



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0051076
(43) 공개일자 2024년04월19일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>H04N 19/44</i> (2014.01) <i>G06N 3/0464</i> (2023.01)
 <i>G06V 10/77</i> (2022.01) <i>G06V 10/82</i> (2022.01)
 <i>G06V 20/40</i> (2022.01) <i>H04N 19/189</i> (2014.01)</p> <p>(52) CPC특허분류
 <i>H04N 19/44</i> (2015.01)
 <i>G06N 3/0464</i> (2023.01)</p> <p>(21) 출원번호 10-2023-7044890
 (22) 출원일자(국제) 2022년06월03일
 심사청구일자 없음
 (85) 번역문제출일자 2023년12월26일
 (86) 국제출원번호 PCT/US2022/032048
 (87) 국제공개번호 WO 2022/260934
 국제공개일자 2022년12월15일
 (30) 우선권주장
 63/197,834 2021년06월07일 미국(US)</p> | <p>(71) 출원인
 오피 솔루션즈, 엘엘씨
 미국 01002 매사추세츠주 암허스트 미들 스트리트 368</p> <p>(72) 발명자
 칼바, 하리
 미국, 플로리다 33431, 보카 러톤, 글라데스 로드 777
 푸르트, 보리보예
 미국, 플로리다 33431, 보카 러톤, 글라데스 로드 777
 아지치, 벨리보르
 미국, 조지아 30144, 캔턴, 벨리 코트 501</p> <p>(74) 대리인
 특허법인 티앤아이</p> |
|---|---|

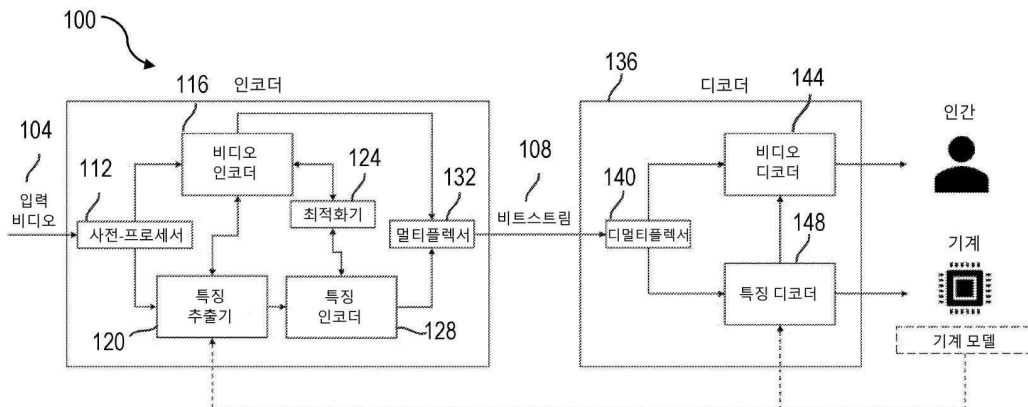
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 기계용 비디오 코딩(VCM)을 위한 인코더 및 디코더

(57) 요약

기계용 비디오 코딩(video coding for machines; VCM) 인코더는 제1 비디오 인코더를 포함하며, 제1 비디오 인코더는 입력 비디오를 비트스트림으로 인코딩하도록 구성된다. VCM 인코더는 특징 추출기를 포함하며, 특징 추출기는 입력 비디오에서 적어도 하나의 특징을 검출하도록 구성된다. VCM 인코더는 제2 인코더를 포함하며, 제2 인코더는 입력 비디오 및 적어도 하나의 특징의 함수로서 특징 비트스트림을 인코딩하도록 구성된다.

대표도



(52) CPC특허분류

G06V 10/7715 (2023.08)

G06V 10/82 (2022.01)

G06V 20/46 (2022.01)

H04N 19/189 (2015.01)

명세서

청구범위

청구항 1

기계용 비디오 코딩(video coding for machines; VCM) 인코더로서, 상기 VCM은,

제1 비디오 인코더로서, 상기 제1 비디오 인코더는 입력 비디오를 비트스트림으로 인코딩하도록 구성된, 상기 제1 비디오 인코더;

특징 추출기로서, 상기 특징 추출기는 상기 입력 비디오에서 적어도 하나의 특징을 검출하도록 구성된, 상기 특징 추출기; 및

제2 인코더로서, 상기 제2 인코더는 상기 입력 비디오 및 적어도 하나의 특징의 함수로서 특징 비트스트림을 인코딩하도록 구성된, 상기 제2 인코더를 포함하는, VCM 인코더.

청구항 2

제1항에 있어서, 상기 특징 추출기는 적어도 하나의 특징 맵을 출력하도록 구성된 기계-학습 모델을 더 포함하는, VCM 인코더.

청구항 3

제2항에 있어서, 상기 기계-학습 모델은 컨볼루션(convolutional) 신경망을 더 포함하는, VCM 인코더.

청구항 4

제3항에 있어서, 상기 컨볼루션 신경망은, 복수의 컨볼루션 계층들; 및

복수의 풀링 계층들을 포함하는, VCM 인코더.

청구항 5

제2항에 있어서, 상기 특징 추출기는 분류기를 더 포함하며, 상기 분류기는 상기 기계-학습 모델의 출력을 적어도 하나의 특징으로 분류하도록 구성되는, VCM 인코더.

청구항 6

제5항에 있어서, 상기 분류기는 심층 신경망을 더 포함하는, VCM 인코더.

청구항 7

제5항에 있어서, 상기 제2 인코더는 분류에 따른 상기 적어도 하나의 특징 맵의 특징 맵들을 적어도 하나의 특징으로 그룹화하도록 더 구성되는, VCM 인코더.

청구항 8

제1항에 있어서, 상기 제2 인코더는 특징 인코더를 더 포함하는, VCM 인코더.

청구항 9

제1항에 있어서, 상기 제2 인코더는 비디오 인코더를 더 포함하는, VCM 인코더.

청구항 10

제1항에 있어서, 상기 제1 비디오 인코더는 상기 특징 추출기에 결합되어 이로부터 특징 신호를 수신하는, VCM 인코더.

청구항 11

제1항에 있어서, 멀티플렉서를 더 포함하며, 상기 멀티플렉서는 상기 비디오 비트스트림과 상기 특징 비트스트림을 결합하도록 구성되는, VCM 인코더.

청구항 12

제1항에 있어서, 상기 특징 추출기는 복수의 특징 맵들을 생성하도록 구성되며, 상기 특징 맵들은 인코딩 이전에 공간적으로 배열되는, VCM 인코더.

청구항 13

제12항에 있어서, 상기 특징 맵들은 상기 특징 맵들의 텍스처 성분에 적어도 부분적으로 기초하여 공간적으로 배열되는, VCM 인코더.

청구항 14

인코딩된 하이브리드 비트스트림을 수신하도록 구성된 VCM 디코더로서, 상기 디코더는,

상기 하이브리드 비트스트림을 수신하는 디멀티플렉서;

특징 디코더로서, 상기 특징 디코더는 상기 디멀티플렉서로부터 인코딩된 특징 비트스트림을 수신하고, 기계 프로세싱을 위해 특징들의 디코딩된 세트를 제공하는, 상기 특징 디코더;

상기 특징 디코더에 결합된 기계 모델; 및

비디오 디코더로서, 상기 비디오 디코더는 상기 디멀티플렉서로부터 인코딩된 비디오 스트림을 수신하고 인간 소비를 위한 디코딩된 비디오 신호를 제공하는, 상기 비디오 디코더를 포함하는, VCM 디코더.

청구항 15

제14항에 있어서, 상기 특징 디코더는, 복수의 공간적으로 배열된 특징 맵들을 포함하는 비트스트림을 수신하고, 상기 공간적으로 배열된 특징 맵들을 디코딩하며, 특징 맵들의 원본 시퀀스를 재구성하도록 구성되는, VCM 디코더.

발명의 설명

기술 분야

[0001] 본 발명은 전반적으로 비디오 인코딩 및 디코딩 분야에 관한 것이다. 구체적으로, 본 발명은 기계용 비디오 코딩(video coding for machines; VCM)을 위한 인코더 및 디코더에 관한 것이다.

배경 기술

[0002] 비디오 코덱은, 디지털 비디오를 압축하거나 또는 압축해제하는 전자 회로 또는 소프트웨어를 포함할 수 있다. 비디오 코덱은 압축되지 않은 비디오를 압축된 포맷으로 변환하거나 또는 그 반대로 변환할 수 있다. 비디오 압축의 맥락에서, 비디오를 압축하는 (및/또는 이의 일부 기능을 수행하는) 디바이스는 전형적으로 인코더로 지칭될 수 있으며, 비디오를 압축해제하는 (및/또는 이의 일부 기능을 수행하는) 디바이스는 디코더로 지칭될 수 있다.

[0003] 압축된 데이터의 포맷은 표준 비디오 압축 사양을 따를 수 있다. 압축은, 압축된 비디오가 원본 비디오 내에 존재하는 일부 정보를 결여한다는 점에서 손실이 있을 수 있다. 그 결과는, 압축해제된 비디오가 압축되지 않은 원본 비디오보다 더 낮은 품질을 가질 수 있다는 것을 포함할 수 있으며, 이는, 원본 비디오를 정확하게 재구성하기에는 정보가 불충분하기 때문이다.

[0004] 비디오 품질, 비디오를 표현하기 위해 사용되는 데이터의 양(예를 들어, 비트 레이트에 의해 결정됨), 인코딩 및 디코딩 알고리즘들의 복잡성, 데이터 손실 및 오류들에 대한 민감도, 편집의 용이성, 랜덤 액세스, 중단-대-중단 지연(예를 들어, 레이턴시(latency)), 및 유사한 것 사이에 복잡한 관계들이 있을 수 있다.

[0005] 모션 보상은, 이전 및/또는 미래 프레임들과 같은 참조 프레임이 주어지면, 비디오 내의 카메라 및/또는 객체들의 모션을 고려함으로써 비디오 프레임 또는 이의 일 부분을 예측하기 위한 접근방식을 포함할 수 있다. 모션 보상은, 비디오 압축에 대한 비디오 데이터의 인코딩 및 디코딩에서, 예를 들어, 동화상 전문가 그룹(Motion

Picture Experts Group; MPEG)의 어드밴스드 비디오 코딩(advanced video coding; AVC) 표준(H.264로도 지칭됨)을 사용하는 인코딩 및 디코딩에서 이용될 수 있다. 모션 보상은 참조 픽처(picture)를 현재 픽처로 변환하는 측면에서 픽처를 설명할 수 있다. 참조 픽처는 현재 픽처와 비교할 때 시간 상 이전일 수 있으며, 또는 현재 픽처와 비교할 때 미래일 수 있다. 이미지들이 이전에 송신된 및/또는 저장된 이미지들로부터 정확하게 합성될 때, 압축 효율이 개선될 수 있다.

[0006] 비디오는 전통적으로 인간이 소비하기 위한 미디어였으며, 비디오 압축 방법론들은 압축해제 이후 인간 시청자에 의해 인지되는 비디오의 충실도를 유지하는 데 초점을 맞췄다. 그러나, 오늘날, 방대한 양의 비디오는 기계들에 의해 분석되고 있다. 이와 같이, 기계 분석을 위해 최적화된 비디오 압축 방법들을 개발하고 최적화할 필요성이 커지고 있다. 애플리케이션에 의존하여, 기계는 분석 및 기능을 수행하기 위해 비디오 콘텐츠와 동일한 정보를 필요로 하지 않을 수 있다. 대신, 비디오 신호 내의 특정 특징들로 충분할 수 있다. 기계용 비디오 코딩(Video coding for machines; "VCM")은, 전통적인 비디오 스트림뿐만 아니라 기계 분석에 매우 적합한 비디오 스트림으로부터 추출된 특징들 둘 모두를 압축함으로써 압축된 비트스트림을 생성하는 접근방식이다.

발명의 내용

해결하려는 과제

과제의 해결 수단

- [0007] 인코더 및 디코더를 포함하는 기계용 비디오 코딩(video coding for machines; VCM)을 위한 시스템이 제공된다. VCM 인코더는, 바람직하게는 입력 비디오를 비트스트림으로 인코딩하도록 구성된 제1 비디오 인코더를 포함한다. VCM 인코더는, 입력 비디오에서 적어도 하나의 특징을 검출하도록 구성된 특징 추출기를 더 포함한다. 제2 인코더는 적어도 하나의 특징 및 입력 비디오의 합수로서 특징 비트스트림을 인코딩하도록 구성된다.
- [0008] 일부 실시예들에서, 비디오 디코더는 특징 추출기에 결합되어 이로부터 특징 신호를 수신한다. 바람직하게는, 기계 모델은 특징 추출기에 포함되거나 또는 이에 제공될 수 있다. 멀티플렉서는 인코딩된 비디오 및 특징 신호들을 디코더로 송신하기 위한 비트스트림으로 결합하기 위해 제공될 수 있다.
- [0009] 일부 바람직한 실시예들에서, 특징 추출기는 적어도 하나의 특징 맵을 출력하도록 구성된 기계-학습 모델을 더 포함한다. 더 바람직하게는, 기계-학습 모델은 컨볼루션(convolutional) 신경망을 더 포함할 수 있다. 일부 실시예들에서, 컨볼루션 신경망은 복수의 컨볼루션 계층들 및 복수의 풀링 계층들을 포함한다.
- [0010] 특징 추출기는 기계-학습 모델의 출력을 적어도 하나의 특징으로 분류하도록 구성된 분류기를 더 포함할 수 있다. 특정 실시예들에서, 분류기는 심층 신경망이다.
- [0011] 특징 추출기는, 복수의 특징 맵들을 생성하고, 인코딩 이전에 복수의 특징 맵들 중 적어도 일 부분을 공간적으로 배열하도록 구성될 수 있다. 특징 맵들은 텍스처와 같은 특징 맵들의 파라미터들에 기초하여 공간적으로 배열될 수 있다.
- [0012] 또 다른 실시예에서, 제2 인코더는 적어도 하나의 특징의 분류에 따라 특징 맵들을 그룹화하도록 더 구성될 수 있다.
- [0013] VCM 디코더는 인코딩된 하이브리드 비트스트림을 수신하도록 구성된다. VCM 디코더는, 하이브리드 비트스트림을 수신하고 이로부터 비디오 비트스트림과 특징 비트스트림을 제공하는 디멀티플렉서를 포함한다. 특징 디코더가 제공된다. 특징 디코더는 디멀티플렉서로부터 인코딩된 특징 비트스트림을 수신하고, 기계 프로세싱을 위해 특징들의 디코딩된 세트를 제공한다. 기계 모델은 바람직하게는 특징 디코더에 결합된다. 비디오 디코더는, 디멀티플렉서로부터 인코딩된 비디오 스트림을 수신하고 인간 소비를 위해 적절한 디코딩된 비디오 신호를 제공하기 위해 제공된다.
- [0014] 일부 실시예들에서, VCM 디코더는 복수의 공간적으로 배열된 특징 맵들을 포함하는 비트스트림을 수신하고, 공간적으로 배열된 특징 맵들을 디코딩하며, 특징 맵들의 원본 시퀀스를 재구성하도록 구성될 수 있다.
- [0015] 본 발명의 비제한적인 실시예들의 이러한 그리고 다른 측면들과 특징들은 첨부된 도면과 함께 본 발명의 특정한 비제한적인 실시예들의 다음의 설명을 검토할 때 당업자들에게 명백해질 것이다.

도면의 간단한 설명

- [0016] 본 발명을 예시하기 위한 목적을 위해, 도면들은 본 발명의 하나 이상의 실시예들의 측면들을 도시한다. 그러나, 본 발명이 도면들에 도시된 정확한 배열들 및 수단들로 한정되지 않는다는 것이 이해되어야 한다.
- 도 1은 VCM 인코더 및 디코더 시스템의 예시적인 실시예를 예시하는 간략화된 블록도이다.
- 도 2는 VCM 인코더 및 디코더 시스템의 예시적인 실시예를 예시하는 간략화된 블록도이다.
- 도 3은 기계 모델의 예시적인 실시예를 도식적으로 예시하는 개략도이다.
- 도 4는 컨볼루션 유닛들의 조립을 위한 프로세스의 예시적인 실시예를 예시하는 도면이다.
- 도 5는 비디오의 예시적인 실시예를 코딩 유닛 및 컨볼루션 유닛 표현들로 예시하는 블록도이다.
- 도 6은 인터 예측 프로세스의 예시적인 실시예를 예시하는 블록도이다.
- 도 7은 인트라 예측 프로세스의 예시적인 실시예를 예시하는 블록도이다.
- 도 8은 비디오 입력 픽처 및 대응하는 컨볼루션 계층의 예시적인 실시예를 예시하는 개략도이다.
- 도 9는 비디오 디코더의 예시적인 실시예를 예시하는 간략화된 블록도이다.
- 도 10은 비디오 인코더의 예시적인 실시예를 예시하는 블록도이다.
- 도 11은 기계-학습 모듈의 예시적인 실시예를 예시하는 블록도이다.
- 도 12는, 본 명세서에서 개시되는 방법론들 중 임의의 하나 이상 및 이의 임의의 하나 이상의 부분들을 구현하기 위해 사용될 수 있는 컴퓨팅 시스템의 블록도이다.
- 도면들이 반드시 축적이 맞춰질 필요는 없으며, 가상 라인들, 도식적 표현들 및 단편적 뷰(view)들에 의해 예시될 수 있다. 특정 경우들에서, 실시예들의 이해를 위해 필요하지 않거나 도는 다른 세부사항들을 인지하기 어렵게 만드는 세부사항들을 생략될 수 있다.

