐 텐서의 번호는 #0에서 # $n-1$ 로 붙여질 수 있다. 한편, 도 19는 객체 탐색 및 객체 추적에서의 피쳐 텐서를 일 예로서 나타낸 도면이다. 이러한 텐서는 시간 인스턴스에서 객체를 탐색하는 태스크에서 사용될 수 있으며 동일한 구조에서 시간 순서(타임라인, time line)에 따른 피쳐를 나열하여 객체 추적(object tracking)에도 사용될 수 있다. 객체 추적의 경우 동일한 객체를 구별하고 탐색하여 움직임을 파악하여야 하므로, 임의의 시간 n 에서 시작하여 $n+1$, $n+2$... N 에 대한 시간 동안 수행될 수 있다. 한편, 도 20은 하나의 텐서를 채널 단위로 분할하여 Y축으로, 시간 순서 순으로 입력되는 X축으로 표시한 것이다.
- [0171] 지금까지 네트워크의 입출력 단위를 설명하였으며, 이를 기반으로 본 개시에서는 각 레이어에서 추출된 피쳐 텐서를 압축하여 부/복호화 하는 과정에서, 디코더 시스템에서 지원될 수 있는 멀티 피쳐 레이어에 대하여 설명한다.
- [0172] **멀티 피쳐 레이어**
- [0173] 이하, 도 21 내지 도 24를 참조하여 본 개시의 실시예에 대하여 설명한다. 도 21은 머신 태스크가 Detectron2인 경우, 하나의 피쳐 텐서(예를 들어, 스템 레이어(Stem Layer))를 부호화 하여 전송한 경우를 나타낼 수 있다. 스템 피쳐(Stem Feature)를 부호화 하여 비트스트림에 전송한 경우, 디코더, 즉 복호화 장치는 이를 복호화 하여 머신 태스크에 입력으로 넣어줄 수 있다. 이러한 경우에는 스템 레이어의 출력 텐서를 전송하였기 때문에 머신 태스크는 스템 레이어 이전의 네트워크 레이어를 수행하지 않아도 될 수 있다. 다시 말해, 스템 레이어의 출력 텐서를 전송하였기 때문에 머신 태스크는 스템 레이어 이전의 네트워크 레이어에 대하여는 필요한 작업을 수행하지 않을 수 있다.
- [0174] 또 다른 예시로서, 도 22는 예시적으로 하나의 피쳐 텐서 레이어 (예를 들어, C3)를 부호화 하는 경우에 대한 도면일 수 있다. 이 경우, 복호화 장치 이후의 머신 태스크는 C3의 이전 레이어를 수행할 필요가 없을 수 있다. 한편, 도 22는, 또 다른 예시로서, 예시적으로 여러 개의 텐서 레이어를 모두 부호화하여 RestNet 레이어들을 모두 수행하지 않고 FPN(Feature Pyramid Network)의 동작만 수행하여도 머신 태스크가 가능한 경우를 나타낼 수도 있다.
- [0175] 한편, 도 22를 지원하기 위해서는 도 23과 같이 각 네트워크 인터널 레이어의 피쳐들을 정의할 수 있고, 도 24와 같이 각 네트워크 인터널 레이어의 피쳐들을 정의할 수 있으며, 도 25와 같이 이를 3차원 공간에 표현할 수 있다.
- [0176] 상기 예시처럼, 피쳐 복호화 장치는 머신 태스크가 어떤 레이어를 수행할 것인지에 대한 여부를 알 수 없기 때문에 피쳐 복호화 장치와 머신 태스크를 연결하는 시스템 레이어에서는 어떠한 경우라도 지원할 수 있도록 동작해야 할 수 있다.
- [0177] 이를 지원하기 위해, 본 개시의 실시예에서는 상기에서 설명한 바와 같은 계층적 구조를 표현하기 위한 신택스를 제안한다.
- [0178] **실시예**
- [0179] 일 실시예로서, 본 개시에서 제안하는 신택스는 다음과 같이 표현될 수 있다.

표 1

[0180]

Feature_layer_parameter_set() {
..
flps_max_internal_layers_minus1
...
for(i = 0; i <= flps_max_internal_layers_minus1; i++) {
flps_internal_layer_id[i]
flps_max_channel_layers_minus1[i]
for(j = 0; j <= flps_max_channel_layers_minus1[i]; j++) {
flps_channel_layer_id[i][j]
}
}
for(i = 0; i <= flps_max_internal_layers_minus1; i++) {
for(j = 0; j <= flps_max_internal_layers_minus1; j++) {
if (i != j) {
flps_internal_layer_dependency_flag[i][j]
}
}
}
...

[0181]

신택스 정보인 flps_max_internal_layers_minus1에 1을 더한 값은 피쳐 레이어 파라미터 세트에 의해 특정된, 피쳐 레이어 파라미터 세트 (FLPS, feature layer parameter set)를 참조하는 각 CFS(Coded Feature Sequence) 내 허용된 최대 인터널 레이어의 수를 특정할 수 있다. 신택스 정보인 flps_internal_layer_id[i]는 i번째 인터널 레이어의 식별자를 의미할 수 있다.

[0182]

신택스 정보인 flps_max_channel_layers_minus1[i]에 1을 더한 값은 i번째 인터널 레이어의 채널 레이어의 수를 의미할 수 있다.

[0183]

신택스 정보인 flps_channel_layer_id[i][j]는 i번째 인터널 레이어의 j번째 채널 레이어의 식별자를 의미할 수 있다.

[0184]

한편, 상기 신택스에서 사용되는 i 인덱스는 피쳐 레이어를 나타내는 인덱스이며, j 인덱스는 채널 레이어를 나타내는 인덱스일 수 있다. 즉, i 인덱스로 피쳐 레이어를 구분하며, j 인덱스로 채널 레이어를 구분할 수 있다.

[0185]

신택스 정보인 flps_internal_layer_dependency_flag[i][j]이 제1 값(예를 들어, 1)이면, i번째 인터널 레이어가 j번째 인터널 레이어에 종속적임을 나타낼 수 있고, 제2 값(예를 들어, 0)이면, i번째 인터널 레이어가 j번째 인터널 레이어에 독립적임을 나타낼 수 있다. 이 신택스가 존재하지 않으면, flps_internal_layer_dependency_flag[i][j] 값은 디폴트 값(예를 들어, 1)로 설정될 수 있다. 보다 상세하게는, flps_internal_layer_dependency_flag를 도 26을 참조하여 상세히 설명될 수 있다. 도 26은 예시적인 레이어간 종속성을 설명하기 위한 도면으로서, 도 26의 네트워크 피쳐를 전송하는 과정에서, 예를 들어, C4 레이어만을 전송하고, C4 이전 레이어를 바이패스(bypass)하는 것을 불가능할 수 있다. 이는 FPN 부분에서 C3의 피쳐 정보가 필요한데 C4만 부복호화 하여 전송하면, C3 정보가 없기 때문에 태스크 동작이 어려울 수 있기 때문이다. 즉, C4는 C3와 종속성이 있음을 의미할 수 있다. flps_internal_layer_dependency_flag는 이러한 정보를 포기하기 위한 플래그일 수 있다. 한편, flps_internal_layer_dependency_flag[i][j] 신택스 정보에서의 i 인덱스 및 j 인덱스는 모두 인터널 레이어를 구분하는 데 사용될 수 있다.

[0186]

한편, 상기에서 설명된 신택스 정보의 명칭은 설명을 명료하게 하기 위한 일 일 예에 불과하므로, 본 개시가 이에 한정되는 것은 아니다.

[0187]

본 개시에 의하면, 다양한 네트워크를 지원하기 위하여, 하나의 피쳐를 각 레이어로 구성하는 프레임워크가 제안되고, 이는 필요에 따라 레이어를 선택적으로 사용할 수 있는 자유도를 지원할 수 있다.