발명을 실시하기 위한 구체적인 내용

- [0017] 실시예들에서, VCM 인코더는 비디오 또는 하이브리드 모드로 동작하는 것이 가능할 수 있다. 어느 모드에서든, VCM 인코더는 사람이 볼 수 있는 출력 비디오로서 디코딩될 비디오 비트스트림을 제공할 수 있다.
- [0018] 이제 도 1을 참조하면, 기계용 비디오 코딩(video coding for machines; VCM)을 위한 인코더의 예시적인 실시예가 예시된다. VCM 인코더(100)는 비제한적으로 디지털 및/또는 아날로그 회로부를 포함하는 임의의 회로부를 사용하여 구현될 수 있다. VCM 인코더(100)는 하드웨어 구성, 소프트웨어 구성, 펌웨어 구성 및/또는 이들의 임의의 조합을 사용하여 구성될 수 있다. VCM 인코더(100)는, 비제한적으로 이하에서 설명되는 바와 같은 임의의 컴퓨팅 디바이스를 포함할 수 있는, 컴퓨팅 디바이스로서 및/또는 컴퓨팅 디바이스의 구성요소로서 구현될 수 있다. 일 실시예에서, VCM 인코더(100)는 입력 비디오(104)를 수신하고 출력 비트스트림(108)을 생성하도록 구성될 수 있다. 입력 비디오(104)의 수신은 이하에서 설명되는 임의의 방식으로 달성될 수 있다. 비트스트림은, 비제한적으로, 이하에서 설명되는 바와 같은 임의의 비트스트림을 포함할 수 있다. VCM 인코더(100)는, 비제한적으로, 사전-프로세서(112), 비디오 인코더(116), 특징 추출기(120), 최적화기(124), 특징 인코더(128), 및/또는 멀티플렉서(132)를 포함할 수 있다. 사전-프로세서(112)는 입력 비디오(104) 스트림을 수신하고 스트림의 비디오, 오디오 및 메타데이터 서브-스트림들을 파싱(parse)할 수 있다. 사전-프로세서(112)는 이하에서 더 상세하게 설명되는 바와 같은 디코더를 포함하거나 및/또는 이와 통신할 수 있다. 다시 말해서, 사전-프로세서(112)는 입력 스트림들을 디코딩하기 위한 능력을 가질 수 있다. 이는, 비제한적인 예에서, 입력 비디오(104)의 디코딩을 가능하게 할 수 있으며, 이는 하류의 픽셀-도메인 분석을 용이하게 할 수 있다.
- [0019] 도 1을 더 참조하면, VCM 인코더(100)는 하이브리드 모드로 및/또는 비디오 모드로 동작할 수 있다. 하이브리드 모드에 있을 때, VCM 인코더(100)는, 인간 소비들을 위해 의도된 시각적 신호를 인코딩하고 기계 소비들을 위해 의도된 특징 신호를 인코딩하도록 구성될 수 있다. 기계 소비들은, 비제한적으로 이하에서 더 상세하게 설명되는 바와 같은 컴퓨팅 디바이스들을 포함하는, 임의의 디바이스들 및/또는 구성요소들을 비제한적으로 포함할 수 있다. 입력 신호는, 예를 들어, 하이브리드 모드에 있을 때, 사전-프로세서(112)를 통해 전달될 수 있다.
- [0020] 계속해서 도 1을 참조하면, 인간 소비를 위한 비디오 신호를 인코딩하는 비디오 인코더(116)는, 비제한적으로,

이하에서 더 상세하게 설명되는 바와 같은 임의의 비디오 인코더(116)를 포함할 수 있다. VCM 인코더(100)가 하이브리드 모드에 있을 때, VCM 인코더(100)는 수정되지 않은 입력 비디오(104)를 비디오 인코더(116)에 전송하고, 동일한 입력 비디오(104)의 카피 및/또는 어떤 방식으로 수정된 입력 비디오(104)를 특징 추출기(120)로 전송할 수 있다. 입력 비디오(104)에 대한 수정들은, 본 개시내용의 전체를 검토할 때 당업자들에게 일어날 수 있는 임의의 스케일링, 변환 또는 다른 수정을 포함할 수 있다. 예를 들어, 그리고 비제한적으로, 입력 비디오(104)가 더 낮은 해상도로 리사이징(resize)될 수 있거나, 입력 비디오(104)의 픽처들의 시퀀스 중 특정한 수의 픽처들이 폐기되어 입력 비디오(104)의 프레임레이트를 감소시키거나, 컬러 정보가, 예를 들어 그리고 비제한적으로, RGB 비디오를 그레이스케일 비디오로 변환함으로써 수정될 수 있거나 하는 등이다.

[0021] 계속해서 도 1을 참조하면, 비디오 인코더(116)와 특징 추출기(120)가 연결되며, 양 방향으로 유용한 정보를 교환할 수 있다. 예를 들어, 그리고 비제한적으로, 비디오 인코더(116)는 모션 추정 정보를 특징 추출기(120)로 전달할 수 있으며, 그 반대도 마찬가지이다. 비디오 인코더(116)는, 비디오 인코더(116) 및/또는 특징 추출기(120)가 식별할 수 있는 관심 영역(regions of interest; ROI)에 기초하여 양자화 매핑 및/또는 이를 설명하는 데이터를 특징 추출기(120)에 제공할 수 있거나, 또는 그 반대일 수 있다. 비디오 인코더(116)는 입력 비디오(104), 입력 신호, 및/또는 임의의 프레임 및/또는 이의 서브프레임에 존재하거나 및/또는 식별된 특징들을 기초하여 하나 이상의 분할 결정들을 설명하는 데이터를 특징 추출기(120)에 제공할 수 있으며; 특징 추출기(120)는 입력 비디오(104), 입력 신호, 및/또는 임의의 프레임 및/또는 이의 서브프레임에 존재하거나 및/또는 식별된 특징들을 기초하여 하나 이상의 분할 결정들을 설명하는 데이터를 비디오 인코더(116)에 제공할 수 있다. 비디오 인코더(116)와 특징 추출기(120)는 픽처들의 최적 그룹(optimal group of pictures; GOP) 결정들에 대한 시간 정보를 서로 공유하거나 및/또는 송신할 수 있다. 이러한 기술들 및/또는 프로세스들 각각은, 비제한적으로, 이하에서 더 상세하게 설명되는 바와 같이 수행될 수 있다.

[0022] 계속해서 도 1을 참조하면, 특징 추출기(120)는 오프라인 모드 또는 온라인 모드로 동작할 수 있다. 특징 추출기(120)는 특징들을 식별하거나 및/또는 그렇지 않으면 특징들에 작용하거나 및/또는 특징들을 조작할 수 있다. 본 개시내용에서 사용되는 바와 같은 "특징"은 데이터의 특정 구조 및/또는 콘텐츠 속성이다. 특징들의 예들은 스케일 불변 특징 변환(scale invariant feature transform; SIFT)들, 오디오 특징들, 컬러 히스토그램(hist), 모션 히스토그램, 스피치 레벨, 라우트니스 레벨, 또는 유사한 것을 포함할 수 있다. 특징들은 타임 스탬핑될 수 있다. 각각의 특징은 프레임들의 그룹의 단일 프레임과 연관될 수 있다. 특징들은 타임스탬프들, 비디오 내의 사람들 및 객체들에 대한 라벨들, 객체들 및/또는 관심 영역들에 대한 좌표들, 영역-기반 양자화에 대한 프레임 마스크들, 및/또는 본 개시내용의 전체를 검토할 때 당업자들에게 일어날 수 있는 임의의 다른 특징과 같은 고 레벨 콘텐츠 특징들을 포함할 수 있다. 추가적인 비제한적 예로서, 특징들은, 프레임 또는 프레임들의 그룹의 공간적 및/또는 시간적 특성들을 설명하는 특징들을 포함할 수 있다. 공간적 및/또는 시간적 특성들을 설명하는 특징들의 예들은 모션, 텍스처, 컬러, 밝기, 에지 카운트, 블러(blur), 블록성(blockiness), 또는 유사한 것을 포함할 수 있다. 오프라인 모드에 있을 때, 이하에서 더 상세하게 설명되는 바와 같은 모든 기계 모델들은 인코더에 및/또는 인코더의 메모리 및/또는 인코더에 액세스가능한 메모리에 저장될 수 있다. 이러한 모델들의 예들은, 비제한적으로, 전체적 또는 부분적 컨볼루션 신경망들, 키폰트 추출기들, 에지 검출기들, 돌출맵 생성기(saliency map constructor)들, 또는 유사한 것을 포함할 수 있다. 온라인 모드에 있을 때, 하나 이상의 모델들은 원격 기계에 의해 실시간으로 또는 추출 전 어떤 시점에 특징 추출기(120)로 통신될 수 있다.

[0023] 계속해서 도 1을 참조하면, 특징 인코더(128)는, 예를 들어 그리고 비제한적으로, 특징 추출기(120)에 의해 생성되는 바와 같은 특징 신호를 인코딩하도록 구성된다. 일 실시예에서, 특징들을 추출한 이후에, 특징 추출기(120)는 추출된 특징들을 특징 인코더(128)로 전달할 수 있다.

[0024] 특징 인코더(128)는 멀티플렉서(132)로 전달될 수 있는 특징 스트림을 생성하기 위해, 예를 들어 그리고 비제한적으로 이하에서 설명되는 바와 같은 엔트로피 인코딩 및/또는 유사한 기술들을 사용할 수 있다.

[0025] 비디오 인코더(116) 및/또는 특징 인코더(128)는 최적화기(124)를 통해 연결될 수 있으며; 최적화기(124)는 이러한 비디오 인코더(116)와 특징 인코더(128) 사이에서 유용한 정보를 교환할 수 있다. 예를 들어 그리고 비제한적으로, 엔트로피 코딩에 대한 코드워드 구성 및/또는 길이에 관한 정보는, 최적 압축을 위해, 최적화기(124)를 통해 교환되고 재사용될 수 있다.

[0026] 일 실시예에서, 그리고 계속해서 도 1을 참조하면, 비디오 인코더(116)는 비디오 스트림을 생성할 수 있으며; 비디오 스트림은 멀티플렉서(132)로 전달될 수 있다. 멀티플렉서(132)는 비디오 스트림을 특징 인코더(128)에 의해 생성된 특징 스트림과 멀티플렉싱할 수 있다. 대안적으로 또는 추가적으로, 비디오 및 특징 비트스트림들

은 별개의 채널들, 별개의 네트워크들을 통해 별개의 디바이스로 및/또는 별개의 시간들 또는 별개의 간격들(시간 멀티플렉싱)로 송신될 수 있다. 비디오 스트림 및 특징 스트림 각각은 본 개시내용에서 설명되는 바와 같은 임의의 비트스트림의 구현에 적절한 임의의 방식으로 구현될 수 있다. 일 실시예에서, 멀티플렉싱된 비디오 스트림 및 특징 스트림은 하이브리드 비트스트림을 생성할 수 있으며, 이는 이하에서 더 상세하게 설명되는 바와 같이 송신될 수 있다.

[0027] 계속해서 도 1을 참조하면, VCM 인코더(100)가 비디오 모드에 있을 때, VCM 인코더(100)는 비디오 및 특징 인코딩 둘 모두를 위해 비디오 인코더(116)를 사용할 수 있다. 특징 추출기(120)는 특징들을 비디오 인코더(116)로 송신할 수 있으며; 비디오 인코더(116)는 특징들을 대응하는 비디오 디코더(144)에 의해 디코딩될 수 있는 비디오 스트림으로 인코딩할 수 있다. VCM 인코더(100)가 비디오 인코딩 및 특징 인코딩 둘 모두를 위해 단일 비디오 인코더(116)를 사용할 수 있고, 이러한 경우에, 비디오 및 특징들에 대해 파라미터들의 상이한 세트를 사용할 수 있으며; 대안적으로, VCM 인코더(100)가 병렬로 동작할 수 있는 2개의 별도의 비디오 인코더들(116)을 사용할 수 있다는 것을 유의해야 한다.

[0028] 계속해서 도 1을 참조하면, 시스템(100)은 VCM 디코더(136)를 포함하거나 및/또는 이와 통신할 수 있다. VCM 디코더(136) 및/또는 이의 요소들은 이상에서 설명된 바와 같은 VCM 인코더(100)의 구성을 위해 적절한 임의의 회로부 및/또는 유형의 구성을 사용하여 구현될 수 있다. VCM 디코더(136)는, 비제한적으로, 디멀티플렉서(140)를 포함할 수 있다. 디멀티플렉서(140)는, 비트스트림이 이상에서 설명된 바와 같이 멀티플렉싱된 경우 비트스트림들을 디멀티플렉싱하도록 동작할 수 있다. 예를 들어 그리고 비제한적으로, 디멀티플렉서(140)는 하나 이상의 비디오 비트스트림들 및 하나 이상의 특징 비트스트림들을 포함하는 멀티플렉싱된 비트스트림을 별도의 비디오 비트스트림 및 특징 비트스트림으로 분리할 수 있다.

[0029] 계속해서 도 1을 참조하면, VCM 디코더(136)는 비디오 디코더(144)를 포함할 수 있다. 비디오 디코더(144)는, 비제한적으로 이하에서 더 상세하게 설명되는 바와 같이 디코더에 대해 적절한 임의의 방식으로 구현될 수 있다. 일 실시예에서, 그리고 비제한적으로, 비디오 디코더(144)는, 사람 또는 시각적 감각 능력들을 갖는 다른 생물 및/또는 디바이스가 볼 수 있는 출력 비디오를 생성할 수 있다.

[0030] 계속해서 도 1을 참조하면, VCM 디코더(136)는 특징 디코더(148)를 포함할 수 있다. 일 실시예에서, 그리고 비제한적으로, 특징 디코더(148)는 하나 이상의 디코딩된 데이터를 기계에 제공하도록 구성될 수 있다. 기계는, 비제한적으로 임의의 마이크로 제어기, 프로세서, 내장 시스템, 시스템 온 칩, 네트워크 노드, 또는 유사한 것을 포함하는, 임의의 컴퓨팅 디바이스를 비제한적으로 포함할 수 있다. 기계는, 이하에서 더 상세하게 설명되는 바와 같이 기계 모델을 동작시키거나, 저장하거나, 트레이닝시키거나, 이로부터 입력을 수신하거나, 이에 대한 출력을 생성하거나, 및/또는 그렇지 않으면 이와 상호작용할 수 있다. 기계는 프로세싱 및 통신 구성요소들을 갖는 객체들의 네트워크로서 정의된 사물 인터넷(Internet of Things; IOT)에 포함될 수 있으며, 이들 중 일부는 데스크탑 컴퓨터들, 랩탑 컴퓨터들, 및/또는 모바일 디바이스들과 같은 기존 컴퓨팅 디바이스들이 아닐 수 있다. IoT 내의 객체들은, 비제한적으로, 내장 마이크로프로세서 및/또는 마이크로제어기와 근거리 네트워크(local area network; LAN) 및/또는 광역 네트워크(wide-area network; WAN)와 인터페이싱하기 위한 하나 이상의 구성요소들을 갖는 임의의 디바이스들을 포함할 수 있으며; 하나 이상의 구성요소들은, 비제한적으로, 전기 전자 기술자 협회(Institute of Electronic and Electrical Engineers; IEEE)에 의해 공표된 IEEE 802.15.4 표준의 ZIGBEE 사양 및/또는 프랑스 루에유말메종 소재의 Schneider Electric SE에 의해 공표된 MODBUS에 따라 동작하는 네트워크 통신 구성요소 및/또는, 워싱턴, 커클랜드 소재의 Bluetooth SIG, Inc.에 의해 공표된 바와 같은 프로토콜들을 따른 BLUETOOTH 트랜시버들과 같은, 예를 들어, 2.4-2.485 GHz 범위에서 통신하는 무선 트랜시버를 포함할 수 있다. 당업자들은, 본 개시내용의 전체를 검토할 때, 본 개시내용과 일치되게 이용될 수 있는 다양한 대안적인 또는 추가적인 통신 프로토콜들 및 이러한 프로토콜들을 지원하는 디바이스들을 인식할 수 있을 것이며, 이들 각각은 본 개시내용의 범위 내에 있는 것으로 고려된다.

[0031] 계속해서 도 1을 참조하면, VCM 인코더(100) 및/또는 VCM 디코더(136) 각각은, 임의의 순서로 그리고 임의의 반복의 정도로 본 개시내용에서 설명되는 임의의 실시예의 임의의 방법, 방법 단계, 또는 방법 단계들의 시퀀스를 수행하도록 설계되거나 및/또는 구성될 수 있다. 예를 들어, VCM 인코더(100) 및/또는 VCM 디코더(136) 각각은 희망되는 또는 명령된 결과가 달성될 때까지 단일 단계 또는 시퀀스를 반복적으로 수행하도록 구성될 수 있으며; 단계 또는 단계들의 시퀀스의 반복은 후속 반복들에 대한 입력들로서 이전 반복들의 출력들을 사용하여 반복적으로 및/또는 회귀적으로 수행될 수 있으며, 이는 반복들의 입력들 및/또는 출력들을 집성하여 집성 결과, 전역 변수들과 같은 하나 이상의 변수들의 감소 또는 감분, 및/또는 더 큰 프로세싱 태스크에 반복적으로 어드레싱되는 더 작은 프로세싱 태스크들의 세트로 분할하는 것을 생성한다. VCM 인코더(100) 및/또는 VCM 디코

터(136) 각각은, 2개 이상의 병렬 스레드들, 프로세서 코어들, 또는 유사한 것을 사용하여 단계를 2번 이상 동시에 및/또는 실질적으로 동시에 수행하는 것과 같이 본 개시내용에서 설명되는 바와 같은 임의의 단계 또는 단계들의 시퀀스를 병렬로 수행할 수 있으며: 병렬 스레드들 및/또는 프로세서들 사이의 태스크들의 분할은 반복들 사이에서의 태스크들의 분할에 적절한 임의의 프로토콜을 사용하여 수행될 수 있다. 당업자는, 본 개시내용의 전체를 검토할 때, 단계들, 단계들의 시퀀스들, 프로세싱 태스크들, 및/또는 데이터가 반복, 회귀, 및/또는 병렬 프로세싱을 사용하여 세분화되거나, 공유되거나, 또는 그렇지 않으면 처리되는 다양한 방법들을 인식할 것이다.

[0032] 이제 도 2를 참조하면, 비디오 모드로 동작하는 VCM 인코더(100)의 예시적인 실시예(200)가 예시된다. VCM 인코더(100)는, VCM 인코더(100)와 상호작용하는 사용자, 프로그램, 메모리 위치, 및/또는 하나 이상의 추가적인 디바이스들(미도시)로부터 명령을 수신할 때 비디오 모드와 하이브리드 모드 사이에서 스위칭하도록 구성될 수 있다. 비디오 모드에서, VCM 인코더(100)는 제2 비디오 인코더(204)를 동작시키도록 구성될 수 있다.

[0033] 제2 비디오 인코더(204)는, 이상에서 설명된 바와 같이 특정 인코더(128) 및/또는 비디오 인코더(116)에 대해 적절한 임의의 방식으로 특징들 및/또는 제2 비디오 스트림을 인코딩할 수 있다. 제2 비디오 인코더(204)는 특징 추출기(120) 및/또는 최적화기(124)로부터 데이터를 수신하거나 및/또는 이들로 데이터를 송신할 수 있다. 계속해서 도 2를 참조하면, 비디오 모드는 시각적 정보로서 표현될 수 있는 임의의 또는 모든 유형의 특징들을 인코딩하기 위해 사용될 수 있다. 비제한적인 예로서, 비디오 모드는 에지들, 라인들 또는 유사한 것을 나타내는 이미지와 같은 돌출 맵들, 필터링된 이미지들, 이하에서 더 상세하게 설명되는 바와 같은 컨볼루션 신경망들의 특징 맵들, 또는 유사한 것을 인코딩하기 위해 사용될 수 있다.

[0034] 이제 도 3을 참조하면, VCM 인코더(100)와 함께 사용하기 위한 컨볼루션 신경망(convolutional neural network; CNN)의 예시적인 실시예(300)가 예시된다. 본 개시내용에서 사용되는 바와 같은, 인공 신경망으로도 알려진 "신경망"은 "노드들"의 네트워크이거나, 또는 하나 이상의 입력들, 하나 이상의 출력들, 및 입력들에 기초하여 출력들을 결정하는 기능을 갖는 데이터 구조체들이다. 이러한 노드들은, 노드들의 입력 계층, 하나 이상의 중간 계층들, 및 노드들의 출력 계층을 포함하는, 비제한적으로 컨볼루션 신경망과 같은 네트워크로 조직될 수 있다. 노드들 사이의 연결들은 네트워크를 "트레이닝시키는" 프로세스를 통해 생성될 수 있으며, 이러한 프로세스에서 트레이닝 데이터세트로부터의 요소들이 입력 노드들에 적용되고, 그런 다음 출력 노드들에서 희망되는 값을 생성하기 위해서 신경망의 인접한 계층들의 노드들 사이의 연결들 및 가중치들을 조정하기 위해 적절한 트레이닝 알고리즘(예컨대 레벤버그-마쿼트(Levenberg-Marquardt), 공액 기울기(conjugate gradient), 모의 담금질(simulated annealing), 또는 다른 알고리즘들)이 사용된다. 이러한 프로세스는 때때로 심층 학습으로 지칭된다.

[0035] 계속해서 도 3을 참조하면, 노드는, 비제한적으로, 노드를 포함하는 신경망에 대한 입력으로부터 및/또는 다른 노드들로부터 수치 값들을 수신할 수 있는 복수의 입력들(X_i)을 포함할 수 있다. 노드는, 개별적인 입력들(X_i)과 곱해지는 가중치들(w_i)을 사용하여 입력들의 가중 합을 수행할 수 있다. 추가적으로 또는 대안적으로, 바이어스(b)는, 계층에 대한 입력과는 독립적인 신경망 계층의 각각의 유닛에 오프셋이 추가되도록 입력들의 가중 합에 추가될 수 있다. 그런 다음, 가중 합은, 하나 이상의 출력들(y)을 생성할 수 있는 함수(ϕ)로 입력될 수 있다. 입력(X_i)에 적용되는 가중치(w_i)는, 입력이, 예를 들어, 큰 수치 값을 갖는 대응하는 가중치에 의해 하나 이상의 출력들(y)에 대해 강한 영향을 갖는다는 것을 나타내는 "흥분성(excitatory)"인지 및/또는 입력이, 예를 들어, 작은 수치 값을 갖는 대응하는 가중치에 의해 하나 이상의 출력들(y)에 대해 약한 영향을 갖는다는 것을 나타내는 "억제성(inhibitory)"인지 여부를 나타낼 수 있다. 가중치들(w_i)의 값들은, 이상에서 설명된 바와 같은 임의의 적절한 프로세스를 사용하여 수행될 수 있는 트레이닝 데이터를 사용하여 신경망을 트레이닝시키는 것에 의해 결정될 수 있다. 일 실시예에서, 그리고 비제한적으로, 신경망은 입력들로서 의미 유닛들을 수신하고 본 개시내용에서 설명되는 바와 같은 기계-학습 프로세서들을 사용하여 도출되는 가중치들(w_i)에 따라 이러한 의미 유닛들을 나타내는 벡터들을 출력한다. 본 개시내용에서 사용되는 바와 같은 "컨볼루션 신경망"은, 적어도 하나의 히든 계층이, 풀링 계층들, 완전 연결 계층들, 및 유사한 것과 같은 하나 이상의 추가적인 계층들과 함께, "커널"로서 알려진 입력들의 서브세트와 해당 층에 대한 입력들을 컨볼루션하는 컨볼루션 계층인 신경망이다. CNN은, 비제한적으로, 심층 신경망(deep neural network; DNN) 확장을 포함할 수 있으며, 여기서 DNN은 2개 이상의 히든 계층들을 갖는 신경망으로서 정의된다.

[0036] 일 실시예에서, 계속해서 도 3을 참조하면, CNN 및/또는 다른 모델들은 입력 픽처들에 대한 표준 비디오 인코더

(116) 요건들에 맞게 특징 맵들을 인코딩하거나, 조정하거나 및/또는 재배열하기 위해 사용될 수 있다. 예시적인 실시예에서, 폭(W) 및 높이(H)의 입력 이미지가 CNN 및/또는 다른 모델들에 입력될 수 있다. 컨볼루션 및 풀링으로 알려진 일련의 동작들을 통해, 입력 픽처 내의 이미지는 하나 이상의 특징 맵들로 변환될 수 있다. 출력으로서 CNN의 각각의 계층은 몇몇 특징 맵들을 포함할 수 있다. 총 n개의 컨볼루션(C) 및 풀링(P) 계층들이 있을 수 있다. 각각의 계층은 폭(w) 및 높이(h)로서 표현되는 컨볼루션 및 풀링 커널의 동일한 또는 상이한 치수들을 가질 수 있으며; 예를 들어, 컨볼루션 계층 1은 폭(C1_w) 및 높이(C1_h)를 가질 수 있고, 풀링 계층 1은 폭(P1_w) 및 높이(P1_h)를 가질 수 있으며, 컨볼루션 계층 2는 폭(C2_w) 및 높이(C2_h)를 가질 수 있고, 풀링 계층 2는 폭(P2_w) 및 높이(P2_h)를 가질 수 있는 등이다.

[0037] 일부 실시예들에서, 그리고 계속해서 도 3을 참조하면, 최종 풀링 동작 이후에, 출력 벡터는, 입력 벡터들의 분류를 위해 사용될 수 있는 비제한적으로 심층 신경망과 같은 분류기로 전달될 수 있다. 본 개시내용에서 사용되는 바와 같은 "분류 알고리즘"은, 입력들을 데이터의 카테고리들 또는 빈(bin)들로 소팅하여 데이터의 카테고리들 또는 빈들 및/또는 이와 연관된 라벨들을 출력하는 이하에서 더 상세하게 설명되는 바와 같은 기계 학습 알고리즘 및/또는 프로세스이며; 본 개시내용에서 사용되는 바와 같은 "분류기"는 이하에서 더 상세하게 설명되는 바와 같은 기계-학습 프로세스들에 의해 생성된 수학적 모델, 신경망, 또는 프로그램과 같은 기계-학습 모델이다. 분류기는, 함께 클러스터링되거나, 이하에서 설명되는 바와 같은 거리 메트릭 하에서 가까운 것으로 발견되거나 하는 등의 데이터의 세트를 라벨링하거나 또는 그렇지 않으면 식별하는 적어도 데이터를 출력하도록 구성될 수 있다. 컴퓨팅 디바이스 및/또는 다른 디바이스는, 컴퓨팅 디바이스가 트레이닝 데이터로부터 분류기를 도출하는 프로세스들로서 정의된 분류 알고리즘을 사용하여 분류기를 생성할 수 있다. DNN 분류기와 같은 신경망 분류기들에 더하여, 분류는, 비제한적으로, 선형 분류기들 예컨대 비제한적으로 로지스틱 회귀 및/또는 나이브 베이즈 분류기(naive Bayes classifier)들과 같은 선형 분류기, k-최근접 이웃 분류기들과 같은 최근접 이웃 분류기들, 서포트 벡터 머신들, 최소 제곱 서포트 벡터 머신들, 피셔의 선형 판별(fisher's linear discriminant), 2차 분류기들, 결정 트리들, 부스팅 트리들, 랜덤 포레스트 분류기들, 학습 벡터 양자화, 및/또는 신경망-기반 분류기들을 사용하여 수행될 수 있다.

[0038] 일 실시예에서, 그리고 계속해서 도 3을 참조하면, 출력 뉴런들의 값들의 조합은 관심이 있는 특정 클래스의 존재 또는 부존재를 결정하기 위해 VCM 인코더(100)에 의해 사용될 수 있다. 예를 들어, 2개의 출력들을 갖는 신경망은, 사람이 검출되는 경우 출력 조합 [0, 1] 그리고 자동차가 검출되는 경우 출력 조합 [1, 0] 등으로 트레이닝될 수 있다. 출력으로서 컨볼루션 및 풀링이 값들의 2-차원 매트릭스를 가질 수 있기 때문에, 이러한 값들의 2-차원 매트릭스는 시각적 픽처로서 표현될 수 있다. 이러한 픽처는 또한 "특징 맵"으로도 지칭될 수 있다. 일부 구현예들에서, 그리고 비제한적으로, VCM 인코더(100)는 완전한 CNN을 포함할 수 있으며, 이러한 경우에 특징 추출기(120)의 출력은 최종 풀링된 벡터, 또는 DNN의 중간 또는 마지막 계층들의 출력을 포함할 수 있다. 다른 구현예들에서, VCM 인코더(100)는 CNN의 일 부분만을 포함할 수 있다. 예를 들어, VCM 인코더(100)는 C1, P1, C2, P2, ..., Ck, Pk로 표시될 수 있는 CNN의 처음 k개의 계층들을 포함할 수 있으며, 여기서 k는 1과 n-1 사이의 임의의 정수를 포함할 수 있다. 이러한 경우에, 특징 추출기(120)의 출력은 최종 계층 (k)로부터의 특징 맵들의 세트를 포함할 수 있다. 일부 실시예들에서, 특징 추출기(120)의 출력이 컨볼루션 및/또는 풀링 계층들로부터의 특징 맵들의 세트를 포함하는 경우, 출력은 비디오 인코더(116)를 사용하여 인코딩될 수 있다. 특징 맵 세트를 비디오 인코더(116)로 전송하기 이전에, 특징 추출기(120)는 각각의 특징 맵을 적응시키거나 및/또는 재배열하여 비디오 인코더(116)에 대한 입력으로서 적절하게 만들 수 있다. 특징 맵들이 일반적으로 비디오 인코더(116)에 의해 사용되는 전형적인 픽처 크기보다 훨씬 더 작기 때문에, 특징 맵들의 세트를 재배열하는 하나의 방법은 이들을 비디오 인코딩에 적절한 큰 직사각형 유닛으로 조립하는 것을 포함할 수 있다.

[0039] 계속해서 도 3을 참조하면, 특징 재배열 및/또는 매핑을 위해 사용되는 CNN, DNN, 및/또는 임의의 다른 모델 및/또는 프로세스는 이하에서 더 상세하게 설명되는 바와 같은 임의의 기계-학습 프로세스를 사용하여 트레이닝될 수 있다.

[0040] 이제 도 4를 참조하면, 공간적 재조립을 위한 프로세스(400)의 예시적인 실시예가 예시된다. 일 실시예에서, 본 명세서에서 설명되는 컨볼루션 계층 및/또는 다른 계층 및/또는 요소로부터의 특징 맵들의 세트는 직사각형 유닛으로 배열될 수 있으며; 유닛은 대안적으로 또는 추가적으로 직사각형 형태들을 슬라이스들 및/또는 타일들로 조합한 것, 및/또는 다각형 및/또는 만곡된 주변부를 갖는 임의의 형상을 포함하는 임의의 다른 적절한 형상을 가질 수 있다. 하나의 이러한 형상은 도넛-형 또는 다른 인클로저를 형성하는 다른 이러한 형상을 둘러쌀 수 있다. 개별적인 맵들의 공간적 위치는, 맵들이 컨볼루션 연산의 그들의 순차적 순서에 대응하는 사분면에 위치되는 단순한 순차적 배열에 의해 결정될 수 있다. 유닛 내의 맵들의 공간적 위치는 대안적으로 또는 추가적으로

결과적인 비디오 코딩이 최적화되도록 배열될 수 있다. 이러한 최적화를 위해 적용될 수 있는 동작의 일 예는, 예를 들어 그리고 이하에서 설명되는 바와 같이 비제한적으로, 유사한 텍스처를 갖는 맵들을 서로 옆에 배치하여 비디오 인코더(116)에 의한 인트라 예측의 효율성을 증가시키는 것이다. 텍스처의 측정은 맵 내의 픽셀 값들의 분산으로서 표현될 수 있다. 또한, 단일 컨볼루션 계층으로부터 조립된 특징 유닛들은, 공간적 또는 시간적 방식으로 다른 컨볼루션 계층들로부터의 특징 유닛들과 조합될 수 있다.

[0041] 이제 도 5를 참조하면, 도 4에서 생성된 바와 같은 배열된 출력들로서 정의되는, 유닛들로부터 조립된 픽처 맵들의 세트의 예시적인 실시예가 예시된다. 비제한적인 예로서, 컨볼루션 유닛들은, 비디오 인코더(116)에 의해 인코딩되는 픽처 내부의 이웃 블록들로서 조립될 수 있다. 일부 경우들에서, 유닛 경계들을 비디오 인코더(116)의 코딩 유닛들의 경계들과 정렬하는 것이 유리할 수 있다는 점을 유의해야 한다. 이는, 비제한적으로, 코딩 유닛들과 매칭되는 크기(예를 들어, 64x64 픽셀, 또는 128x128 픽셀)의 컨볼루션 커널들을 사용함으로써, 또는 코딩 유닛 크기와 매칭되도록 컨볼루션 유닛들을 리스케일링함으로써 달성될 수 있다. 리스케일링은, 컨볼루션 유닛의 픽셀들에 적용되는 단순한 선형 또는 바이큐빅(bicubic) 리스케일링 기술을 사용하여 달성될 수 있다.

[0042] 계속해서 도 5를 참조하면, 픽처의 단일 코딩 유닛은, 예를 들어, 도 5의 우측 하단에 도시된 바와 같이, 하나 이상의 컨볼루션 유닛들을 포함할 수 있다. 특정 배열 내의 컨볼루션 유닛의 공간적 위치는, 비제한적으로, 컨볼루션 유닛들의 순차적 순서에 기초하여, 또는 텍스처와 같은 유사한 특성들을 갖는 유닛들을 근접하게 배치하는 최적화 알고리즘들을 사용함으로써 계산될 수 있다. 비디오 코딩의 인트라 예측에서의 개선 이외에, 비디오 인코더(116)가 유닛들에 시간적 예측 및/또는 인트라 예측을 적용할 수도 있기 때문에, 공간적 위치들에 대한 시간적 매핑도 고려될 수 있다.

[0043] 이제 도 6을 참조하면, 모션 추정을 사용하여 계산되고 모션 벡터(motion vector; MV)를 야기하는 인트라 예측의 예시적인 실시예의 묘사가 예시된다. 컨볼루션 맵들의 동질성 및 그에 따른 연속적인 픽처들에서의 유닛들의 유사성이 입력 비디오(104)의 변화에 의존하기 때문에, 비디오 인코더(116)는 최적 압축을 달성하기 위해 동적 해상도 변경을 적용할 수 있다. 예를 들어, 비디오 인코더(116)는, 관련 컨볼루션 맵들의 수가 감소할 때 픽처들의 해상도를 감소시킬 수 있다. 이는, 입력 비디오(104)가 다수의 객체들 및 관심 영역들을 포함하는 장면으로부터 하나의 객체 및/또는 더 적은 관심 영역을 갖는 장면으로 전환하는 경우에 발생할 수 있다. 이러한 경우에 컨볼루션 유닛들의 수는 감소될 수 있으며; 결과적으로, 코딩 픽처의 해상도를 감소시키는 것이 유익할 수 있다. 해상도의 변경은, 비제한적으로 헤더 정보를 사용하는 것을 포함하여, 본 개시내용의 전체를 검토할 때 당업자들에게 일어날 수 있는 임의의 방식으로 디코더에 시그널링될 수 있다. 컨볼루션 맵들에 관한 다른 적절한 정보 및/또는 파라미터들은 보충 향상 정보(supplemental enhancement information; SEI) 스트림 또는 다른 유사한 메타데이터 스트림을 사용하여 인코딩될 수 있다. SEI와 같은 메타데이터 스트림은 기계 모델을 설명하는 관련 정보를 인코딩하기 위해서도 사용될 수 있다.