- [0188] 도 27은 본 개시의 일 실시예에 따른 피쳐 복호화 장치가 수행하는 방법의 일 예를 나타낸 도면이다. 일 예로서, 피쳐 복호화 장치는 상기에서 설명한 영상 복호화 장치를 포함하거나, 영상 복호화 장치일 수 있으며, 상기에서 설명한 다른 실시예도 실시할 수 있음은 자명하다.
- [0189] 도 27을 참조하면, 피쳐 복호화 장치는 CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 대한 제1 정보를 비트스트림으로부터 획득(S2701)할 수 있다. 일 예로서, CFS는 피쳐 파라미터 세트(feature parameter set)를 참조할 수 있다. 일 예로서, 제1 정보는 선택스 요소로서, `flps_max_internal_layers_minus1` 일 수 있다. 제1 정보에 기초하여, 인터널 레이어 각각에 대한 식별자에 관한 제2 정보 및 인터널 레이어의 채널 레이어의 수에 대한 제3 정보가 비트스트림으로부터 획득(S2702)될 수 있다. 일 예로서, 제2 정보는 선택스 요소로서, `flps_internal_layer_id[i]` 일 수 있으며, 인터널 레이어의 수 만큼 획득될 수 있다. 일 예로서, 제3 정보는 선택스 요소로서, `flps_max_channel_layers_minus1[i]` 일 수 있으며, 인터널 레이어의 수 만큼 획득될 수 있다. 이후, 제3 정보에 기초하여 상기 채널 레이어 각각에 대한 식별자에 대한 제4 정보를 비트스트림으로부터 획득(S2703)될 수 있다. 일 예로서, 제4 정보는 선택스 요소로서 `flps_channel_layer_id[i][j]` 일 수 있으며 각 인터널 레이어 마다 채널 레이어의 수 만큼 획득될 수 있다. 또한, 제4 정보에 기초하여 채널 레이어가 복원될 수 있다. 일 예로서, 제1 정보는 최대 인터널 레이어의 수에서 1을 감한 값을 나타낼 수 있다. 제3 정보는 채널 레이어의 수에서 1을 감한 값을 나타낼 수 있다.
- [0190] 한편, 일 예로서, 도 27의 방법에 더하여, 피쳐 복호화 장치가 수행하는 방법은 제1 정보 및 제3 정보에 기초하여 인터널 레이어 간의 종속성에 대한 제5 정보가 비트스트림으로부터 더 획득될 수 있다. 일 예로서, 제5 정보가 제1 값(예를 들어, 1)이면, 인터널 레이어 간의 종속성이 있음을 나타낼 수 있다. 제5 정보가 제2 값(예를 들어, 0)이면, 인터널 레이어 간의 종속성이 없음을 나타낼 수 있다. 일 예로서, 제5 정보는 인터널 레이어 인덱스가 서로 다른 인터널 레이어에 대하여만 획득될 수 있다. 한편, 제5 정보가 비트스트림에 존재하지 않는 경우, 제5 정보는 제1 값(예를 들어, 1)으로 유도될 수 있다. 한편, 제1 정보 내지 제5 정보는 모두 피쳐 파라미터 세트들을 포함한 파라미터 세트에 포함될 수 있다.
- [0191] 한편, 도 27은 본 개시의 일 실시예에 해당하므로, 설명한 선택스 요소의 이름으로 본 개시가 한정되는 것은 아니며, 선택스 요소가 획득되는 순서 일부가 변경되거나, 일부 선택스 요소는 획득되지 않을 수도 있다.
- [0192] 도 28은 본 개시의 일 실시예에 따른 피쳐 부호화 장치가 수행하는 방법의 일 예를 나타낸 도면이다. 일 예로서, 피쳐 부호화 장치는 상기에서 설명한 영상 부호화 장치를 포함하거나, 영상 부호화 장치일 수 있으며, 상기에서 설명한 다른 실시예도 실시할 수 있음은 자명하다.
- [0193] 도 28을 참조하면, 피쳐 부호화 장치는 CFS(coded feature sequence) 내 허용된 최대 인터널 레이어의 수에 대한 제1 정보를 결정(S2801)하고, 인터널 레이어 각각에 대한 식별자에 대한 제2 정보 및 인터널 레이어의 채널 레이어의 수에 대한 제3 정보를 결정(S2802)할 수 있다. 일 예로서, CFS는 피쳐 파라미터 세트를 참조할 수 있다. 또한, 채널 레이어 각각에 대한 식별자에 대한 제4 정보를 결정(S2803)할 수 있다. 일 예로서, 제1 정보, 제2 정보, 제3 정보 및 제4 정보는 비트스트림으로 부호화될 수 있다.
- [0194] 일 예로서, 제1 정보는 선택스 요소로서, `flps_max_internal_layers_minus1` 일 수 있다. 일 예로서, 제2 정보는 선택스 요소로서, `flps_internal_layer_id[i]` 일 수 있으며, 인터널 레이어의 수 만큼 결정될 수 있다. 일 예로서, 제3 정보는 선택스 요소로서, `flps_max_channel_layers_minus1[i]` 일 수 있으며, 인터널 레이어의 수 만큼 결정될 수 있다. 일 예로서, 제4 정보는 선택스 요소로서 `flps_channel_layer_id[i][j]` 일 수 있으며 각 인터널 레이어 마다 채널 레이어의 수 만큼 결정될 수 있다. 또한, 제4 정보에 기초하여 채널 레이어가 부호화될 수 있다.
- [0195] 일 예로서, 제1 정보는 최대 인터널 레이어의 수에서 1을 감한 값을 나타낼 수 있다. 제3 정보는 채널 레이어의 수에서 1을 감한 값을 나타낼 수 있다. 한편, 일 예로서, 도 28의 방법에 더하여, 피쳐 부호화 장치는 인터널 레이어 간의 종속성에 대한 제5 정보를 더 결정할 수 있으며, 제5 정보를 비트스트림으로 부호화할 수 있다. 일 예로서, 제5 정보가 제1 값(예를 들어, 1)이면, 인터널 레이어 간의 종속성이 있음을 나타낼 수 있다. 제5 정보가 제2 값(예를 들어, 0)이면, 인터널 레이어 간의 종속성이 없음을 나타낼 수 있다. 일 예로서, 제5 정보는 인터널 레이어 인덱스가 서로 다른 인터널 레이어에 대하여만 결정될 수 있다. 한편, 제5 정보가 부호화되지 않는 경우, 비트스트림의 수신 측에서는 제5 정보는 기본 값으로 유도할 수 있는데, 기본 값은 예를 들어, 1일 수 있다. 한편, 제1 정보 내지 제5 정보는 모두 피쳐 파라미터 세트들을 포함한 파라미터 세트에 포함될 수 있다.
- [0196] 한편, 도 28은 본 개시의 일 실시예에 해당하므로, 설명한 선택스 요소의 이름으로 본 개시가 한정되는 것은 아

니며, 선택 요소가 동시에 결정되거나, 선택 요소가 결정되는 순서 일부가 변경되거나, 일부 선택 요소는 결정되지 않을 수도 있다.

- [0197] 본 개시에 따르면, 스템 피처(Stem Feature)를 부호화 하여 비트스트림에 전송한 경우, 복호화 장치는 이를 복호화하여 머신 태스크(Machine Task)에 입력으로 넣어줄 수 있다. 이러한 경우에는 스템 레이어(stem Layer)의 출력 텐서(Tensor)를 전송하였기 때문에 머신 태스크(Machine Task)는 스템 레이어 이전의 네트워크 레이어(Network Layer)를 수행하지 않아도 될 수 있으므로, 보다 높은 자유도를 지원할 수 있다.
- [0198] 또한, 본 개시에 의하면, 머신 태스크가 어떤 레이어를 수행할 것인지에 대한 여부를 알 수 없는 경우도 지원할 수 있어, 부호화/복호화 효율이 향상될 수 있다.
- [0199] 본 개시의 예시적인 방법들은 설명의 명확성을 위해서 동작의 시리즈로 표현되어 있지만, 이는 단계가 수행되는 순서를 제한하기 위한 것은 아니며, 필요한 경우에는 각각의 단계가 동시에 또는 상이한 순서로 수행될 수도 있다. 본 개시에 따른 방법을 구현하기 위해서, 예시하는 단계에 추가적으로 다른 단계를 포함하거나, 일부의 단계를 제외하고 나머지 단계를 포함하거나, 또는 일부의 단계를 제외하고 추가적인 다른 단계를 포함할 수도 있다.
- [0200] 본 개시에 있어서, 소정의 동작(단계)을 수행하는 영상 부호화 장치 또는 영상 복호화 장치는 해당 동작(단계)의 수행 조건이나 상황을 확인하는 동작(단계)을 수행할 수 있다. 예를 들어, 소정의 조건이 만족되는 경우 소정의 동작을 수행한다고 기재된 경우, 영상 부호화 장치 또는 영상 복호화 장치는 상기 소정의 조건이 만족되는지 여부를 확인하는 동작을 수행한 후, 상기 소정의 동작을 수행할 수 있다.
- [0201] 본 개시의 다양한 실시예는 모든 가능한 조합을 나열한 것이 아니고 본 개시의 대표적인 양상을 설명하기 위한 것이며, 다양한 실시예에서 설명하는 사항들은 독립적으로 적용되거나 또는 둘 이상의 조합으로 적용될 수도 있다.
- [0202] 본 개시에서 설명한 실시예들은 프로세서, 마이크로 프로세서, 컨트롤러 또는 칩 상에서 구현되어 수행될 수 있다. 예를 들어, 각 도면에서 도시한 기능 유닛들은 컴퓨터, 프로세서, 마이크로 프로세서, 컨트롤러 또는 칩 상에서 구현되어 수행될 수 있다. 이 경우 구현을 위한 정보(e.g., information on instructions) 또는 알고리즘이 디지털 저장 매체에 저장될 수 있다.
- [0203] 또한, 본 개시의 실시예(들)이 적용되는 디코더(디코딩 장치) 및 인코더(인코딩 장치)는 멀티미디어 방송 송수신 장치, 모바일 통신 단말, 홈 시네마 비디오 장치, 디지털 시네마 비디오 장치, 감시용 카메라, 비디오 대화 장치, 비디오 통신과 같은 실시간 통신 장치, 모바일 스트리밍 장치, 저장 매체, 캠코더, 주문형 비디오(VoD) 서비스 제공 장치, OTT 비디오(Over the top video) 장치, 인터넷 스트리밍 서비스 제공 장치, 3차원(3D) 비디오 장치, VR(virtual reality) 장치, AR(argumente reality) 장치, 화상 전화 비디오 장치, 운송 수단 단말(ex. 차량(자율주행차량 포함) 단말, 로봇트 단말, 비행기 단말, 선박 단말 등) 및 의료용 비디오 장치 등에 포함될 수 있으며, 비디오 신호 또는 데이터 신호를 처리하기 위해 사용될 수 있다. 예를 들어, OTT 비디오(Over the top video) 장치로는 게임 콘솔, 블루레이 플레이어, 인터넷 접속 TV, 홈시어터 시스템, 스마트폰, 태블릿 PC, DVR(Digital Video Recorder) 등을 포함할 수 있다.
- [0204] 또한, 본 개시의 실시예(들)이 적용되는 처리 방법은 컴퓨터로 실행되는 프로그램의 형태로 생산될 수 있으며, 컴퓨터가 판독할 수 있는 기록 매체에 저장될 수 있다. 본 문서의 실시예(들)에 따른 데이터 구조를 가지는 멀티미디어 데이터도 또한 컴퓨터가 판독할 수 있는 기록 매체에 저장될 수 있다. 컴퓨터가 판독할 수 있는 기록 매체는 컴퓨터로 읽을 수 있는 데이터가 저장되는 모든 종류의 저장 장치 및 분산 저장 장치를 포함한다. 컴퓨터가 판독할 수 있는 기록 매체는, 예를 들어, 블루레이 디스크(BD), 범용 직렬 버스(USB), ROM, PROM, EPROM, EEPROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크 및 광학적 데이터 저장 장치를 포함할 수 있다. 또한, 컴퓨터가 판독할 수 있는 기록 매체는 반송파(예를 들어, 인터넷을 통한 전송)의 형태로 구현된 미디어를 포함한다. 또한, 인코딩 방법으로 생성된 비트스트림이 컴퓨터가 판독할 수 있는 기록 매체에 저장되거나 유무선 통신 네트워크를 통해 전송될 수 있다.
- [0205] 또한, 본 개시의 실시예(들)은 프로그램 코드에 의한 컴퓨터 프로그램 제품으로 구현될 수 있고, 프로그램 코드는 본 개시의 실시예(들)에 의해 컴퓨터에서 수행될 수 있다. 프로그램 코드는 컴퓨터에 의해 판독가능한 캐리어 상에 저장될 수 있다.
- [0206] 도 29는 본 개시의 실시예들이 적용될 수 있는 콘텐츠 스트리밍 시스템의 일 예를 나타내는 도면이다.

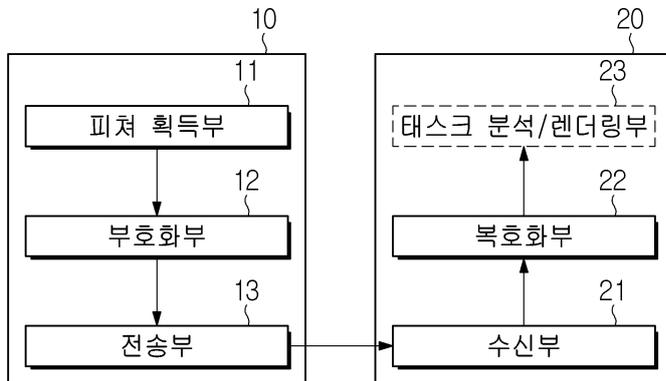
- [0207] 도 29를 참조하면, 본 개시의 실시예가 적용된 콘텐츠 스트리밍 시스템은 크게 인코딩 서버, 스트리밍 서버, 웹 서버, 미디어 저장소, 사용자 장치 및 멀티미디어 입력 장치를 포함할 수 있다.
- [0208] 인코딩 서버는 스마트폰, 카메라, 캠코더 등과 같은 멀티미디어 입력 장치들로부터 입력된 콘텐츠를 디지털 데이터로 압축하여 비트스트림을 생성하고 이를 스트리밍 서버로 전송하는 역할을 한다. 다른 예로, 스마트폰, 카메라, 캠코더 등과 같은 멀티미디어 입력 장치들이 비트스트림을 직접 생성하는 경우, 인코딩 서버는 생략될 수 있다.
- [0209] 비트스트림은 본 개시의 실시예가 적용된 영상 부호화 방법 및/또는 영상 부호화 장치에 의해 생성될 수 있고, 스트리밍 서버는 비트스트림을 전송 또는 수신하는 과정에서 일시적으로 비트스트림을 저장할 수 있다.
- [0210] 스트리밍 서버는 웹 서버를 통한 사용자 요청에 기반하여 멀티미디어 데이터를 사용자 장치에 전송하고, 웹 서버는 사용자에게 어떠한 서비스가 있는지를 알려주는 매개체 역할을 할 수 있다. 사용자가 웹 서버에 원하는 서비스를 요청하면, 웹 서버는 이를 스트리밍 서버에 전달하고, 스트리밍 서버는 사용자에게 멀티미디어 데이터를 전송할 수 있다. 이 때, 콘텐츠 스트리밍 시스템은 별도의 제어 서버를 포함할 수 있고, 이 경우 제어 서버는 콘텐츠 스트리밍 시스템 내 각 장치 간 명령/응답을 제어하는 역할을 수행할 수 있다.
- [0211] 스트리밍 서버는 미디어 저장소 및/또는 인코딩 서버로부터 콘텐츠를 수신할 수 있다. 예를 들어, 인코딩 서버로부터 콘텐츠를 수신하는 경우, 콘텐츠를 실시간으로 수신할 수 있다. 이 경우, 원활한 스트리밍 서비스를 제공하기 위하여 스트리밍 서버는 비트스트림을 일정 시간동안 저장할 수 있다.
- [0212] 사용자 장치의 예로는, 휴대폰, 스마트 폰(smart phone), 노트북 컴퓨터(laptop computer), 디지털방송용 단말기, PDA(personal digital assistants), PMP(portable multimedia player), 네비게이션, 슬레이트 PC(slate PC), 태블릿 PC(tablet PC), 울트라북(ultrabook), 웨어러블 디바이스(wearable device, 예를 들어, 위치형 단말기 (smartwatch), 글래스형 단말기 (smart glass), HMD(head mounted display)), 디지털 TV, 데스크탑 컴퓨터, 디지털 사이니지 등이 있을 수 있다.
- [0213] 콘텐츠 스트리밍 시스템 내 각 서버들은 분산 서버로 운영될 수 있으며, 이 경우 각 서버에서 수신하는 데이터는 분산 처리될 수 있다.
- [0214] 도 30은 본 개시의 실시예들이 적용될 수 있는 콘텐츠 스트리밍 시스템의 다른 예를 나타내는 도면이다.
- [0215] 도 30을 참조하면, VCM과 같은 실시예에서, 기기의 성능, 사용자의 요청, 수행하고자 하는 태스크의 특성 등에 따라 사용자 단말에서 태스크를 수행할 수도 있고 외부 기기(e.g., 스트리밍 서버, 분석 서버 등)에서 태스크를 수행할 수도 있다. 이와 같이, 태스크 수행에 필요한 정보를 외부 기기로 전송하기 위하여, 사용자 단말은 태스크 수행에 필요한 정보(e.g., 태스크, 신경망 네트워크 및/또는 용도와 같은 정보)를 포함하는 비트스트림을 직접 또는 인코딩 서버를 통해 생성할 수 있다.
- [0216] 분석 서버는 사용자 단말로부터(또는, 인코딩 서버로부터) 전송받은 부호화된 정보를 복호화한 후, 사용자 단말의 요청 태스크를 수행할 수 있다. 분석 서버는 태스크 수행을 통해 얻은 결과를 사용자 단말에게 다시 전송하거나 연계된 다른 서비스 서버(e.g., 웹 서버)에 전송할 수 있다. 예를 들어, 분석 서버는 화재를 판별하는 태스크를 수행하여 얻은 결과를 소방 관련 서버로 전송할 수 있다. 분석 서버는 별도의 제어 서버를 포함할 수 있고, 이 경우 제어 서버는 분석 서버와 연관된 각 장치와 서버 간 명령/응답을 제어하는 역할을 할 수 있다. 또한, 분석 서버는 사용자 기기가 수행하고자 하는 태스크와 수행할 수 있는 태스크 정보를 기반으로 웹 서버에게 원하는 정보를 요청할 수도 있다. 분석 서버가 웹 서버에 원하는 서비스를 요청하면, 웹 서버는 이를 분석 서버에 전달하고, 분석 서버는 사용자 단말로 그에 대한 데이터를 전송할 수 있다. 이 때, 콘텐츠 스트리밍 시스템의 제어 서버는 스트리밍 시스템 내 각 장치 간 명령/응답을 제어하는 역할을 수행할 수 있다.

산업상 이용가능성

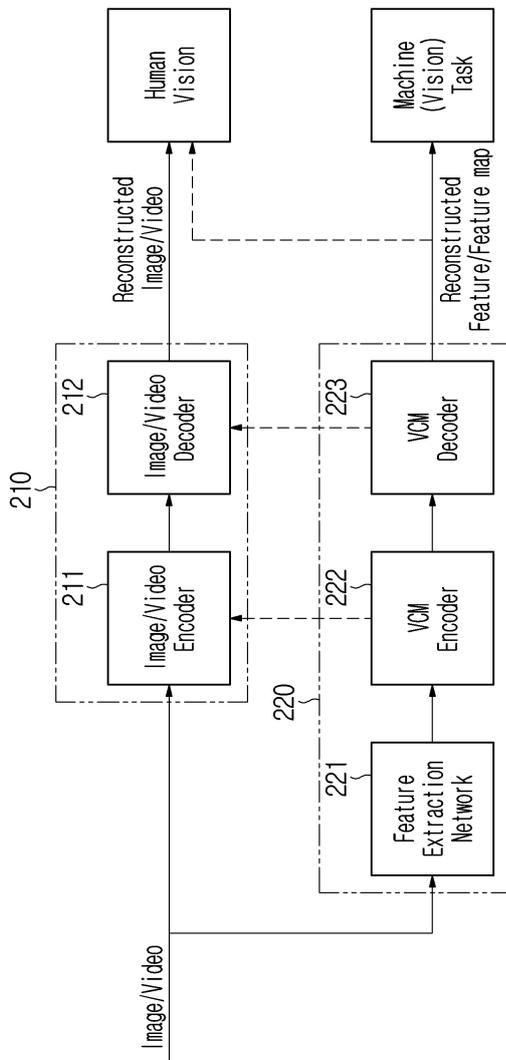
- [0217] 본 개시에 따른 실시예는 피쳐/피쳐맵을 부호화/복호화하는데 이용될 수 있다.