[0044] 계속해서 도 6을 참조하면, 비디오 인코더(116)의 최적 성능은 대부분의 표준 인코더들에 내재될 수 있는 레이트-왜곡 최적화(rate-distortion optimization; RDO)에 의해 주어진 입력 픽처들에 대해 보장될 수 있지만, 일부 파라미터들은 특정 추출기(120)에 의해 조정되고 업데이트될 수 있다. 이러한 조정의 이유는, 비디오 인코더(116)가 인코딩된 비디오의 시각적 품질에 대해 최적화될 수 있는 반면, 기계의 유틸리티 함수는 특정 경우들에서 해당 척도로부터 벗어날 수 있다는 사실을 포함할 수 있다. 예를 들어, 그리고 비제한적으로, CNN에서 중요한 가중치들의 값들과 연관된 특징 맵들은 더 중요한 것으로서 시그널링될 수 있으며, 이러한 의미에서, 비디오 인코더(116)에 의한 이들에 대한 압축은 조정되어야 할 필요가 있을 수 있다. 이러한 기능의 하나의 비제한적인 예는, 인코더의 양자화 레벨을 주어진 컨볼루션 유닛과 연관된 가중치(들)의 크기에 반비례하도록 업데이트하는 것일 수 있다. 가중치들의 크기는 CNN 내의 모든 가중치들과 비교한 상대적 값으로서 표현될 수 있다.

[0045] 이제 도 7을 참조하면, 모션 벡터 매핑을 위한 프로세스의 예시적인 실시예가 예시된다. 일 실시예에서, 모션 벡터 매핑은 특정 추출기(120)와 비디오 인코더(116) 사이의 상호작용을 포함할 수 있거나 및/또는 이에 의해 드라이브될 수 있으며; 이러한 상호작용은 하이브리드 모드 및/또는 비디오 모드에서 발생할 수 있다. 일 실시예에서, 모션 추정 정보는 특정 추출기(120)로부터 비디오 인코더(116)로 및/또는 비디오 인코더(116)로부터 특정 추출기(120)로 전송될 수 있다. 연속적인 픽처들이 어떤 실제 세계 움직임을 나타내는 시간 종속성을 갖는 입력 비디오(104)의 경우에, 모션은 모션 추정을 사용하는 인트라 예측을 사용하여 추정될 수 있다. 결과적인 모션 벡터(motion vector; MV)들은 연속적인 프레임들 사이의 예측 유닛들의 변위를 나타낼 수 있다.

[0046] 일 예로서, 그리고 계속해서 도 7을 참조하면, 단순한 병진 모션은, 현재 프레임 내의 및/또는 하나의 프레임으로부터 다음 프레임으로의 블록들, 코딩 유닛들, 코딩 트리 유닛들, 컨볼루션 유닛들, 및/또는 픽셀들의 변위를

설명하는 2개의 성분들(MVx, MVy)을 갖는 모션 벡터(MV)를 사용하여 설명될 수 있다. 회전, 주밍(zooming), 및/또는 래핑(warping)과 같은 더 복잡한 모션은 아핀(affine) 모션 벡터를 사용하여 설명될 수 있으며, 여기서 본 개시내용에서 사용되는 바와 같은 "아핀 모션 벡터"는 모션 동안 외관 형상이 변화하지 않으면서 비디오 내의 뷰에 걸쳐 이동하는 객체를 예시하는 픽셀들의 세트와 같은, 비디오 픽처 및/또는 픽처에 표현되는 포인트들 또는 픽셀들의 세트의 균일한 변위를 설명하는 벡터이다. 비디오 인코딩 및/또는 디코딩에 대한 일부 접근방식들은 인터 픽처 코딩에서 모션 보상을 위해 4-파라미터 또는 6-파라미터 아핀 모델들을 사용할 수 있다.

[0047] 예를 들어, 그리고 계속해서 도 7을 참조하면, 6-파라미터 아핀 모션은 다음과 같이 설명될 수 있다:

[0048] $x' = ax + by + c$ $y' = dx + ey + f$

[0049] 그리고 4-파라미터 아핀 모션은 다음과 같이 설명될 수 있다:

[0050] $x' = ax + by + cy' = -bx + ay + f$

[0051] 여기서 (x,y) 및 (x',y')는 각각 현재 픽처 및 참조 픽처에서의 픽셀 위치들이며; a, b, c, d, e, 및 f는 아핀 모션 모델의 파라미터들이다.

[0052] 계속해서 도 7을 참조하면, 그리고 이상에서 설명된 바와 같이, 모션 추정은 컨볼루션 유닛들에 대해 및/또는 더 일반적으로는 특징 맵들에 대해 수행될 수 있다. 비디오 모드의 경우에, 이러한 추정은 비디오 인코더(116)에 의해 수행될 수 있다. 그러나, 일부 경우들에서, 빠르고 단순한 모션 추정은 특징 추출기(120) 자체에 구현될 수 있으며, 따라서 특징 맵들을 특징 인코더(128)로 전송하기 이전에 특징 맵들 사이의 시간적 종속성들의 제거를 가능하게 할 수 있다. 이러한 방식으로 획득된 모션 정보는 입력 비디오(104)를 인코딩하기 위해 비디오 인코더(116)에 의해 재사용될 수 있다. 더 낮은 해상도의 특징 맵들에 대한 모션 추정이 훨씬 더 효율적일 수 있기 때문에, 이는 비디오 인코더(116)의 복잡성을 상당히 감소시킬 수 있다. 다시 말해서, 특징 맵들은 이상에서 설명된 바와 같이 생성될 수 있으며, 모션 벡터들은 특징 맵들로부터 도출될 수 있고, 모션 벡터들은 비디오 인코더(116)로 시그널링될 수 있다. 따라서, 비디오 인코더(116)는, 예를 들어 그리고 비제한적으로 이하에서 설명되는 바와 같이, 비디오 비트스트림을 인코딩하기 위해 시그널링된 모션 벡터들을 사용할 수 있다.

[0053] 일부 경우들에서, 그리고 계속해서 도 7을 참조하면, 모션 벡터 매핑은 비디오 인코더(116)에 의해 계산되고, 특징 추출기(120)로 전송될 수 있다. 이는 특징 모달리티(feature modality)에서 모션 벡터들의 정밀도를 증가시키기 위해 수행될 수 있다. 모션 벡터들은, 컨볼루션 유닛과 픽처 예측 유닛의 해상도 차이에 비례하는 적절한 스케일링 상수를 적용함으로써, 특징 모달리티와 비디오 모달리티 사이와 같이 모달리티들 사이에서 전송될 수 있다. 이제 도 8을 참조하면, 특징들로부터 비디오로 매핑될 수 있는, ROI들에 기초하는 양자화 매핑을 위한 프로세스의 예시적인 실시예가 예시된다. 일 실시예에서, 비디오 인코더(116)는 주어진 비트 예산에 대해 코딩 유닛의 시각적 품질을 최대화하기 위해 RDO에 기초하여 코딩 유닛에 양자화를 적용할 수 있다. 일부 경우들에서, 픽처의 특정 부분들은 다른 부분들보다 지각적으로 더 중요할 수 있으며, 더 높은 품질로 코딩될 수 있고, 반면 픽처의 나머지 부분은 다소 더 낮은 품질로 코딩될 수 있다. 지각적으로 중요한 부분들의 예들은, 사람 얼굴들, 객체들, 낮은 텍스처, 또는 유사한 것을 포함하는 것들이다. 픽처의 부분들의 중요성은 유틸리티 함수에 의해 결정될 수도 있다. 예를 들어, 감시 비디오에서, 얼굴들 및 관심이 있는 작은 객체들에 대한 세부사항들을 보존하는 것이 가장 중요할 수 있다. 특징 추출기(120)에 의해 획득된 정보를 사용하면, 픽처의 이러한 부분들을 지정하는 것이 가능할 수 있다. 이는 경계 박스들로서 표현될 수 있는 관심 영역들을 사용하여 수행될 수 있으며; 경계 박스들은, 비제한적으로, 그 상단 좌측 코너와 같은 픽처 및/또는 특징 맵 내의 위치의 좌표들(x, y), 및 픽셀 값들로서 표현되는 바와 같은, 폭과 높이(w, h)를 포함하는, 임의의 적절한 방식으로 정의될 수 있다. 이러한 공간적 정보는, ROI 내의 모든 코딩 유닛들에 대한 RDO 계산에서 더 낮은 왜곡 및 더 높은 레이트를 할당할 수 있는 비디오 인코더(116)로 전달될 수 있다.

[0054] 도 8을 더 참조하면, 중요성은, 예컨대 비제한적으로 컨볼루션 유닛, 코딩 유닛, 코딩 트리 유닛, 또는 유사한 것으로서 정의되는 것과 같은 픽처 내의 구역(area)의 특성들에 기초하여 계산되거나 및/또는 외부 전문가에 의해 공급될 수 있는 중요성 계수(S_N)에 따라 결정되거나, 저장되거나, 및/또는 시그널링될 수 있다. 본 명세서에서 사용되는 바와 같은 구역의 "특성"은 그 콘텐츠에 기초하여 결정되는 구역의 측정가능 속성이며; 특성은 제1 구역에 대해 수행된 하나 이상의 계산들의 출력을 사용하여 수치적으로 표현될 수 있다. 하나 이상의 계산들은 제1 구역에 의해 표현되는 임의의 신호의 임의의 분석을 포함할 수 있다. 하나의 비제한적인 예는 품질 모델링 애플리케이션들에서 매끄러운 배경을 갖는 구역에 대해 더 높은 S_N 을 할당하고 덜 매끄러운 배경을 갖는 구역에 대해 더 낮은 S_N 을 할당하는 것을 포함할 수 있으며; 비제한적인 예로서, 매끄러움은 에지들의 수의 결정하기 위

한 캐니(Canny) 에지 검출을 사용하여 결정될 수 있고, 여기서 수가 낮을수록 더 큰 정도의 매끄러움을 나타낸다. 자동 매끄러움 검출의 추가적인 예는 구역에 걸친 공간적 변수들의 신호에 대한 고속 푸리에 변환(fast Fourier transform; FFT)들의 사용을 포함할 수 있으며, 여기서 신호는 임의의 2차원 좌표계에 걸쳐, 그리고 적색-녹색-청색 컬러 값들을 나타내는 채널들에 걸쳐 또는 유사한 것에 걸쳐 분석될 수 있고; FFT를 사용하여 계산된 주파수 영역에서의 더 낮은 주파수 성분들의 더 큰 상대적인 우세는 더 큰 정도의 매끄러움을 나타내며, 반면 높은 주파수들의 더 큰 상대적인 우세는 배경 구역에 걸쳐 컬러 및/또는 음영 값들의 더 빈번하고 빠른 전환들을 나타낼 수 있고 이는 더 낮은 매끄러움 점수를 초래할 수 있으며; 의미상 중요한 객체들은 사용자 입력에 의해 식별될 수 있다. 의미론적 중요성은 대안적으로 또는 추가적으로 에지 구성, 및/또는 텍스처 패턴에 따라 검출될 수 있다. 배경은, 비제한적으로, 의미상 중요한 객체를 비제한적으로 포함하는, 얼굴 또는 다른 아이템과 같은 중요한 또는 "전경" 객체를 나타내는 구역의 일 부분을 수신함으로써 및/또는 검출함으로써 식별될 수 있다. 다른 예는 사람 얼굴과 같은 의미상 중요한 객체들을 포함하는 구역들에 대해 더 높은 S_N 을 할당하는 것을 포함할 수 있다.

[0055] 예시적인 실시예에서, 그리고 계속해서 도 8을 참조하면, 본 개시내용에서 설명되는 CNN 또는 다른 요소는 제1 특징 맵 내의 계층(Cn)에서 얼굴을 검출할 수 있다. 대응하는 경계 박스는 비디오 인코더(116)에 의해 인코딩된 픽처에 매핑될 수 있다. 지정된 코딩 유닛은 더 높은 우선순위를 할당 받을 수 있으며 적절한 RDO 업데이트로 코딩될 수 있다. 다른 예들에서, 도출은, 키포인트 추출기들, 에지 검출기들, 돌출 맵 생성기들, 또는 유사한 것과 같은 다른 특징 모델들로부터 수행될 수 있다.

[0056] 계속해서 도 8을 참조하면, VCM 인코더(100)는, 예를 들어 그리고 비제한적으로, 특징 추출기(120)로부터 비디오 인코더(116)에 제공되는 정보를 사용하여, 특징들에 기초하여 분할 결정들을 수행할 수 있다. 일 실시예에서, 비디오 인코더(116)는 분할과 같은 다른 인코딩 파라미터들을 업데이트하기 위해 특징 추출기(120)로부터 수신된 관련 정보를 사용할 수 있다. 예를 들어, 그리고 비제한적으로, 코딩 유닛 트리의 깊이는 비제한적으로 경계 박스와 같은 관심 영역 내의 유닛들에 대해 증가될 수 있다. 다른 예에서, 비디오 인코더(116)는, 예측 유닛들 사이의 블록 경계들에 도입되는 왜곡들을 회피하고 가능한 한 많은 세부사항들을 보존하기 위해 가장 작은 코딩 유닛의 크기를 관련 경계 박스의 크기에 대해 정렬할 수 있다.

[0057] 계속해서 도 8을 참조하면, 특징 추출기(120)로부터 비디오 인코더(116)로 또는 그 반대로 송신되는 정보는 시간적 정보를 포함할 수 있다. 비제한적인 예에서, 특징 추출기(120)는 비제한적으로 장면 변화와 같은 입력 비디오(104)의 중요한 변화들을 검출하고 비디오 인코더(116)에 타임스탬프를 시그널링하기 위해 사용될 수 있으며; 비디오 인코더(116)는 하나 이상의 픽처들의 그룹(group of pictures; GOP)들의 구조 및 길이를 최적으로 결정하기 위해 이러한 정보를 사용할 수 있다. 키-프레임들, 또는 다시 말해서 인트라 코딩된 프레임들은 비제한적인 예에서 장면의 첫 번째 픽처에 대응할 수 있다. 일 실시예에서, 연속적인 픽처들의 변동성은, 인트라(I)와 인터(P, B) 사이에 GOP에서 사용되는 프레임들의 최적 유형뿐만 아니라, 이러한 프레임들의 수 및/또는 시퀀스를 결정할 수 있다. 반면, 비디오 인코더(116)에 의해 획득된 정보는, 예를 들어, 특징 추출기가 모션 추정을 포함하지 않는 경우에, 특징 추출기(120)에 의해 사용될 수 있다. 이러한 경우들에서, 비디오 인코더(116)의 모션 추정은 특징 추출기(120)에 의한 모션 추적 및 활동 검출을 개선하기 위해 사용될 수 있다.

[0058] 도 9는, 압축된 하이브리드 비트스트림으로부터 비디오 및 특징들을 디코딩하기 위한 비디오 디코더(144) 및/또는 특징 디코더(148)를 구현하는 데 적절할 수 있는 예시적인 디코더를 예시하는 시스템 블록도이다. 디코더(900)는, 엔트로피 디코더 프로세서(904), 역양자화 및 역변환 프로세스(908), 디블로킹(deblocking) 필터(912), 프레임 버퍼(916), 모션 보상 프로세서(920) 및/또는 인트라 예측 프로세서(924)를 포함할 수 있다.

[0059] 동작 시에, 그리고 계속해서 도 9를 참조하면, 비트 스트림(928)은 디코더(900)에 의해 수신되고 엔트로피 디코더 프로세서(904)에 입력될 수 있으며, 엔트로피 디코더 프로세서는 비트 스트림의 부분들을 양자화된 계수들로 엔트로피 디코딩할 수 있다. 양자화된 계수들은, 프로세싱 모드에 따라 모션 압축 프로세서(920) 또는 인트라 예측 프로세서(924)의 출력에 추가될 수 있는 잔차 신호를 생성하기 위해 역양자화 및 역변환을 수행할 수 있는 역양자화 및 역변환 프로세서(908)에 제공될 수 있다. 모션 보상 프로세서(920) 및 인트라 예측 프로세서(924)의 출력은 이전에 디코딩된 블록에 기초하는 블록 예측을 포함할 수 있다. 예측과 잔차의 합은 디블로킹 필터(912)에 의해 프로세싱되어 프레임 버퍼(916)에 저장될 수 있다.

[0060] 일 실시예에서, 그리고 계속해서 도 9를 참조하면, 디코더(900)는, 임의의 순서로 그리고 임의의 정도의 반복으로, 이상에서 설명된 바와 같은 임의의 실시예의 이상에서 설명된 바와 같은 임의의 동작들을 구현하도록 구성된 회로부를 포함할 수 있다. 예를 들어, 디코더(900)는 희망되는 또는 명령된 결과가 달성될 때까지 단일 단계

또는 시퀀스를 반복적으로 수행하도록 구성될 수 있으며; 단계 또는 단계들의 시퀀스의 반복은 후속 반복들에 대한 입력들로서 이전 반복들의 출력들을 사용하여 반복적으로 및/또는 회귀적으로 수행될 수 있으며, 이는 반복들의 입력들 및/또는 출력들을 집성하여 집성 결과, 전역 변수들과 같은 하나 이상의 변수들의 감소 또는 감소, 및/또는 더 큰 프로세싱 태스크에 반복적으로 어드레싱되는 더 작은 프로세싱 태스크들의 세트로 분할하는 것을 생성한다. 디코더는, 2개 이상의 병렬 스트림들, 프로세서 코어들, 또는 유사한 것을 사용하여 단계를 2번 이상 동시에 및/또는 실질적으로 동시에 수행하는 것과 같이 본 개시내용에서 설명되는 바와 같은 임의의 단계 또는 단계들의 시퀀스를 병렬로 수행할 수 있으며: 병렬 스트림들 및/또는 프로세서들 사이의 태스크들의 분할은 반복들 사이에서의 태스크들의 분할에 적절한 임의의 프로토콜을 사용하여 수행될 수 있다. 당업자는, 본 개시내용의 전체를 검토할 때, 단계들, 단계들의 시퀀스들, 프로세싱 태스크들, 및/또는 데이터가 반복, 회귀, 및/또는 병렬 프로세싱을 사용하여 세분화되거나, 공유되거나, 또는 그렇지 않으면 처리되는 다양한 방법들을 인식할 것이다.

[0061] 도 10은, 적응적 크로핑(cropping)이 가능한 예시적인 비디오 인코더(116, 1000)를 예시하는 시스템 블록도이다. 예시적인 비디오 인코더(116, 1000)은, 트리-구조형 매크로 블록 분할 기법(예를 들어, 쿼드-트리 플러스 이진 트리)와 같은 프로세싱 기법에 따라 초기에 세그먼트화되거나 또는 분할될 수 있는 입력 비디오(104, 1004)를 수신할 수 있다. 트리-구조형 매크로 블록 분할 기법의 일 예는 픽처 프레임을 큰 블록 요소들, 소위 코딩 트리 유닛(coding tree unit; CTU)들로 분할하는 것을 포함할 수 있다. 일부 구현예들에서, 각각의 CTU는 다수의 서브-블록들, 소위 코딩 유닛(coding unit; CU)들로 한 번 이상 추가로 분할될 수 있다. 이러한 분할의 최종 결과는 예측 유닛(predictive unit; PU)들로 지칭될 수 있는 서브-블록들의 그룹을 포함할 수 있다. 변환 유닛(transform unit; TU)들도 사용될 수 있다.

[0062] 계속해서 도 10을 참조하면, 예시적인 비디오 인코더(116, 1000)는 인트라 예측 프로세서(1008), 모션 벡터 후보 리스트에 전역 모션 벡터 후보를 추가하는 것을 포함하여 모션 벡터 후보 리스트를 구성할 수 있는 인트라 예측 프로세서로도 지칭될 수 있는 모션 추정/보상 프로세서(1012), 변환/양자화 프로세서(1016), 역양자화/역변환 프로세서(1020), 인-루프(in-loop) 필터(1024), 디코딩된 픽처 버퍼(1028), 및/또는 엔트로피 코딩 프로세서(1032)를 포함할 수 있다. 비트 스트림 파라미터들은 출력 비트 스트림(1036)에 포함시키기 위해 엔트로피 코딩 프로세서(1032)에 입력될 수 있다.

[0063] 동작 시에, 그리고 계속해서 도 10을 참조하면, 입력 비디오(104, 1004)의 프레임의 각각의 블록에 대해, 인트라 픽처 예측을 통해 블록을 프로세싱할지 또는 모션 추정/보상을 사용하여 블록을 프로세싱할지 여부가 결정될 수 있다. 블록은 인트라 예측 프로세서(1008) 및/또는 모션 추정/보상 프로세서(1012)에 제공될 수 있다. 블록이 인트라 예측을 통해 프로세싱되는 경우, 인트라 예측 프로세서(1008)는 예측자를 출력하기 위한 프로세싱을 수행할 수 있다. 블록이 모션 추정/보상을 통해 프로세싱되는 경우, 모션 추정/보상 프로세서(1012)는, 적용가능한 경우, 모션 벡터 후보 리스트에 전역 모션 벡터 후보를 추가하는 것을 포함하여 모션 벡터 후보 리스트를 구성하는 것을 포함하는 프로세싱을 수행할 수 있다.

[0064] 도 10을 더 참조하면, 잔차는 입력 비디오(104)로부터 예측자를 빼는 것에 의해 형성될 수 있다. 잔차는 변환/양자화 프로세서(1016)에 의해 수신될 수 있으며, 변환/양자화 프로세서는 양자화될 수 있는 계수들을 생성하기 위해 변환 프로세싱(예를 들어, 이산 코사인 변환(discrete cosine transform; DCT))을 수행할 수 있다. 양자화된 계수들 및 임의의 연관된 시그널링 정보는 엔트로피 인코딩 및 출력 비트 스트림(1036) 내의 포함을 위해 엔트로피 코딩 프로세서(1032)에 제공될 수 있다. 엔트로피 인코딩 프로세서(1032)는 현재 블록을 인코딩하는 것과 관련된 시그널링 정보의 인코딩을 지원할 수 있다. 또한, 양자화 계수들은 역양자화/역변환 프로세서(1020)에 제공될 수 있고, 역양자화/역변환 프로세서는 예측자와 조합될 수 있으며 인 루프 필터(1024)에 의해 프로세싱될 수 있는 픽셀들을 재현할 수 있고, 역양자화/역변환 프로세서의 출력은, 모션 벡터 후보 리스트에 전역 모션 벡터 후보를 추가하는 것을 포함하여 모션 벡터 후보 리스트를 구성할 수 있는 모션 추정/보상 프로세서(1012)에 의한 사용을 위해 디코딩된 픽처 버퍼(1028)에 저장될 수 있다.

[0065] 계속해서 도 10을 참조하면, 몇몇 변형들이 이상에서 상세하게 설명되었지만, 다른 수정들 또는 추가들이 가능하다. 예를 들어, 일부 구현예들에서, 현재 블록들은 임의의 대칭적 블록들(8x8, 16x16, 32x32, 64x64, 128 x 128, 및 유사한 것)뿐만 아니라 임의의 비대칭적 블록(8x4, 16x8, 및 유사한 것)을 포함할 수 있다.

[0066] 일부 구현예들에서, 그리고 계속해서 도 10을 참조하면, 쿼드트리 플러스 이진 결정 트리(quadtrees plus binary decision tree; QTBT)가 구현될 수 있다. QTBT에서, 코딩 트리 유닛 레벨에서, QTBT의 분할 파라미터들은 임의의 오버헤드를 송신하지 않으면서 로컬 특성에 맞춰 적응되도록 동적으로 도출될 수 있다. 그 후에, 코딩 유닛

레벨에서, 조인트-분류기 결정 트리 구조는 불필요한 반복들을 제거하고 잘못된 예측의 위험을 제어할 수 있다. 일부 구현예들에서, LTR 프레임 블록 업데이트 모드는, QTBT의 모든 리프(leaf) 노드에서 이용가능한 추가적인 옵션으로서 이용가능할 수 있다.

[0067] 일부 구현예들에서, 그리고 계속해서 도 10을 참조하면, 추가적인 선택스(syntax) 요소들은 비트스트림의 상이한 계층 레벨들에서 시그널링될 수 있다. 예를 들어, 플래그는, 시퀀스 파라미터 세트(Sequence Parameter Set; SPS)에 코딩된 인에이블 플래그를 포함시킴으로써 전체 시퀀스에 대해 인에이블될 수 있다. 또한, CTU 플래그는 코딩 트리 유닛(coding tree unit; CTU) 레벨로 코딩될 수 있다.

[0068] 일부 실시예들은, 하나 이상의 컴퓨팅 시스템들의 하나 이상의 데이터 프로세서들에 의해 실행될 때, 적어도 하나의 프로세서가 본 명세서의 동작들을 수행하게 하는 명령어들을 저장하는 비-일시적 컴퓨터 프로그램 제품들(즉, 물리적으로 구현된 컴퓨터 프로그램 제품들)을 포함할 수 있다.

[0069] 계속해서 도 10을 참조하면, 인코더(1000)는, 임의의 순서로 그리고 임의의 정도의 반복으로, 임의의 실시예의 이상에서 설명된 바와 같은 임의의 동작들을 구현하도록 구성된 회로부를 포함할 수 있다. 예를 들어, 인코더(1000)는 희망되는 또는 명령된 결과가 달성될 때까지 단일 단계 또는 시퀀스를 반복적으로 수행하도록 구성될 수 있으며; 단계 또는 단계들의 시퀀스의 반복은 후속 반복들에 대한 입력들로서 이전 반복들의 출력들을 사용하여 반복적으로 및/또는 회귀적으로 수행될 수 있으며, 이는 반복들의 입력들 및/또는 출력들을 집성하여 집성 결과, 전역 변수들과 같은 하나 이상의 변수들의 감소 또는 감분, 및/또는 더 큰 프로세싱 태스크에 반복적으로 어드레싱되는 더 작은 프로세싱 태스크들의 세트로 분할하는 것을 생성한다. 인코더(1000)는, 2개 이상의 병렬 스레드들, 프로세서 코어들, 또는 유사한 것을 사용하여 단계를 2번 이상 동시에 및/또는 실질적으로 동시에 수행하는 것과 같이 본 개시내용에서 설명되는 바와 같은 임의의 단계 또는 단계들의 시퀀스를 병렬로 수행할 수 있으며; 병렬 스레드들 및/또는 프로세스들 사이의 태스크들의 분할은 반복들 사이에서의 태스크들의 분할에 적절한 임의의 프로토콜을 사용하여 수행될 수 있다. 당업자는, 본 개시내용의 전체를 검토할 때, 단계들, 단계들의 시퀀스들, 프로세싱 태스크들, 및/또는 데이터가 반복, 회귀, 및/또는 병렬 프로세싱을 사용하여 세분화되거나, 공유되거나, 또는 그렇지 않으면 처리되는 다양한 방법들을 인식할 것이다.

[0070] 계속해서 도 10을 참조하면, 비-일시적 컴퓨터 프로그램 제품들(즉, 물리적으로 구현된 컴퓨터 프로그램 제품들)은, 하나 이상의 컴퓨팅 시스템들의 하나 이상의 데이터 프로세서들에 의해 실행될 때, 적어도 하나의 프로세서가, 비제한적으로 이상에서 설명된 동작들 및/또는 디코더(900) 및/또는 인코더(1000)가 수행하도록 구성될 수 있는 임의의 동작들을 포함하여, 본 개시내용에서 설명되는 동작들 및/또는 이의 단계들을 수행하게 하는 명령어들을 저장할 수 있다. 유사하게, 컴퓨터 시스템들은 또한, 하나 이상의 데이터 프로세서 및 하나 이상의 데이터 프로세서들에 결합된 메모리를 포함할 수 있는 것으로 설명된다. 메모리는, 적어도 하나의 프로세서가 본 명세서에서 설명되는 동작들 중 하나 이상을 수행하게 하는 명령어들을 일시적으로 또는 영구적으로 저장할 수 있다. 또한, 방법들은, 단일 컴퓨팅 시스템 내의 또는 2개 이상의 컴퓨팅 시스템들 사이에 분산된 하나 이상의 데이터 프로세서들에 의해 구현될 수 있다. 이러한 컴퓨팅 시스템들은, 네트워크(예를 들어, 인터넷, 무선 광역 네트워크, 근거리 네트워크, 광역 네트워크, 유선 네트워크, 또는 유사한 것)를 통한 연결을 포함하는 하나 이상의 연결들을 통해, 다수의 컴퓨팅 시스템들 중 하나 이상 사이의 직접 연결을 통해, 또는 유사한 것을 통해 연결될 수 있으며, 이를 통해 데이터 및/또는 명령들 또는 다른 명령어들 또는 유사한 것을 교환할 수 있다.

[0071] 이제 도 11을 참조하면, 본 개시내용에서 설명되는 바와 같은 하나 이상의 기계-학습 프로세스들을 수행할 수 있는 기계-학습 모듈(1100)의 예시적인 실시예가 예시된다. 기계-학습 모듈은 기계 학습 프로세스들을 사용하여 본 개시내용에서 설명되는 바와 같은 결정들, 분류, 및/또는 분석 단계들, 방법들, 프로세스들, 또는 유사한 것을 수행할 수 있다. 본 개시내용에서 사용되는 바와 같은 "기계 학습 프로세스"는, 입력들(1112)로서 제공되는 데이터가 주어지면 출력들(1108)을 생성하기 위해 컴퓨팅 디바이스/모듈에 의해 수행될 알고리즘을 생성하기 위해 트레이닝 데이터(1104)를 자동으로 사용하는 프로세스이다. 이는, 실행될 명령들이 일반적으로 사용자에게 의해 미리 결정되고 프로그래밍 언어로 작성되는 비-기계 학습 소프트웨어와는 대조된다.

[0072] 계속해서 도 11을 참조하면, 본 명세서에서 사용되는 바와 같은 "트레이닝 데이터"는, 기계-학습 프로세스가 데이터 요소들의 2개 이상의 카테고리들 사이의 관계를 모델링하기 위해 사용할 수 있는 상관관계들을 포함하는 데이터이다. 예를 들어, 그리고 비제한적으로, 트레이닝 데이터(1104)는 복수의 데이터 엔트리들을 포함할 수 있으며, 각각의 엔트리는 함께 기록되거나, 수신되거나, 및/또는 생성된 데이터 요소들의 세트를 나타내고; 데이터 요소들은 주어진 데이터 엔트리 내의 공유된 존재에 의해, 주어진 데이터 엔트리 내의 근접성에 의해, 또

는 유사한 것에 의해 상관될 수 있다. 트레이닝 데이터(1104) 내의 다수의 데이터 엔트리들은 데이터 요소들의 카테고리들 사이의 상관관계에서 하나 이상의 추세들을 나타낼 수 있으며; 예를 들어, 그리고 비제한적으로, 데이터 요소의 제1 카테고리에 속하는 제1 데이터 요소의 더 높은 값은 데이터 요소의 제2 카테고리에 속하는 제2 데이터 요소의 더 높은 값과 상관되는 경향이 있을 수 있고, 이는 2개의 카테고리들에 속하는 값들을 링크하는 가능한 비례적 또는 다른 수학적 관계를 나타낸다. 데이터 요소들의 다수의 카테고리들은 다양한 상관관계들에 따라 트레이닝 데이터(1104) 내에서 관련될 수 있으며; 상관관계들은 데이터 요소들의 카테고리들 사이의 원인 및/또는 예측 링크들을 나타낼 수 있고, 이는 이하에서 더 상세하게 설명되는 바와 같이 기계-학습 프로세스들에 의해 수학적 관계들과 같은 관계들로서 모델링될 수 있다. 트레이닝 데이터(1104)는, 예를 들어, 데이터 요소들을 데이터 요소들의 카테고리들에 대응하는 하나 이상의 설명자(descriptor)들과 연관시킴으로써, 데이터 요소들의 카테고리들에 의해 포맷팅되거나 및/또는 조직될 수 있다. 비제한적인 예로서, 트레이닝 데이터(1104)는 사람들 또는 프로세스들에 의해 표준화된 형식들로 입력된 데이터를 포함할 수 있어서, 형식 내의 주어진 필드 내의 주어진 데이터 요소의 엔트리는 카테고리들의 하나 이상의 설명자들에 매핑될 수 있다. 트레이닝 데이터(1104) 내의 요소들은 태그들, 토큰들, 또는 다른 데이터 요소들에 의해 카테고리들의 설명자들에 링크될 수 있으며; 예를 들어, 그리고 비제한적으로, 트레이닝 데이터(1104)는, 길이-고정 포맷들, 카테고리들에 대해 데이터의 위치들을 링크하는 포맷들 예컨대 콤마-분리 값(comma-separated value; CSV) 포맷들 및/또는 자체-설명 포맷들 예컨대 확장 마크업 언어(extensible markup language; XML), 자바스크립트 객체 표기(JavaScript Object Notation; JSON), 또는 유사한 것으로 제공될 수 있으며, 이는 프로세스들 또는 디바이스들의 데이터의 카테고리들을 검출하는 것을 가능하게 한다.

[0073] 대안적으로 또는 추가적으로, 그리고 계속해서 도 11을 참조하면, 트레이닝 데이터(1104)는 카테고리화되지 않은 하나 이상의 요소들을 포함할 수 있으며; 즉, 트레이닝 데이터(1104)는 데이터의 일부 요소들에 대해 포맷팅되지 않을 수 있거나 또는 이에 대한 설명자들을 포함하지 않을 수 있다. 기계-학습 알고리즘들 및/또는 다른 프로세스들은, 예를 들어, 자연 언어 프로세싱 알고리즘들, 토큰화, 원시 데이터 내의 상관된 값들의 검출 및 유사한 것을 사용하여 하나 이상의 카테고리화에 따라 트레이닝 데이터(1104)를 소팅할 수 있으며; 카테고리들은 상관관계 및/또는 다른 프로세싱 알고리즘들을 사용하여 생성될 수 있다. 비제한적인 예로서, 텍스트 자료에서, 다른 명사들에 의해 수식되는 명사들과 같은 "n"개의 복합어들을 구성하는 문구들은 특정한 순서로 이러한 단어들을 포함하는 n-gram들의 통계적으로 중요한 보급률(statistically significant prevalence)에 따라 식별될 수 있으며; 이러한 n-gram은 단일 단어들과 유사하게 추적될 "단어"와 같은 언어의 요소로서 카테고리화되며, 이는 통계적 분석의 결과로서 새로운 카테고리를 생성한다. 유사하게, 일부 텍스트 데이터를 포함하는 데이터 엔트리에서, 사람의 이름은 리스트, 사진, 또는 다른 용어들의 개요를 참조하여 식별될 수 있으며, 이는 기계-학습 알고리즘들에 의한 애드-혹(ad-hoc) 카테고리화, 및/또는 데이터 엔트리 내의 데이터를 설명자들과 자동으로 연관시키는 것 또는 주어진 포맷으로 자동으로 연관시키는 것을 허용한다. 데이터 엔트리들을 자동으로 분류하기 위한 능력은, 이하에서 더 상세하게 설명되는 바와 같이 동일한 트레이닝 데이터(1104)를 2개 이상의 개별 기계-학습 알고리즘들에 적용가능하게 만드는 것을 가능하게 할 수 있다. 기계-학습 모듈(1100)에 의해 사용되는 트레이닝 데이터(1104)는 본 개시내용에서 설명되는 바와 같은 임의의 입력 데이터를 본 개시내용에서 설명되는 바와 같은 임의의 출력 데이터에 상관시킬 수 있다.