도면

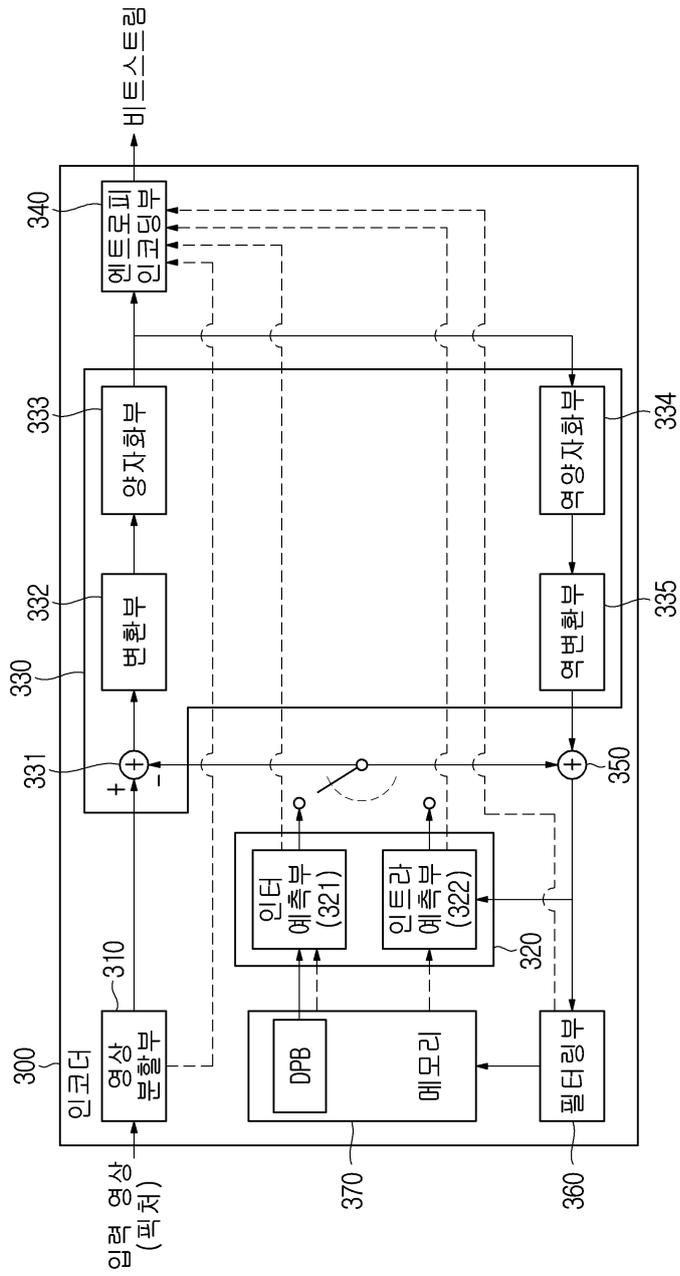
도면1



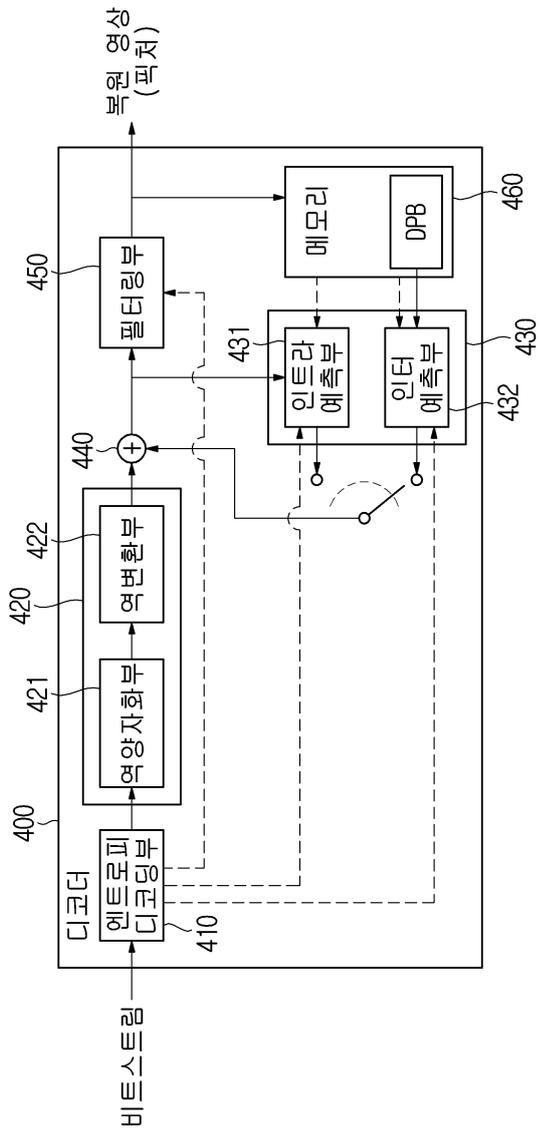
도면2



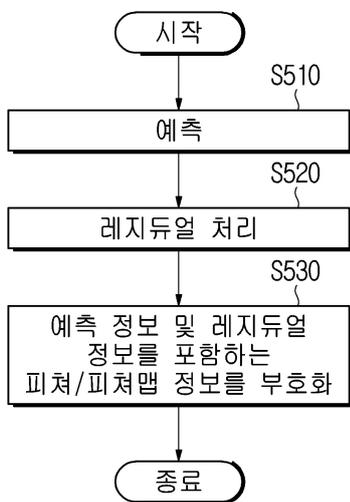
도면3



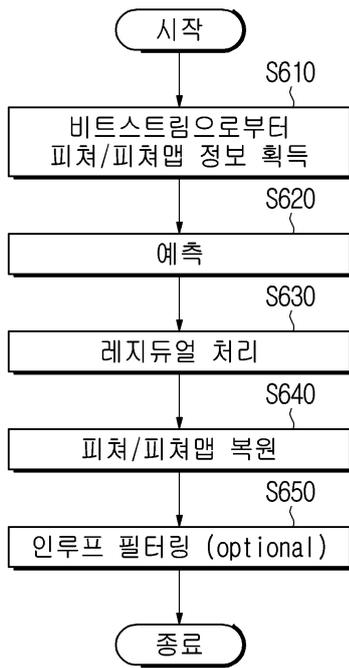
도면4



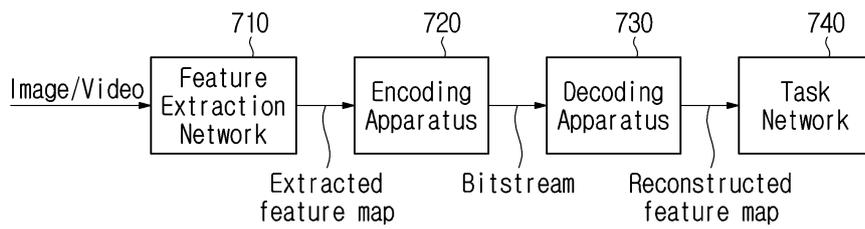
도면5



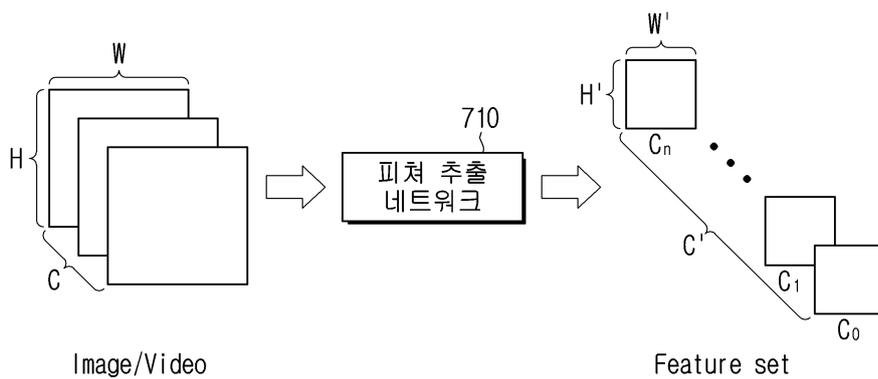
도면6



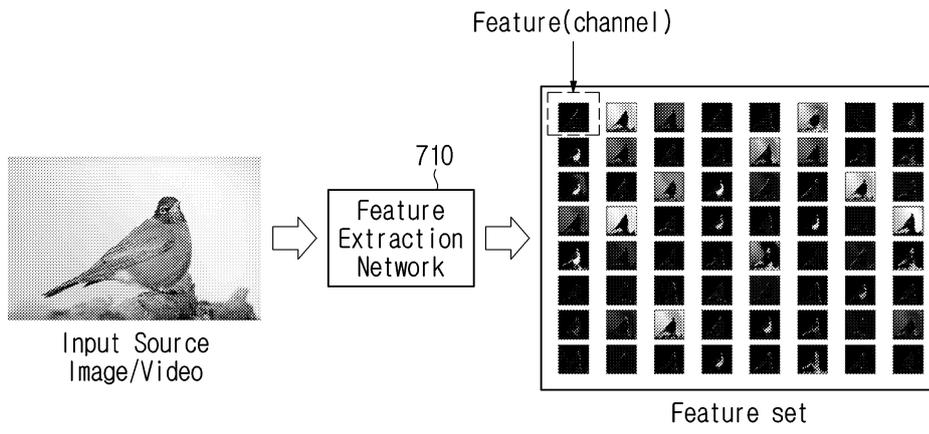
도면7



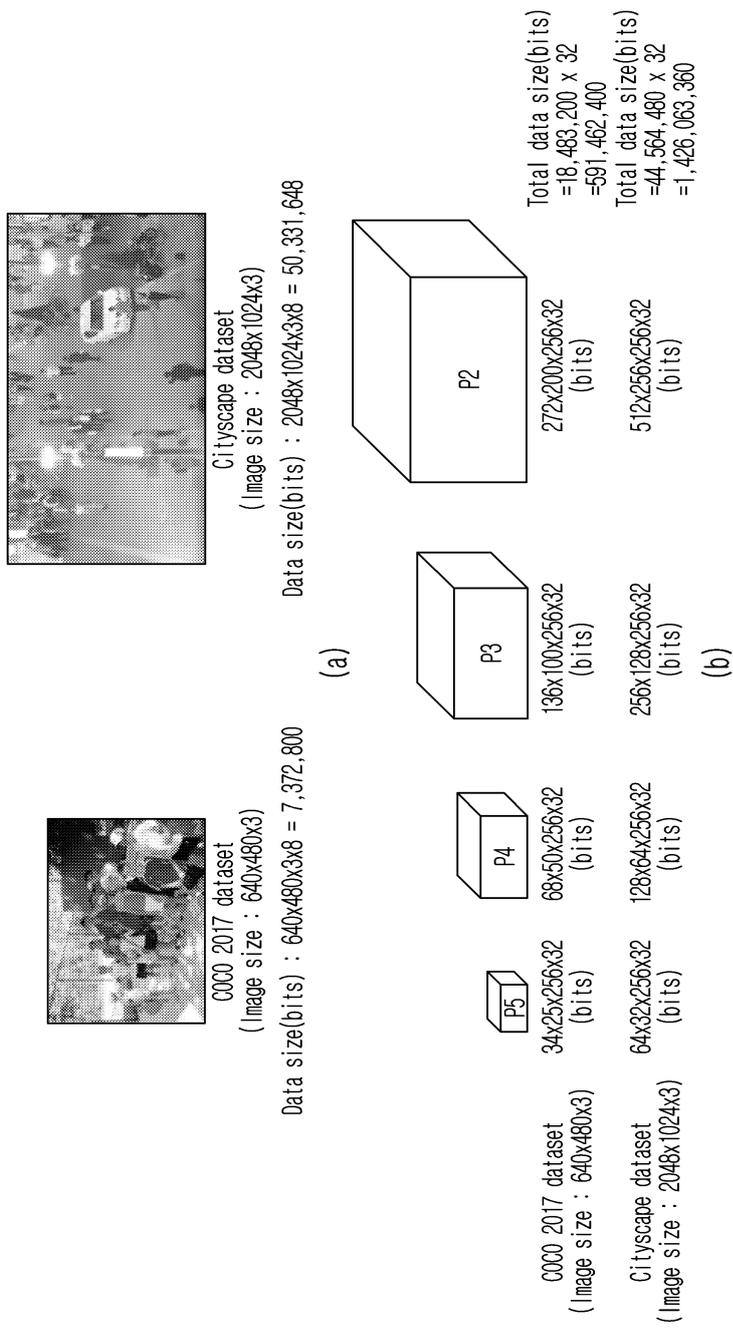
도면8



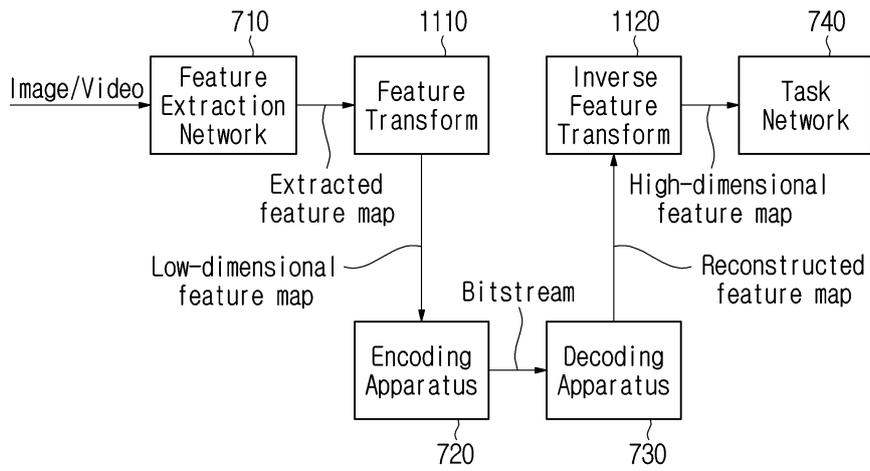
도면9



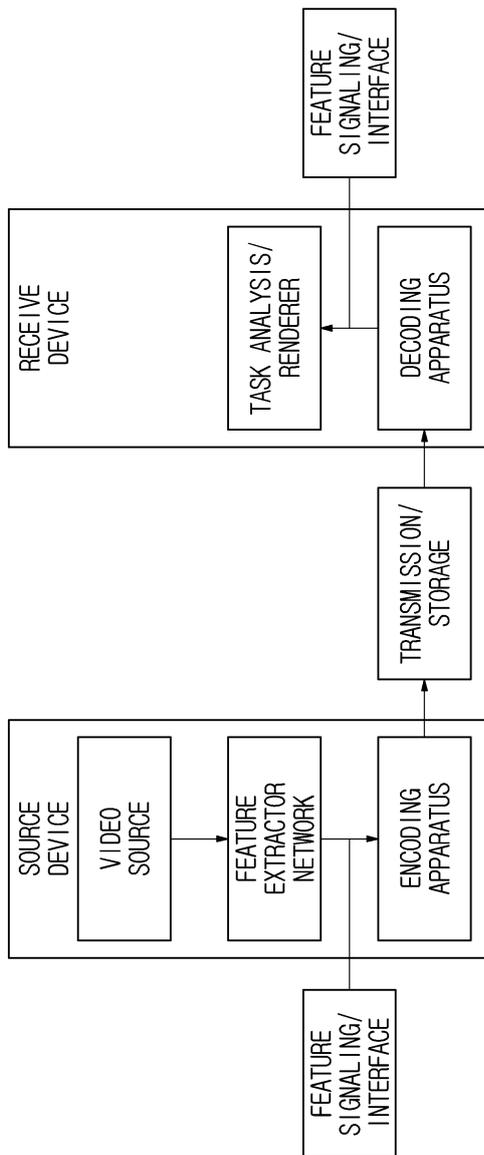
도면10



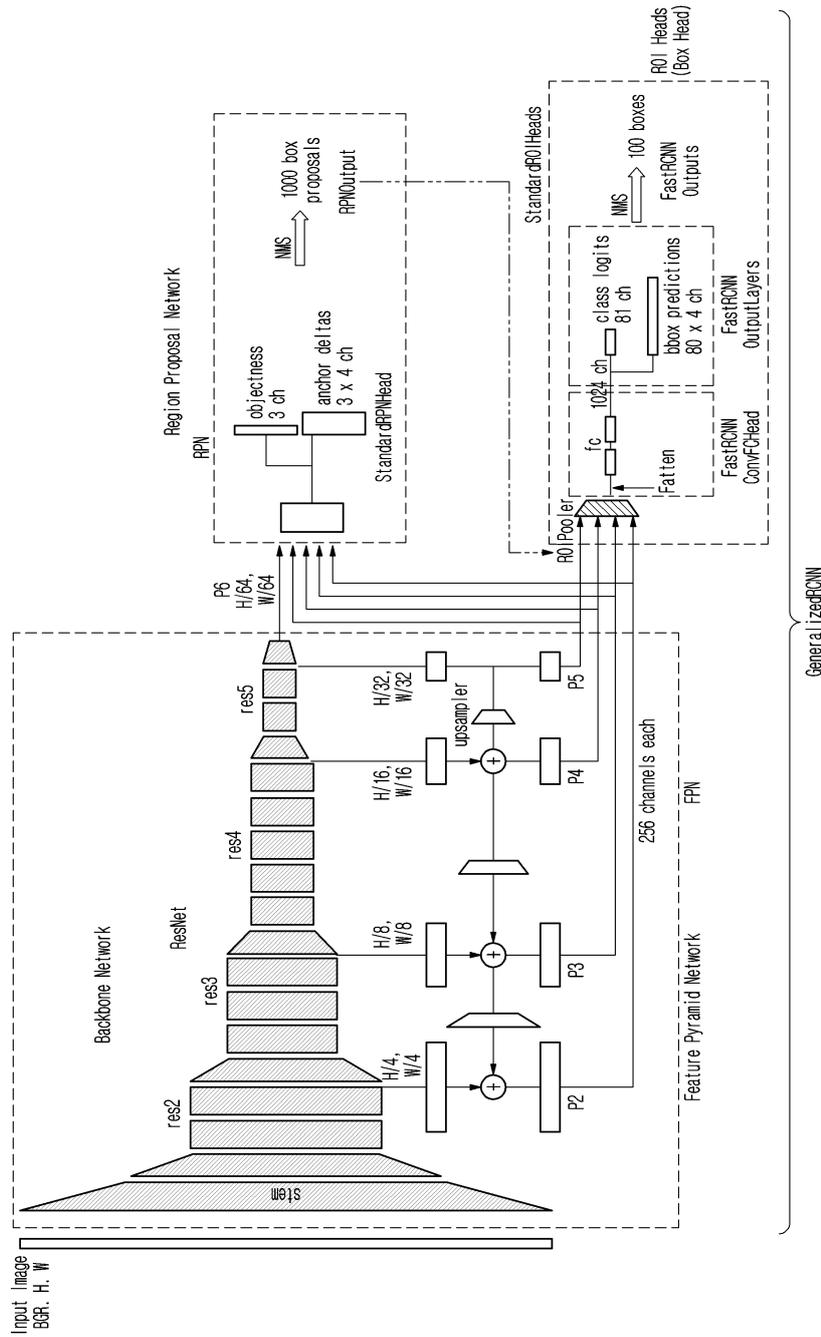
도면11



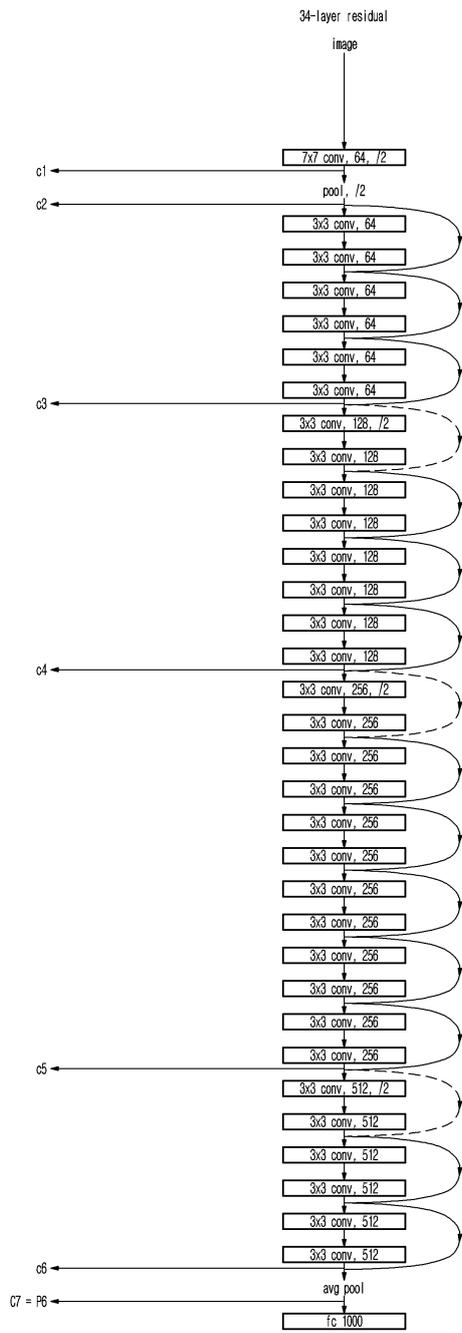
도면12



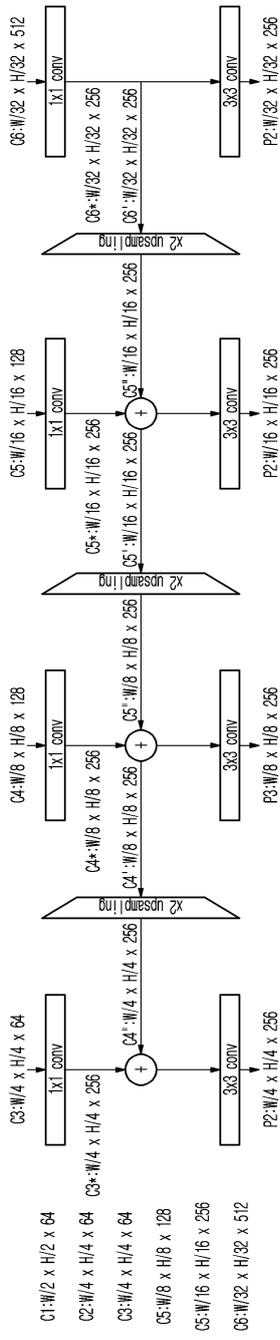
도면13



도면14



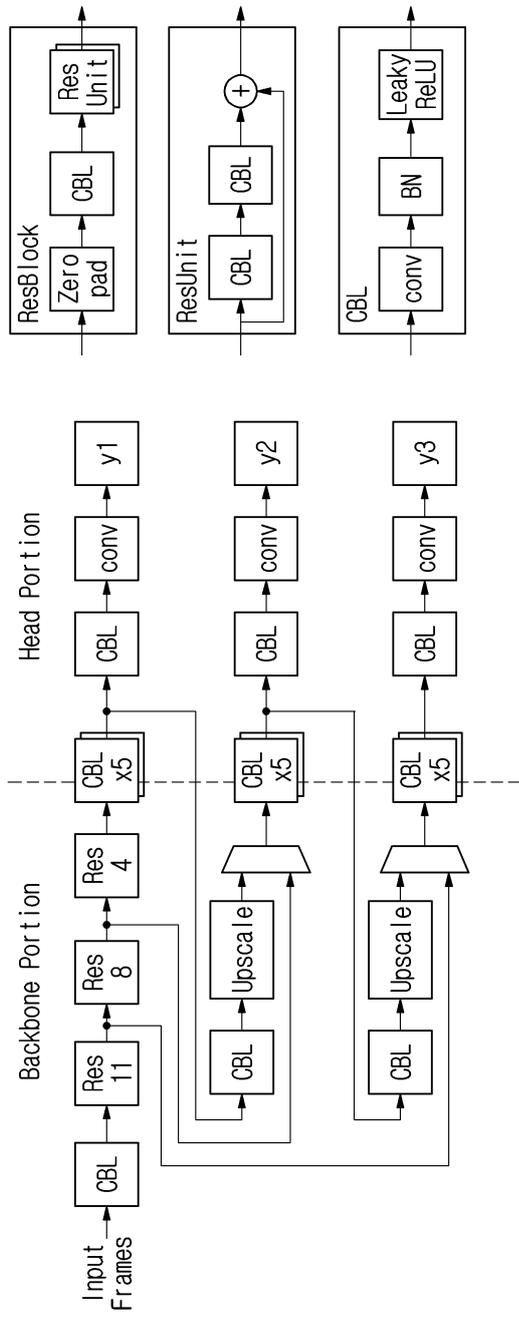
도면15