[0074] 도 11을 더 참조하면, 트레이닝 데이터는 이하에서 더 상세하게 설명되는 바와 같은 하나 이상의 지도 및/또는 비지도 기계-학습 프로세스들 및/또는 모델들을 사용하여 필터링되거나, 소팅되거나 및/또는 선택될 수 있으며; 이러한 모델들은 비제한적으로 트레이닝 데이터 분류기(1116)를 포함할 수 있다. 트레이닝 데이터 분류기(1116)는, 입력들을 데이터의 카테고리들 또는 빈들로 소팅하여 데이터의 카테고리들 또는 빈들 및/또는 이와 연관된 라벨들을 출력하는, 이하에서 더 상세하게 설명되는 바와 같은, "분류 알고리즘"으로 알려진 기계 학습 알고리즘에 의해 생성된 수학적 모델, 신경망, 또는 프로그램과 같은, 본 개시내용에서 사용되는 바와 같이 이하에서 정의되는 바와 같은 기계-학습 모델인 "분류기"를 포함할 수 있다. 분류기는, 함께 클러스터링되거나, 이하에서 설명되는 바와 같은 거리 메트릭 하에서 가까운 것으로 발견되거나 하는 등의 데이터의 세트를 라벨링하거나 또는 그렇지 않으면 식별하는 적어도 데이터를 출력하도록 구성될 수 있다. 기계-학습 모듈(1100)은, 이에 의해 컴퓨팅 디바이스 및/또는 그 위에서 동작하는 임의의 모듈 및/또는 구성요소가 트레이닝 데이터(1104)로부터 분류기를 도출하는 프로세스로서 정의된 분류 알고리즘을 사용하여 분류기를 생성할 수 있다.

[0075] 분류는, 비제한적으로, 선형 분류기들 예컨대 비제한적으로 로지스틱 회귀 및/또는 나이브 베이즈 분류기들과 같은 선형 분류기, k-최근접 이웃 분류기들과 같은 최근접 이웃 분류기들, 서포트 벡터 머신들, 최소 제곱 서포트 벡터 머신들, 피셔의 선형 판별, 2차 분류기들, 결정 트리들, 부스팅 트리들, 랜덤 포레스트 분류기들, 학습

벡터 양자화, 및/또는 신경망-기반 분류기들을 사용하여 수행될 수 있다.

[0076] 계속해서 도 11을 참조하면, 기계-학습 모듈(1100)은, 대안적으로 "레이지 로딩(lazy loading)" 또는 "필요시-호출(call-when-needed)" 프로세스 및/또는 프로토콜로 지칭될 수 있고, 이에 의해 기계 학습이, 온 디맨드(on demand)로 출력을 생성하기 위해 사용될 알고리즘을 도출하기 위해 입력 및 트레이닝 세트를 조합함으로써, 출력으로 변환될 입력의 수신 시에 수행되는 프로세스일 수 있는 레이저-학습 프로세스(1120) 및/또는 프로토콜을 수행하도록 구성될 수 있다. 예를 들어, 시뮬레이션들의 초기 세트는 출력 및/또는 관계에서의 초기 휴리스틱(heuristic) 및/또는 "제1 추측(guess)"을 커버하기 위해 수행될 수 있다. 비제한적인 예로서, 초기 휴리스틱은 입력들과 트레이닝 데이터(1104)의 요소들 사이의 연관들의 랭킹을 포함할 수 있다. 휴리스틱은 어떤 수의 가장 높은 랭킹의 연관들 및/또는 트레이닝 데이터(1104) 요소들을 선택하는 것을 포함할 수 있다. 레이저 학습은, 비제한적으로, K-최근접 이웃 알고리즘, 레이저 나이브 베이즈 알고리즘, 또는 유사한 것을 포함하는 임의의 적절한 레이저 학습 알고리즘을 구현할 수 있으며; 당업자들은, 본 개시내용 전체를 검토할 때, 비제한적으로, 이하에서 더 상세하게 설명되는 바와 같이 기계-학습 알고리즘들의 레이저 학습 애플리케이션을 포함하는, 본 개시내용에서 설명되는 바와 같은 출력들을 생성하기 위해 적용될 수 있는 다양한 레이저-학습 알고리즘들을 인식할 것이다.

[0077] 대안적으로 또는 추가적으로, 그리고 계속해서 도 11을 참조하면, 본 개시내용에서 설명되는 바와 같은 기계-학습 프로세스들은 기계-학습 모델들(1124)을 생성하기 위해 사용될 수 있다. 본 개시내용에서 사용되는 바와 같은 "기계-학습 모델"은, 비제한적으로 이상에서 설명된 임의의 프로세스를 포함하는 임의의 기계-학습 프로세스를 사용하여 생성되고 메모리에 저장되는 바와 같은, 입력들과 출력들 사이의 관계의 수학적 및/또는 알고리즘적 표현이며; 일단 기계-학습 모델(1124)이 생성되면 입력이 기계-학습 모델에 제출되며, 기계-학습 모델은 도출된 관계에 기초하여 출력을 생성한다. 예를 들어, 그리고 비제한적으로, 선형 회귀 알고리즘을 사용하여 생성된 선형 회귀 모델은 출력 데이터를 계산하기 위해 기계-학습 프로세스들 동안 도출된 계수들을 사용하여 입력 데이터의 선형 조합을 계산할 수 있다. 추가적인 비제한적인 예로서, 기계-학습 모델(1124)은, 노드들의 입력 계층, 하나 이상의 중간 계층들, 및 노드들의 출력 계층을 포함하는 컨볼루션 신경망과 같은 인공 신경망을 생성함으로써 생성될 수 있다. 노드들 사이의 연결들은 네트워크를 "트레이닝시키는" 프로세스를 통해 생성될 수 있으며, 이러한 프로세스에서 트레이닝 데이터(1104) 세트로부터의 요소들이 입력 노드들에 적용되고, 그런 다음 출력 노드들에서 희망되는 값을 생성하기 위해서 신경망의 인접한 계층들의 노드들 사이의 연결들 및 가중치들을 조정하기 위해 적절한 트레이닝 알고리즘(예컨대 레벤버그-마쿼트, 공역 기울기, 모의 담금질, 또는 다른 알고리즘들)이 사용된다. 이러한 프로세스는 때때로 심층 학습으로 지칭된다.

[0078] 계속해서 도 11을 참조하면, 기계-학습 알고리즘들은 적어도 지도 기계-학습 프로세스(1128)를 포함할 수 있다. 본 명세서에서 정의되는 바와 같은 적어도 지도 기계-학습 프로세스(1128)는, 다수의 입력들을 다수의 출력들에 관련시키는 트레이닝 세트를 수신하고 입력들을 출력들에 관련시키는 하나 이상의 수학적 관계들을 찾는 것을 추구하는 알고리즘들을 포함하며, 여기서 하나 이상의 수학적 관계들 각각은 일부 스코어링 함수를 사용하는 알고리즘에 특정한 일부 기준에 따라 최적이다. 예를 들어, 지도 학습 알고리즘은, 입력들로서 본 개시내용에서 설명된 바와 같은 입력들, 출력들로서 본 개시내용에서 설명된 바와 같은 출력들, 및 입력들과 출력들 사이에 검출되는 희망되는 관계의 형태를 나타내는 스코어링 함수를 포함할 수 있으며; 스코어링 함수는, 예를 들어, 주어진 입력이 주어진 출력과 연관되지 않을 가능성을 최소화하기 위해 주어진 입력 및/또는 요소들 입력들의 조합이 주어진 출력과 연관될 가능성을 최대화하는 것을 추구할 수 있다. 스코어링 함수는 입력들을 출력들에 관련시키는 알고리즘의 "예상된 손실"을 나타내는 위험 함수로서 표현될 수 있으며, 여기서 손실은, 트레이닝 데이터(1104)에서 제공된 주어진 입력-출력 쌍과 비교할 때, 해당 관계에 의해 생성된 예측이 부정확할 정도를 나타내는 오류 함수로서 계산된다. 당업자들은, 본 개시내용 전체를 검토할 때, 입력들과 출력들 사이의 관계를 결정하기 위해 사용될 수 있는 적어도 지도 기계-학습 프로세스(1128)의 다양하고 가능한 변형들을 인식할 것이다. 지도 기계-학습 프로세스들은 이상에서 정의된 바와 같은 분류 알고리즘들을 포함할 수 있다.

[0079] 도 11을 더 참조하면, 기계-학습 프로세스들은 적어도 비지도 기계-학습 프로세스(1132)를 포함할 수 있다. 본 명세서에서 사용되는 바와 같은 비지도 기계-학습 프로세스는, 라벨들과 관계없이 데이터세트들에서 추론들을 도출하는 프로세스이며; 결과적으로, 비지도 기계-학습 프로세스는 데이터에 제공된 임의의 구조, 관계, 및/또는 상관관계를 자유롭게 발견할 수 있다. 비지도 프로세스들은 응답 변수를 필요로 하지 않을 수 있으며; 비지도 프로세스들은 2개 이상의 변수들 사이의 상관관계의 정도 또는 유사한 것을 결정하기 위해 변수들 사이의 흥미로운 패턴들 및/또는 추론들을 찾기 위해 사용될 수 있다.

[0080] 계속해서 도 11을 참조하면, 기계-학습 모듈(1100)은, 선형 회귀 모델들의 개발을 위한 기술들을 사용하여 기계

-학습 모델(1124)을 생성하도록 설계되고 구성될 수 있다. 선형 회귀 모델들은, 이러한 차이를 측정하기 위한 적절한 노름(norm)(예를 들어, 벡터-공간 거리 노름)에 따라 예측된 결과들과 실제 결과들 사이의 차이의 제곱을 최소화하는 것을 목적으로 하는 일반적인 최소 제곱 회귀를 포함할 수 있으며; 결과적인 선형 방정식의 계수들은 최소화를 개선하기 위해 수정될 수 있다. 선형 회귀 모델들은 리지(ridge) 회귀 방법들을 포함할 수 있으며, 여기서 최소화될 함수는 큰 계수들에 페널티를 가하기 위해 스칼라 양을 각각의 계수의 제곱에 곱하는 항을 최소-제곱 함수에 더한 것을 포함한다. 선형 회귀 모델들은 최소 절대 수축 및 선택 연산자(least absolute shrinkage and selection operator; LASSO) 모델들을 포함할 수 있으며, 여기서 리지 회귀는, 최소-제곱 항에 1의 인자를 곱하고 샘플들의 수의 2배로 나눈 값과 조합된다. 선형 회귀 모델들은 멀티-태스크 라쏘 모델(multi-task lasso model)을 포함할 수 있으며, 여기서 라쏘 모델의 최소-제곱 항에 적용되는 노름은 모든 항들의 제곱들의 합의 제곱근에 달하는 프로베니우스 노름(Frobenius norm)이다. 선형 회귀 모델들은, 탄성망 모델, 멀티-태스크 탄성망 모델, 최소 각도 회귀 모델, LARS 라쏘 모델, 직교 매칭 추적 모델, 베이지안 회귀 모델, 로지스틱 회귀 모델, 확률적 경사 하강 모델, 퍼셉트론(perceptron) 모델, 수동적 공격 알고리즘, 견고성 회귀 모델, 후버(Huber) 회귀 모델, 또는 본 개시내용 전체를 검토할 때 당업자들에게 일어날 수 있는 임의의 다른 적절한 모델을 포함할 수 있다. 선형 회귀 모델들은 일 실시예에서 다항식 회귀 모델들로 일반화될 수 있고, 이에 의해 최상 예측된 출력/실제 출력 맞춤을 제공하는 다항식 방정식(예를 들어, 2차, 3차, 또는 더 고차 방정식)이 추구되며; 이상에서 설명된 것들과 유사한 방법들은, 본 개시내용 전체를 검토할 때 당업자들에게 명백할 바와 같이 오류 함수들을 최소화하기 위해 적용될 수 있다.

[0081] 계속해서 도 11을 참조하면, 기계-학습 알고리즘들은, 비제한적으로, 선형 판별 분석을 포함할 수 있다. 기계-학습 알고리즘들은 2차 판별 분석을 포함할 수 있다. 기계-학습 알고리즘들은 커널 리지 회귀를 포함할 수 있다. 기계-학습 알고리즘들은, 비제한적으로, 지원 벡터 분류-기반 회귀 프로세스들을 포함하는 지원 벡터 기계들을 포함할 수 있다. 기계-학습 알고리즘들은, 확률적 경사 하강에 기초하는 분류 및 회귀 알고리즘들을 포함하는, 확률적 경사 하강 알고리즘들을 포함할 수 있다. 기계-학습 알고리즘들은 최근접 이웃 알고리즘들을 포함할 수 있다. 기계-학습 알고리즘들은 변형 정규화와 같은 다양한 형태들의 잠재적 공간 정규화를 포함할 수 있다. 기계-학습 알고리즘들은 가우시안 프로세스 회귀와 같은 가우시안 프로세스들을 포함할 수 있다. 기계-학습 알고리즘들은, 부분 최소 제곱들 및/또는 표준 상관 분석들을 포함하는, 교차-분해 알고리즘들을 포함할 수 있다. 기계-학습 알고리즘들은 나이브 베이즈 방법들을 포함할 수 있다. 기계-학습 알고리즘들은, 결정 트리 분류 및 회귀 알고리즘들과 같은 결정 트리들에 기초하는 알고리즘들을 포함할 수 있다. 기계-학습 알고리즘들은, 배깅 메타-추정기(bagging meta-estimator), 랜덤화된 트리들의 포레스트, 에이다부스트(AdaBoost), 경사 트리 부스팅, 및/또는 다수결 분류기(voting classifier) 방법들과 같은 앙상블 방법들을 포함할 수 있다. 기계-학습 알고리즘들은, 컨볼루션 신경망 프로세스들을 포함하는, 신경망 알고리즘들을 포함할 수 있다.

[0082] 본 명세서에서 설명되는 측면들 및 실시예들 중 임의의 하나 이상은, 컴퓨터 기술분야의 당업자들에게 명백할 바와 같이, 본 명세서의 교시들에 따라 프로그래밍된 하나 이상의 기계들(예를 들어, 전자 문서에 대한 사용자 컴퓨팅 디바이스, 하나 이상의 서버 디바이스들, 예컨대 문서 서버, 등으로서 사용되는 하나 이상의 컴퓨팅 디바이스들)을 사용하여 편리하게 구현될 수 있다는 것을 유의해야 한다. 적절한 소프트웨어 코딩은, 소프트웨어 기술분야의 당업자들에게 명백할 바와 같이, 본 개시내용의 교시들에 기초하여 숙련된 프로그래머들에 의해 쉽게 준비될 수 있다. 소프트웨어 및/또는 소프트웨어 모듈들을 이용하는 이상에서 논의된 측면들 및 구현예들은 또한, 소프트웨어 및/또는 소프트웨어 모듈의 기계 실행가능 명령어들의 구현을 보조하기 위한 적절한 하드웨어를 포함할 수 있다.

[0083] 이러한 소프트웨어는, 기계-판독가능 저장 매체를 이용하는 컴퓨터 프로그램 제품일 수 있다. 기계-판독가능 저장 매체는, 기계(예를 들어, 컴퓨팅 디바이스)에 의한 실행을 위한 명령어들의 시퀀스를 저장하거나 및/또는 인코딩할 수 있으며 기계가 본 명세서에서 설명된 방법론들 및/또는 실시예들 중 임의의 하나를 수행하게 하는 임의의 매체일 수 있다. 기계-판독가능 저장 매체의 예들은, 비제한적으로, 자기 디스크, 광학 디스크(예를 들어, CD, CD-R, DVD, DVD-R, 등), 자기-광학 디스크, 판독-전용 메모리("ROM") 디바이스, 랜덤-액세스 메모리("RAM"), 자기 카드, 광학 카드, 고체-상태 메모리 디바이스, EPROM, EEPROM, 및 이들의 임의의 조합들을 포함한다. 본 명세서에서 사용되는 바와 같은 기계-판독가능 매체는, 단일 매체뿐만 아니라, 예를 들어, 컴퓨터 메모리와 조합된 콤팩트 디스크들 또는 하나 이상의 하드 디스크 드라이브들의 모음과 같은 물리적으로 별도의 매체의 모음을 포함하도록 의도된다. 본 명세서에서 사용되는 바와 같은 기계-판독가능 저장 매체는 일시적 형태의 신호 송신을 포함하지 않는다.

[0084] 이러한 소프트웨어는 또한 반송파와 같은 데이터 캐리어 상에서 데이터 신호로서 운반되는 정보(예를 들어, 데

이터)를 포함할 수 있다. 예를 들어, 기계-판독가능 정보는, 신호가, 기계가 본 명세서에서 설명된 방법론들 및/또는 실시예들 중 임의의 것을 수행하게 하는 기계(예를 들어, 컴퓨팅 디바이스)에 의한 실행을 위한 명령어의 시퀀스, 또는 이의 부분, 및 임의의 관련 정보를 인코딩하는, 데이터 캐리어에 구현된 데이터-운반 신호로서 포함될 수 있다.

[0085] 컴퓨팅 디바이스의 예들은, 비제한적으로, 전자 책 리딩 디바이스, 컴퓨터 워크스테이션, 단말 컴퓨터, 서버 컴퓨터, 핸드헬드 디바이스(예를 들어, 태블릿 컴퓨터, 스마트폰, 등), 웹 전기기기, 네트워크 라우터, 네트워크 스위치, 네트워크 브리지, 해당 기계가 취할 액션을 지정하는 명령어들의 시퀀스를 실행할 수 있는 임의의 기계, 및 이들의 임의의 조합을 포함한다. 일 예에서, 컴퓨팅 디바이스는 키오스크를 포함하거나 및/또는 이에 포함될 수 있다.