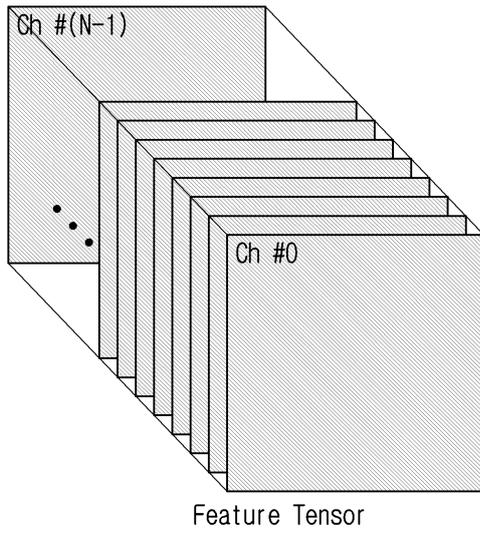
도면16

	Type	Filters	Size	Output
	Convolutional	32	3 x 3	256 x 256
	Convolutional	64	3 x 3 / 2	128 x 128
1 x	Convolutional	32	1 x 1	
	Convolutional	64	3 x 3	
	Residual			128 x 128
	Convolutional	128	3 x 3 / 2	64 x 64
2 x	Convolutional	64	1 x 1	
	Convolutional	128	3 x 3	
	Residual			64 x 64
	Convolutional	256	3 x 3 / 2	32 x 32
8 x	Convolutional	128	1 x 1	
	Convolutional	256	3 x 3	
	Residual			32 x 32
	Convolutional	512	3 x 3 / 2	16 x 16
8 x	Convolutional	256	1 x 1	
	Convolutional	512	3 x 3	
	Residual			16 x 16
	Convolutional	1024	3 x 3 / 2	8 x 8
4 x	Convolutional	512	1 x 1	
	Convolutional	1024	3 x 3	
	Residual			8 x 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

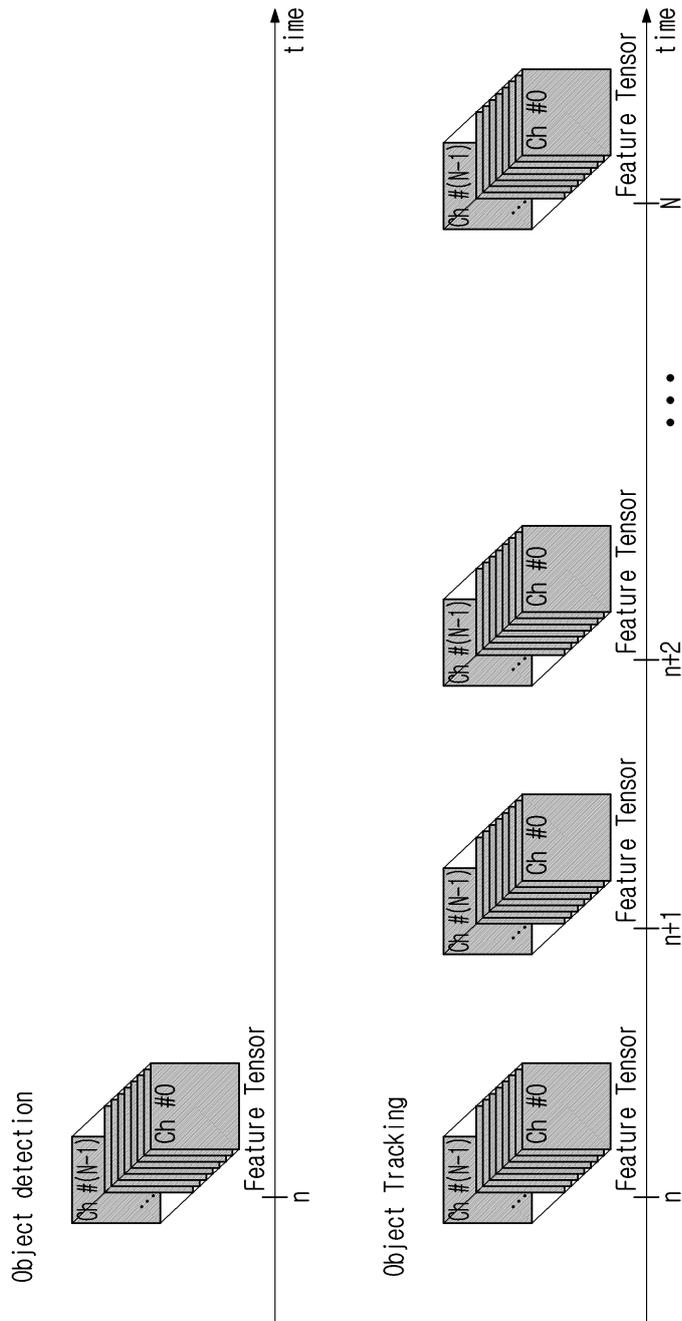
도면17



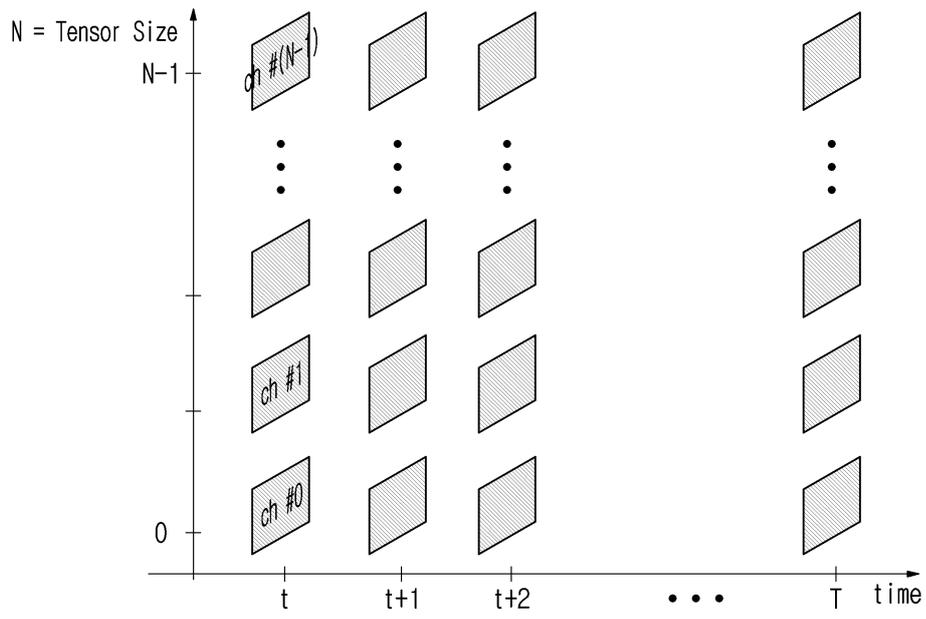
도면18



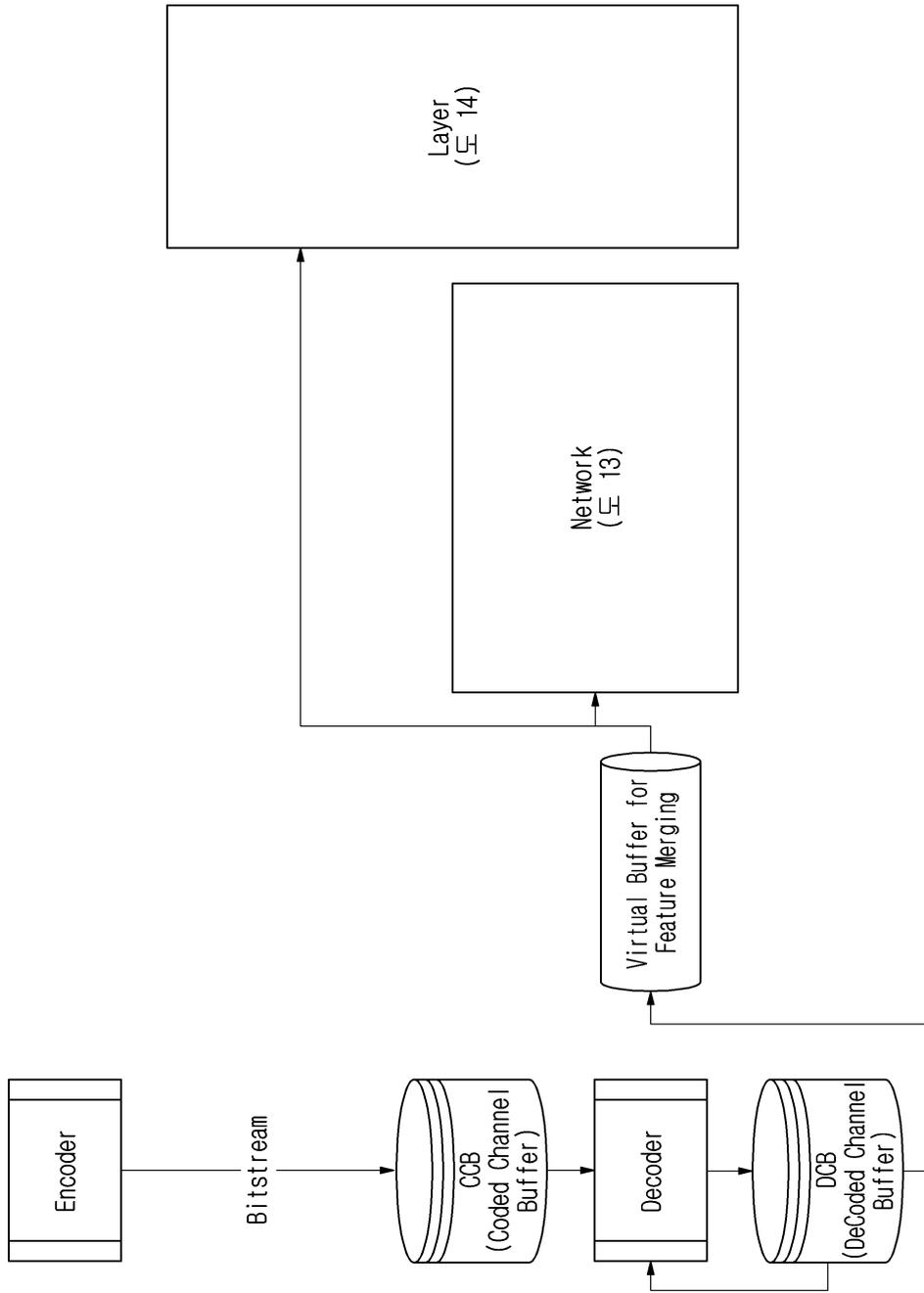
도면19



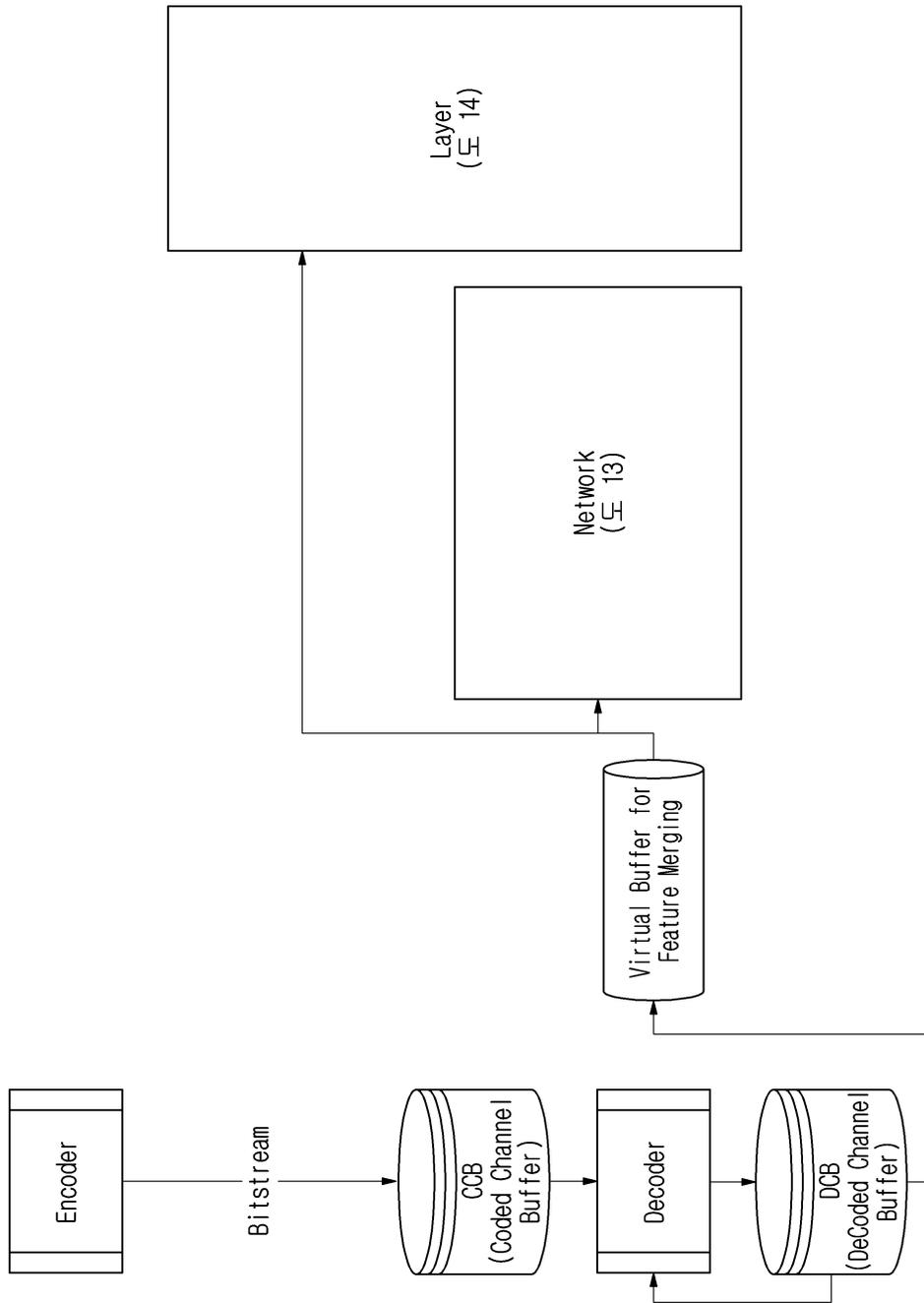