[0086] 도 12는, 제어 시스템이 본 개시내용의 측면들 및/또는 방법론들 중 임의의 하나 이상을 수행하게 하기 위한 명령어들의 세트가 실행될 수 있는 예시적인 컴퓨터 시스템(1200) 형태의 컴퓨팅 디바이스의 일 실시예의 도시적 표현을 도시한다. 다수의 컴퓨팅 디바이스들이, 디바이스들 중 하나 이상이 본 개시내용의 측면들 및/또는 방법론들 중 임의의 하나 이상을 수행하게 하기 위한 명령어들의 특별히 구성된 세트를 구현하기 위해 사용될 수 있다는 것이 또한 고려된다. 컴퓨터 시스템(1200)은, 버스(1212)를 통해 서로 통신하고 다른 구성요소들과 통신하는 프로세서(1204) 및 메모리(1208)를 포함한다. 버스(1212)는, 비제한적으로, 다양한 버스 아키텍처들 중 임의의 것을 사용하는, 메모리 버스, 메모리 제어기, 주변기기 버스, 로컬 버스, 및 이들의 임의의 조합을 포함하는 몇몇 유형들의 버스 구조들 중 임의의 것을 포함할 수 있다.

[0087] 프로세서(1204)는, 비제한적으로, 상태 머신에 의해 조절되고 메모리 및/또는 센서들로부터의 연산 입력들에 의해 지시될 수 있는 산술 논리 유닛(arithmetic and logic unit; ALU)과 같은 산술 및 논리 연산들을 수행하기 위한 논리 회로부를 통합하는 프로세서와 같은 임의의 적절한 프로세서를 포함할 수 있으며; 프로세서(1204)는 비제한적인 예로서 폰 노이만 및/또는 하버드 아키텍처에 따라 조직될 수 있다. 프로세서(1204)는, 비제한적으로, 마이크로제어기, 마이크로프로세서, 디지털 신호 프로세서(digital signal processor; DSP), 필드 프로그램 가능 게이트 어레이(Field Programmable Gate Array; FPGA), 복합 프로그램가능 논리 디바이스(Complex Programmable Logic Device; CPLD), 그래픽 프로세싱 유닛(Graphical Processing Unit; GPU), 범용 GPU, 텐서 프로세싱 유닛(Tensor Processing Unit; TPU), 아날로그 또는 혼합 신호 프로세서, 신뢰 플랫폼 모듈(Trusted Platform Module; TPM), 부동-소수점 유닛(floating-point unit; FPU), 및/또는 시스템 온 칩(system on a chip; SoC)을 포함하거나, 이를 통합하거나, 및/또는 이에 통합될 수 있다.

[0088] 메모리(1208)는, 비제한적으로, 랜덤-액세스 메모리 구성요소, 판독 전용 구성요소, 및 이들의 임의의 조합을 포함하는 다양한 구성요소들(예를 들어, 기계-판독가능 매체)을 포함할 수 있다. 일 예에서, 예컨대 기동 동안에 컴퓨터 시스템(1200) 내의 요소들 사이에서 정보를 전송하는 것을 돕는 기본 루틴들을 포함하는 기본 입력/출력 시스템(basic input/output system; BIOS)(1216)은 메모리(1208)에 저장될 수 있다. 메모리(1208)는 또한, 본 개시내용의 측면들 및/또는 방법론들 중 임의의 하나 이상을 구현하는 (예를 들어, 하나 이상의 컴퓨터-판독가능 매체에 저장된) 명령어들(예를 들어, 소프트웨어)(1220)을 포함할 수 있다. 다른 예에서, 메모리(1208)는, 비제한적으로, 운영 시스템, 하나 이상의 애플리케이션 프로그램들, 다른 프로그램 모듈들, 프로그램 데이터, 및 이들의 임의의 조합들을 포함하는 임의의 수의 프로그램 모듈들을 더 포함할 수 있다.

[0089] 컴퓨터 시스템(1200)은 또한 저장 디바이스(1224)를 포함할 수 있다. 저장 디바이스(예를 들어, 저장 디바이스(1224))의 예들은, 비제한적으로, 하드 디스크 드라이브, 자기 디스크 드라이브, 광학 매체와 조합된 광학 디스크 드라이브, 고체-상태 메모리 디바이스, 및 이들의 임의의 조합들을 포함한다. 저장 디바이스(1224)는 적절한 인터페이스(도시되지 않음)에 의해 버스(1212)에 연결될 수 있다. 예시적인 인터페이스들은, 비제한적으로, SCSI, 고급 기술 부착(advanced technology attachment; ATA), 직렬 ATA, 범용 직렬 버스(universal serial bus; USB), IEEE 1394(FIREWIRE), 및 이들의 임의의 조합들을 포함한다. 일 예에서, 저장 디바이스(1224)(또는 이의 하나 이상의 구성요소들)는 (예를 들어, 외부 포트 커넥터(도시되지 않음)를 통해) 컴퓨터 시스템(1200)과 제거가능하게 인터페이스될 수 있다. 특히, 저장 디바이스(1224) 및 연관된 기계-판독가능 매체(1228)는, 컴퓨터 시스템(1200)에 대한 기계-판독가능 명령어들, 데이터 구조들, 프로그램 모듈들, 및/또는 다른 데이터의 비휘발성 및 휘발성 저장을 제공할 수 있다. 일 예에서, 소프트웨어(1220)는 기계-판독가능 매체(1228) 내에 완전히 또는 부분적으로 상주할 수 있다. 다른 예에서, 소프트웨어(1220)는 프로세서(1204) 내에 완전히 또는 부분적으로 상주할 수 있다.

[0090] 컴퓨터 시스템(1200)은 또한 입력 디바이스(1232)를 포함할 수 있다. 일 예에서, 컴퓨터 시스템(1200)의 사용자

는 입력 디바이스(1232)를 통해 컴퓨터 시스템(1200)으로 명령들 및/또는 다른 정보를 입력할 수 있다. 입력 디바이스(1232)의 예들은, 비제한적으로, 문자-숫자 입력 디바이스(예를 들어, 키보드), 포인팅 디바이스, 조이스틱, 게임패드, 오디오 입력 디바이스(예를 들어, 마이크, 음성 응답 디바이스, 등), 커서 제어 디바이스(예를 들어, 마우스), 터치패드, 광학 스캐너, 비디오 캡처 디바이스(예를 들어, 스틸 카메라, 비디오 카메라), 터치 스크린, 및 이들의 임의의 조합들을 포함한다. 입력 디바이스(1232)는, 비제한적으로, 직렬 인터페이스, 병렬 인터페이스, 게임 포트, USB 인터페이스, FIREWIRE 인터페이스, 버스(1212)에 대한 직접 인터페이스, 및 이들의 임의의 조합들을 포함하는, 다양한 인터페이스들(도시되지 않음) 중 임의의 것을 통해 버스(1212)에 인터페이스될 수 있다. 입력 디바이스(1232)는, 이하에서 추가로 논의되는, 디스플레이(1236)의 부분이거나 또는 이와는 별개일 수 있는 터치 스크린 인터페이스를 포함할 수 있다. 입력 디바이스(1232)는, 이상에서 설명된 바와 같은 그래픽 인터페이스에서 하나 이상의 그래픽 표현들을 선택하기 위한 사용자 선택 디바이스로서 사용될 수 있다.

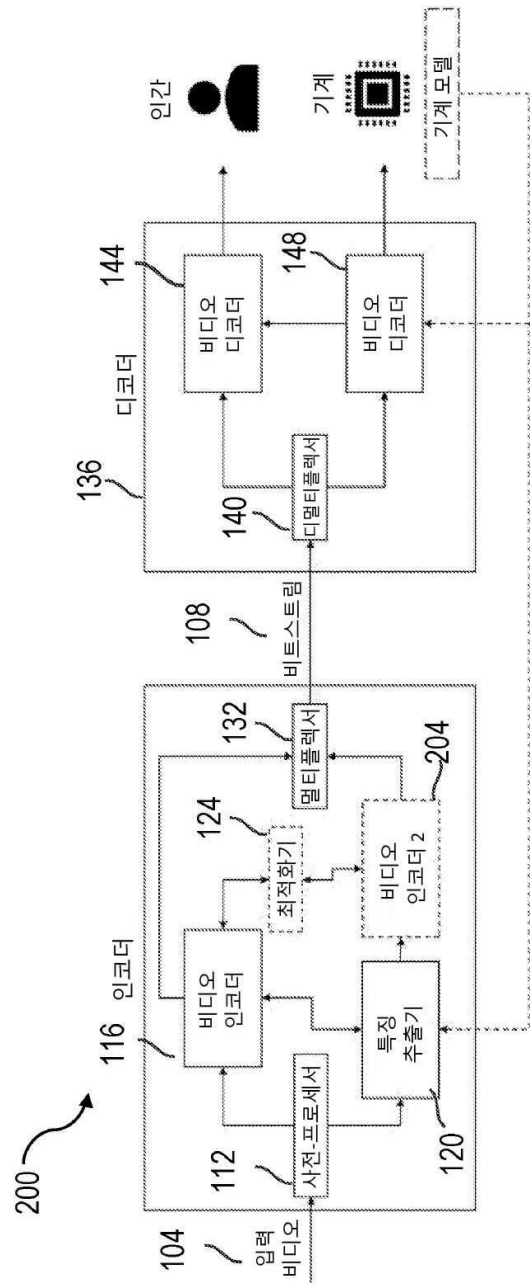
[0091] 사용자는 또한, 저장 디바이스(1224)(예를 들어, 제거가능 디스크 드라이브, 플래시 드라이브, 등) 및/또는 네트워크 인터페이스 디바이스(1240)를 통해 컴퓨터 시스템(1200)에 명령들 및/또는 다른 정보를 입력할 수 있다. 네트워크 인터페이스 디바이스(1240)와 같은 네트워크 인터페이스 디바이스는 컴퓨터 시스템(1200)을 네트워크(1244)와 같은 다양한 네트워크들 중 하나 이상 및 이에 연결된 하나 이상의 원격 디바이스들(1248)에 연결하기 위해 사용될 수 있다. 네트워크 인터페이스 디바이스의 예들은, 비제한적으로, 네트워크 인터페이스 카드(예를 들어, 모바일 네트워크 인터페이스 카드, LAN 카드), 모뎀, 및 이들의 임의의 조합을 포함한다. 네트워크의 예들은, 비제한적으로, 광역 네트워크(예를 들어, 인터넷, 기업 네트워크), 근거리 네트워크(예를 들어, 사무실, 건물, 캠퍼스 또는 다른 상대적으로 작은 지리적 공간과 연관된 네트워크), 전화 네트워크, 전화/음성 제공자와 연관된 데이터 네트워크(예를 들어, 모바일 통신 제공자 데이터 및/또는 음성 네트워크), 2개의 컴퓨팅 디바이스들 사이의 직접 연결, 및 이들의 임의의 조합들을 포함한다. 네트워크(1244)와 같은 네트워크는 무선 및/또는 유선 통신 모드를 이용할 수 있다. 일반적으로, 임의의 네트워크 토폴로지가 사용될 수 있다. 정보(예를 들어, 데이터, 소프트웨어(1220), 등)는 네트워크 인터페이스 디바이스(1240)를 통해 컴퓨터 시스템(1200)으로 및/또는 이로부터 통신될 수 있다. 컴퓨터 시스템(1200)은, 디스플레이 디바이스(1236)와 같은 디스플레이 디바이스에 디스플레이가능 이미지를 통신하기 위한 비디오 디스플레이 어댑터(1252)를 더 포함할 수 있다. 디스플레이 디바이스의 예들은, 비제한적으로, 액정 디스플레이(liquid crystal display; LCD), 음극선관(cathode ray tube; CRT), 플라즈마 디스플레이, 발광 다이오드(light emitting diode; LED) 디스플레이, 및 이들의 임의의 조합들을 포함한다.

[0092] 디스플레이 어댑터(1252)와 디스플레이 디바이스(1236)는 본 개시내용의 측면들의 그래픽 표현들을 제공하기 위해 프로세서(1204)와 조합되어 사용될 수 있다. 디스플레이 디바이스에 더하여, 컴퓨터 시스템(1200)은, 비제한적으로, 오디오 스피커, 프린터, 및 이들의 임의의 조합들을 포함하는 하나 이상의 다른 주변 출력 디바이스들을 포함할 수 있다. 이러한 주변 출력 디바이스들은 주변기기 인터페이스(1256)를 통해 버스(1212)에 연결될 수 있다. 주변기기 인터페이스의 예들은, 비제한적으로, 직렬 포트, USB 연결, FIREWIRE 연결, 병렬 연결, 및 이들의 임의의 조합들을 포함한다.

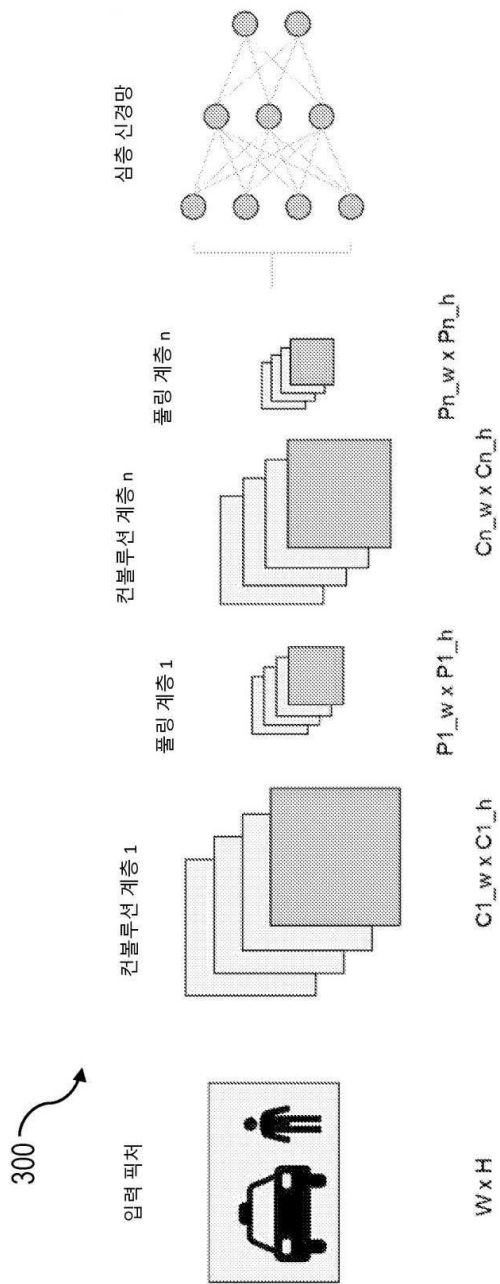
[0093] 이상의 내용은 본 발명의 예시적인 실시예들의 상세한 설명이었다. 다양한 수정들 및 추가들이 본 발명의 사상 및 범위로부터 벗어나지 않고 이루어질 수 있다. 이상에서 설명된 다양한 실시예들 각각의 특징들은 연관된 새로운 실시예들에서 다수의 특징 조합들을 제공하기 위해 적절하게 다른 설명된 실시예들의 특징들과 조합될 수 있다. 또한, 이상의 내용이 다수의 별도의 실시예들을 설명하지만, 본 명세서에서 설명된 것들은 단지 본 발명의 원리들의 적용의 예시일 뿐이다. 추가적으로, 본 명세서의 특정 방법들이 특정 순서로 수행되는 것으로 예시되거나 및/또는 설명될 수 있지만, 그 순서는 본 개시내용에 따른 방법들, 시스템들, 및 소프트웨어를 달성하기 위해 당업자들 내에서 매우 가변적이다. 따라서, 이러한 설명은 단지 예로서 취해진 것이며, 본 발명의 범위를 달리 제한하려는 것이 아니다.

[0094] 예시적인 실시예는 이상에서 개시되었고, 첨부된 도면들에 예시되었다. 본 발명의 사상 및 범위로부터 벗어나지 않고 본 명세서에 구체적으로 개시된 것들에 대해 다양한 변경들, 생략들 및 추가들이 이루어질 수 있다는 것이 당업자들에 의해 이해될 것이다.