도면20



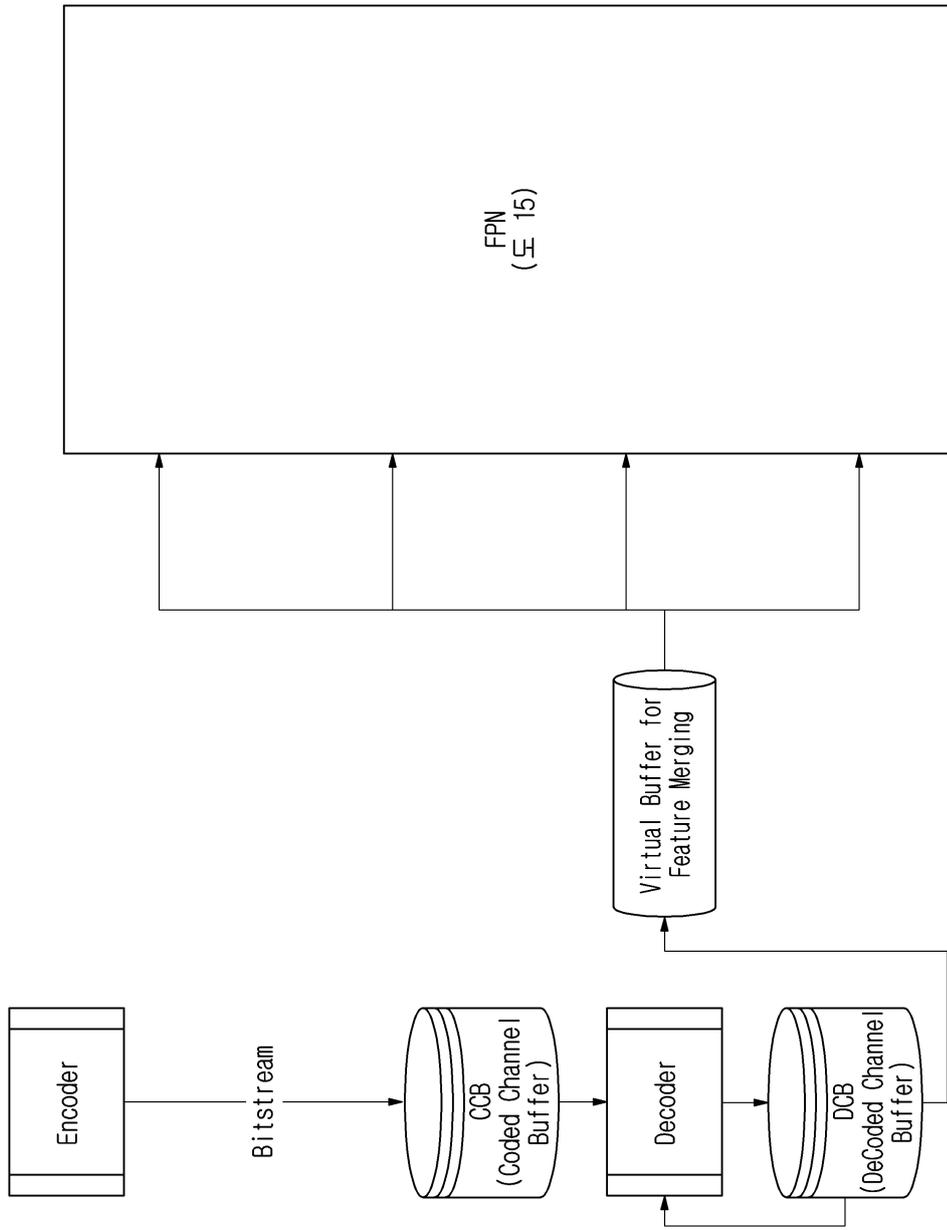
도면21



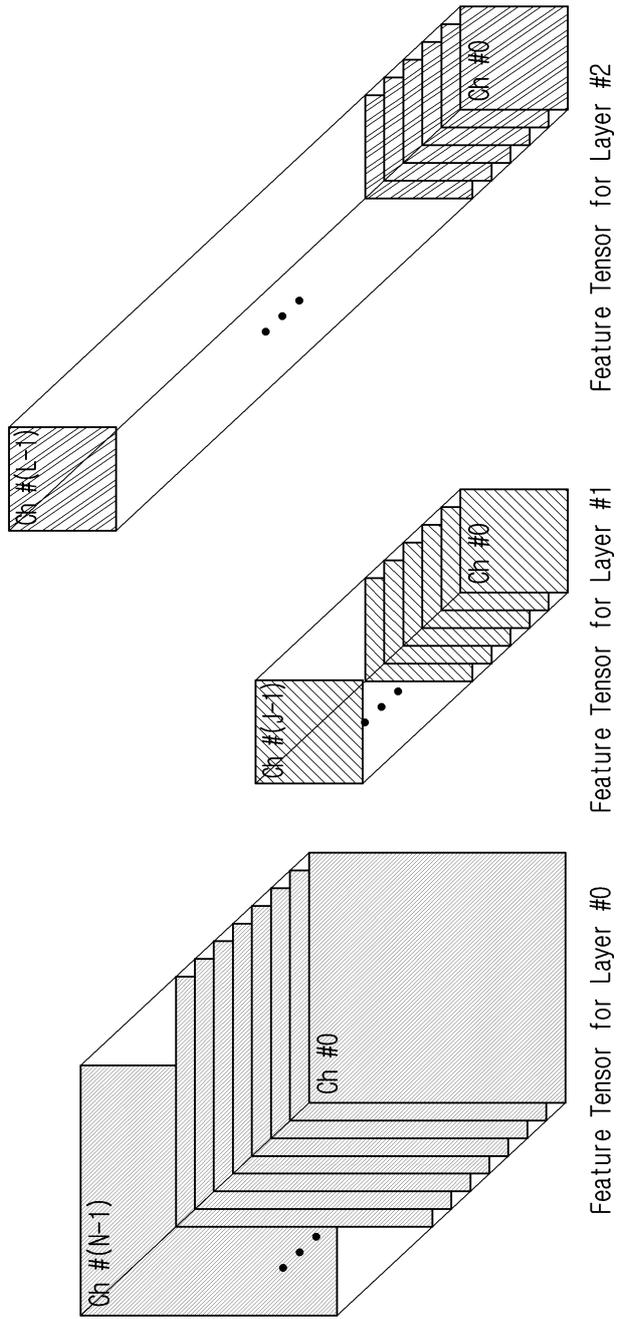
도면22



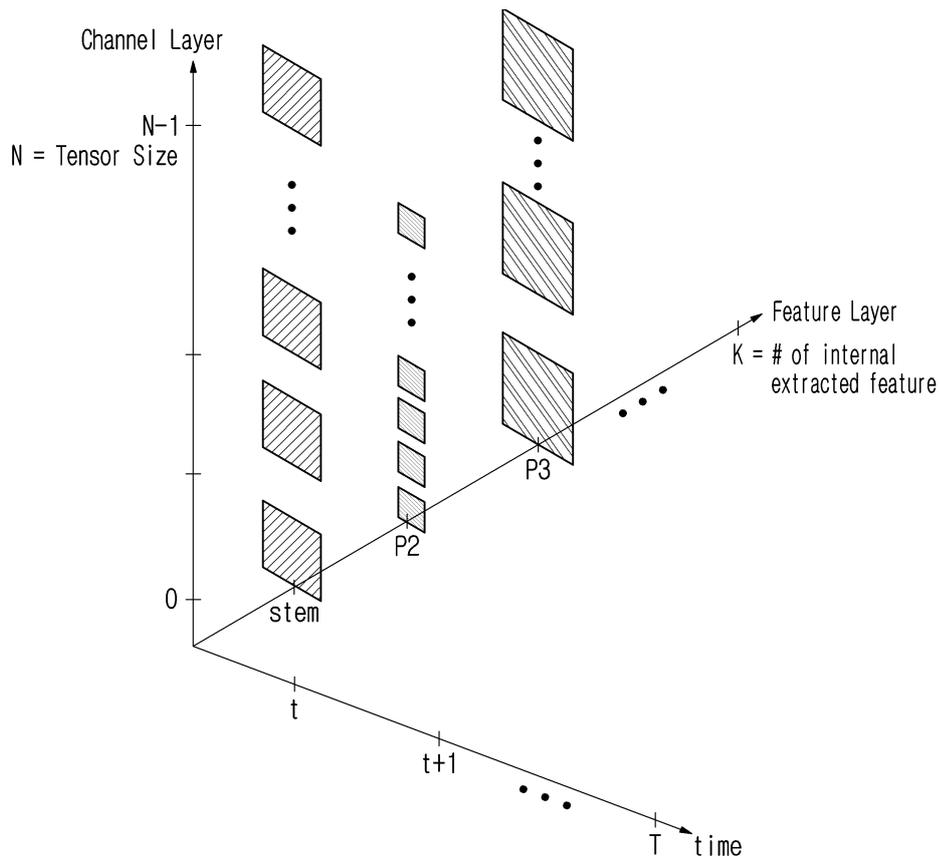
도면23



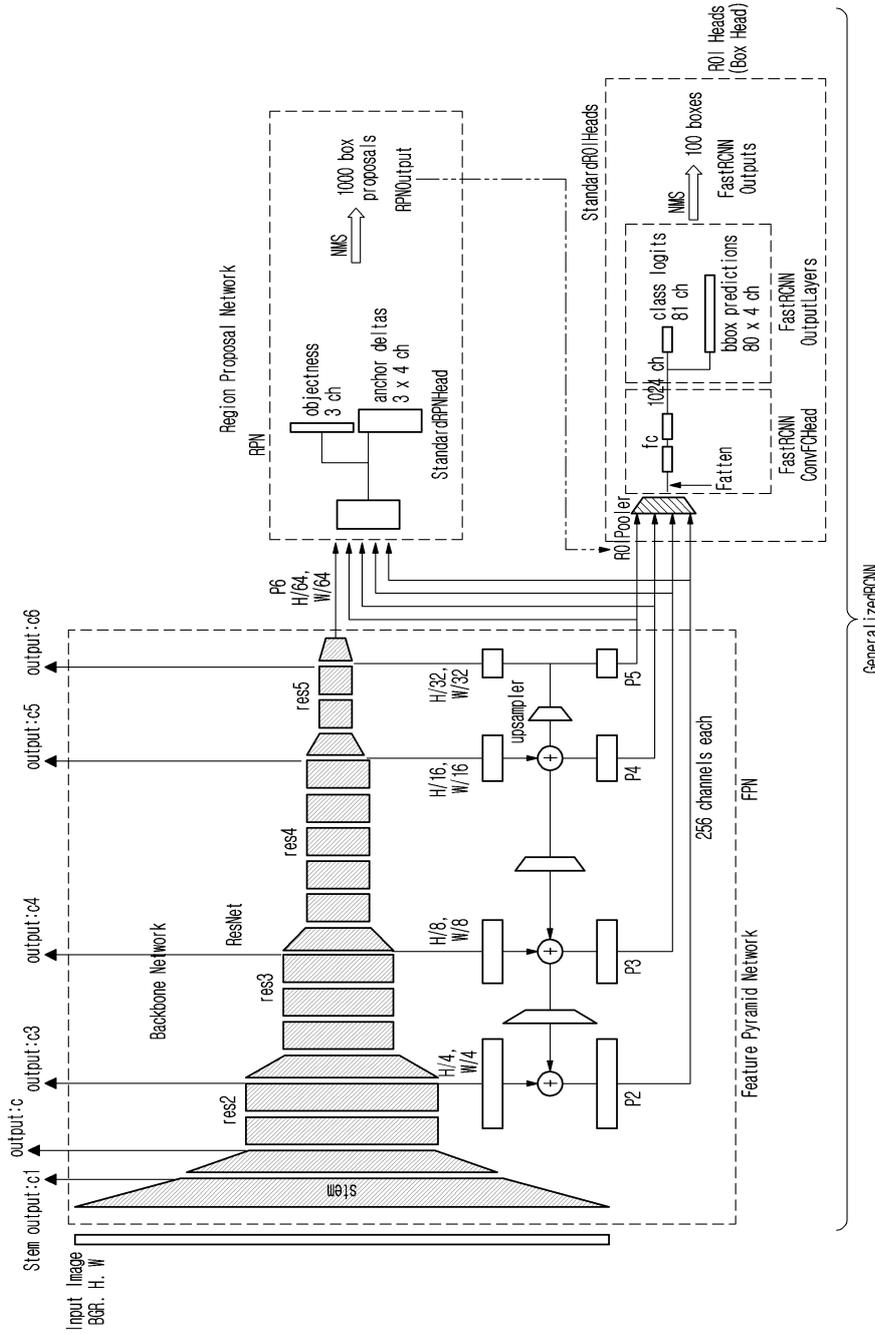
도면24



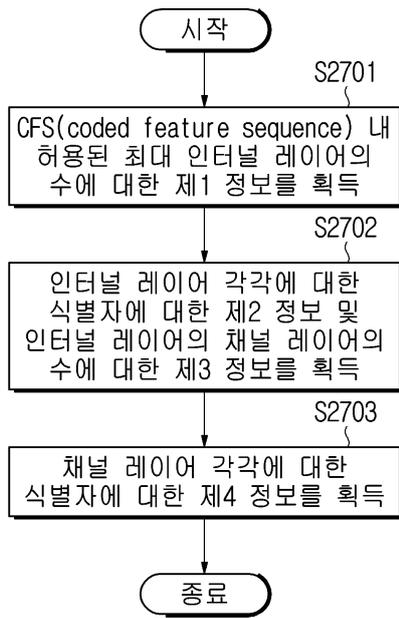
도면25



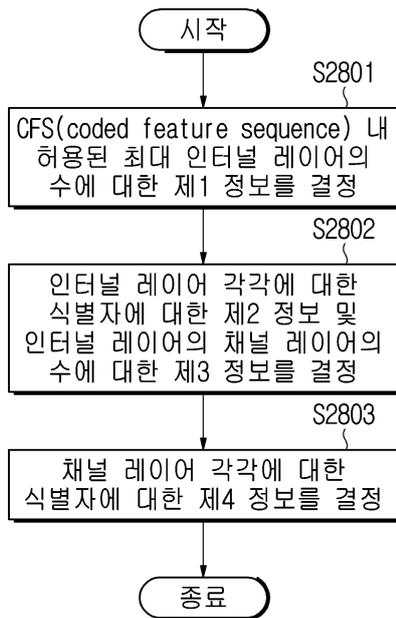
도면26



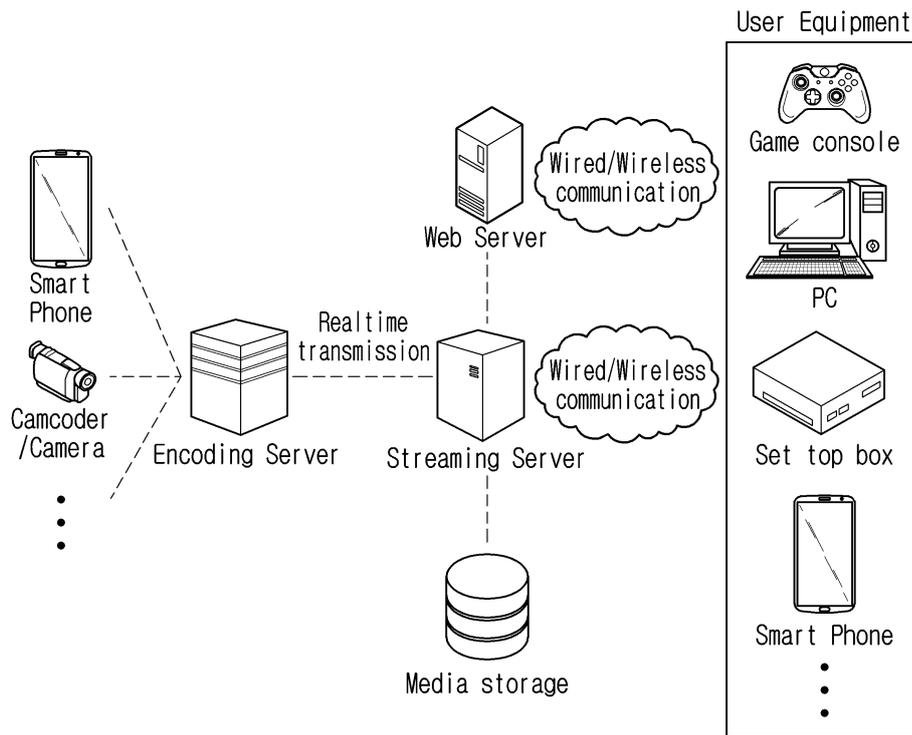
도면27



도면28



도면29



도면30