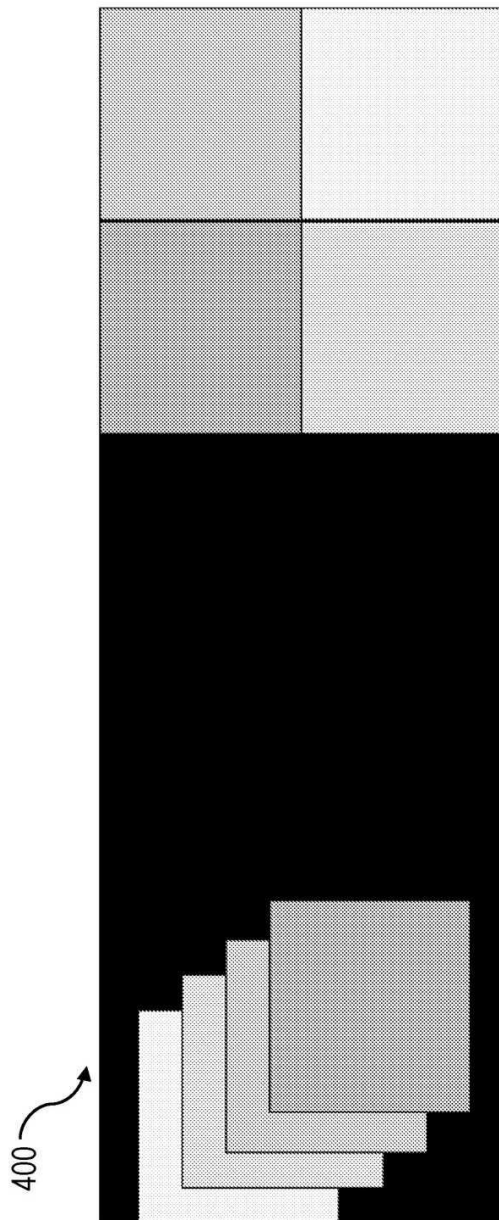
도면2



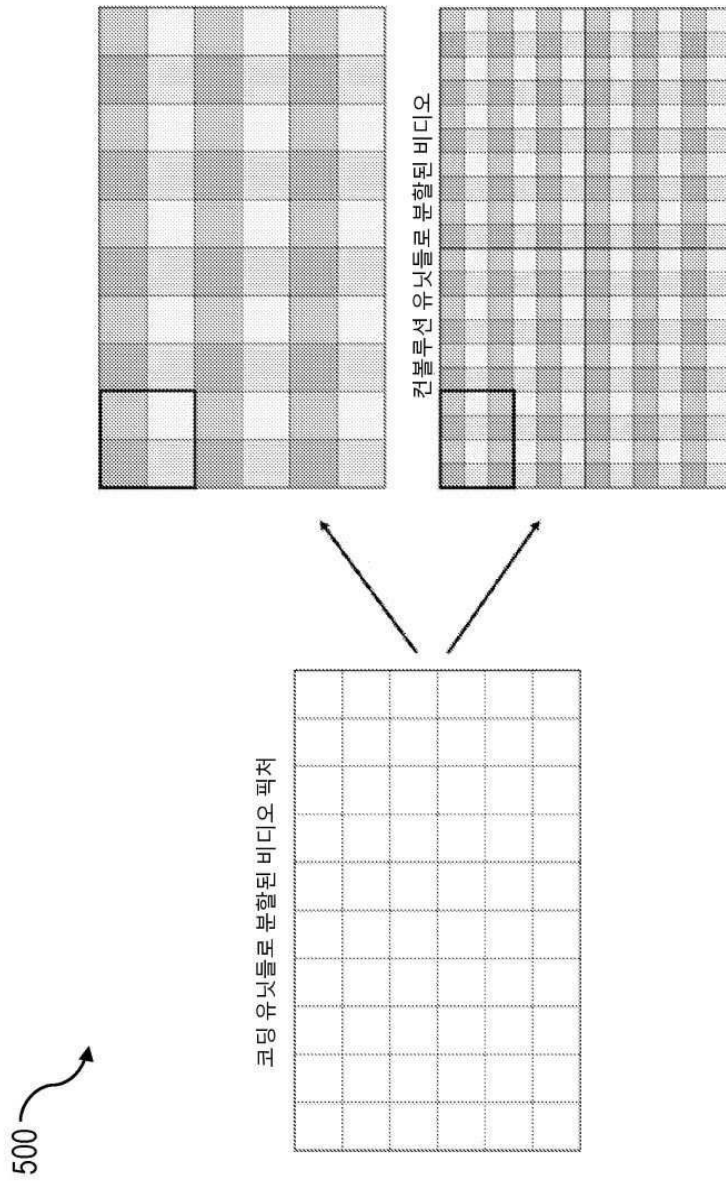
도면3



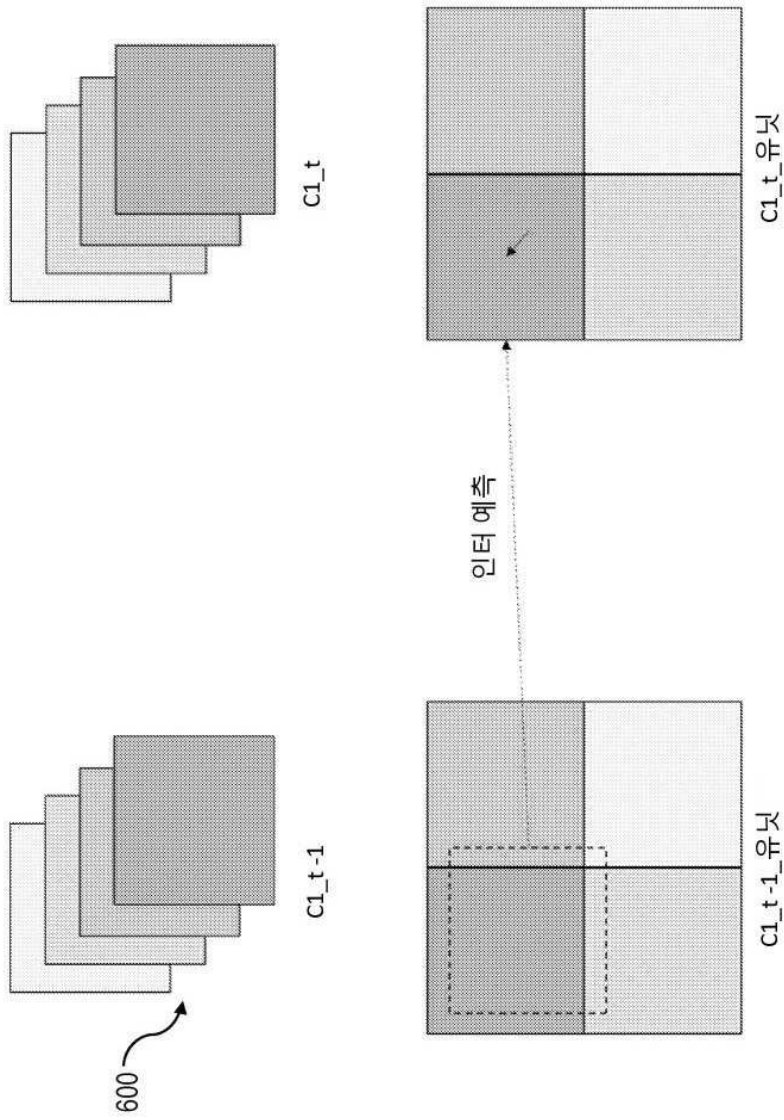
도면4



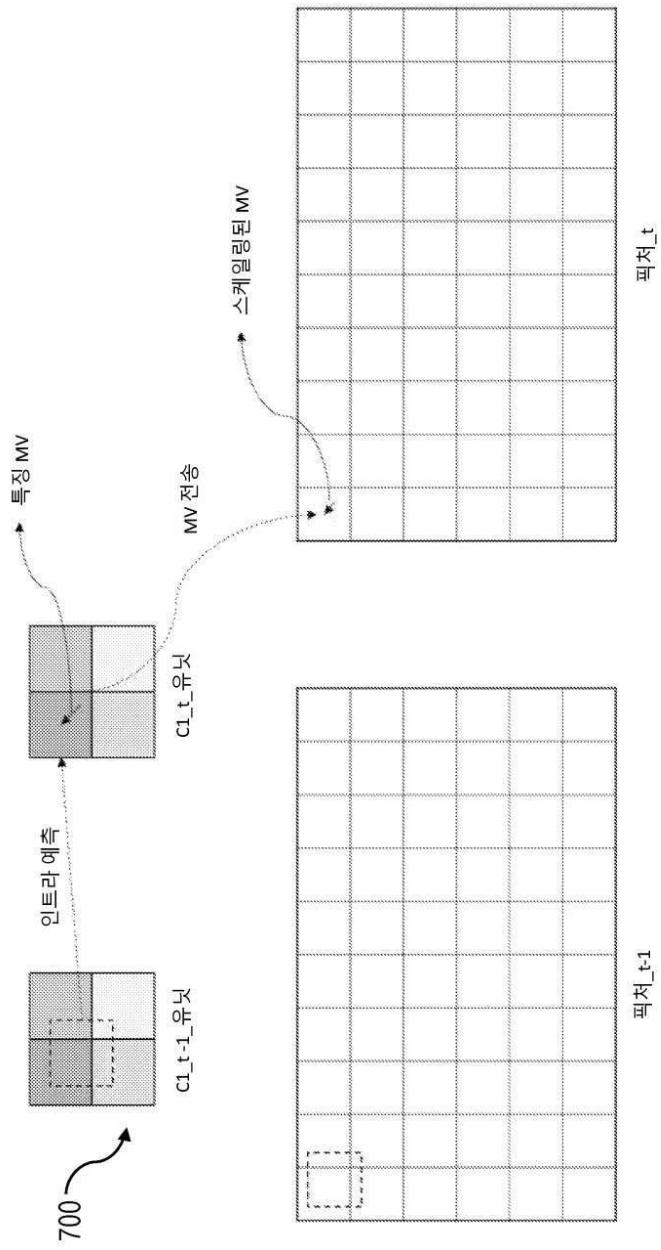
도면5



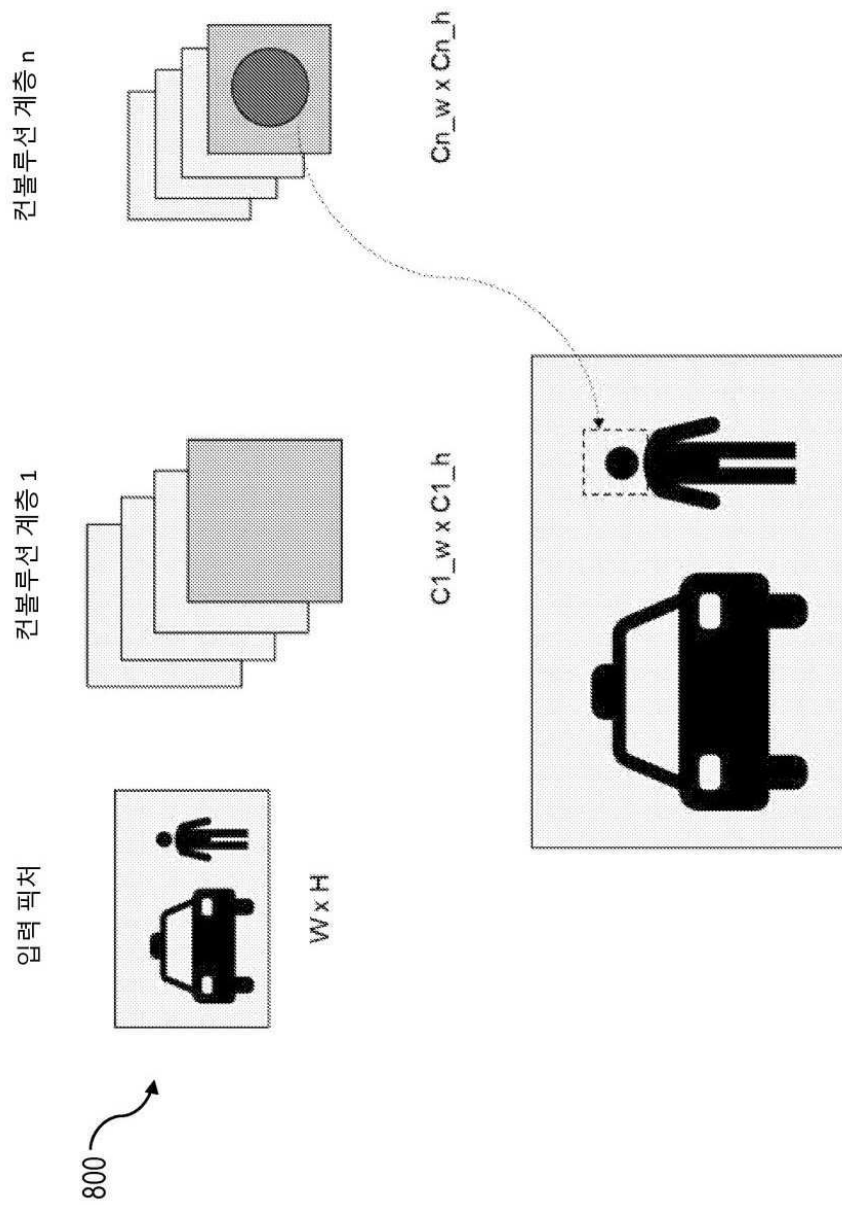
도면6



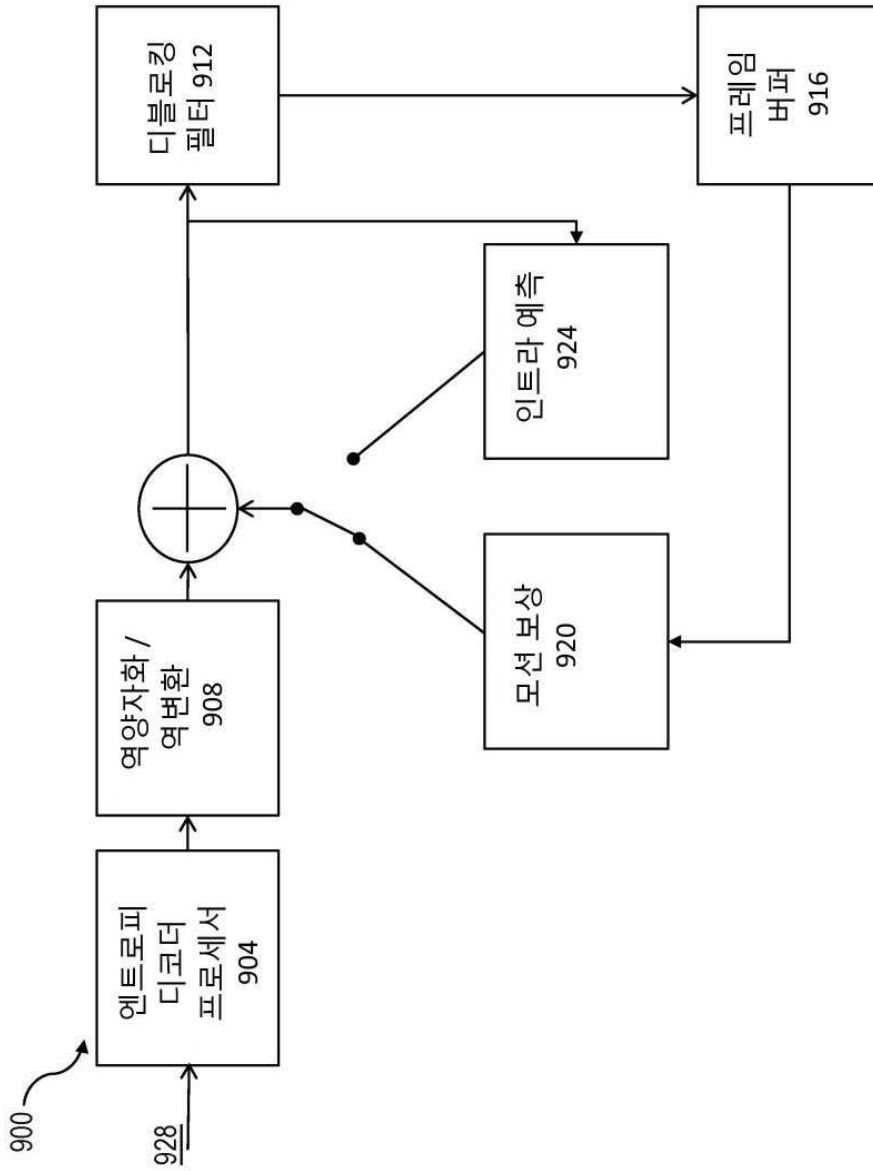
도면7



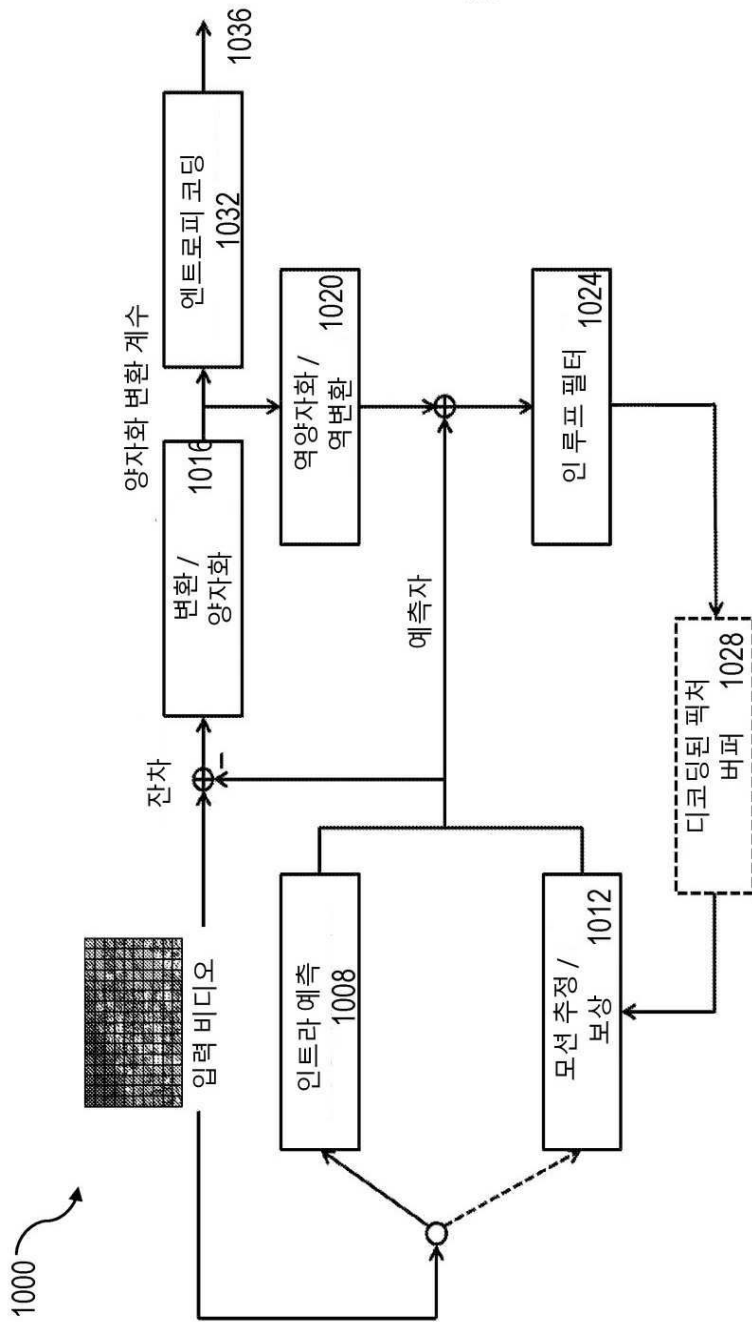
도면8



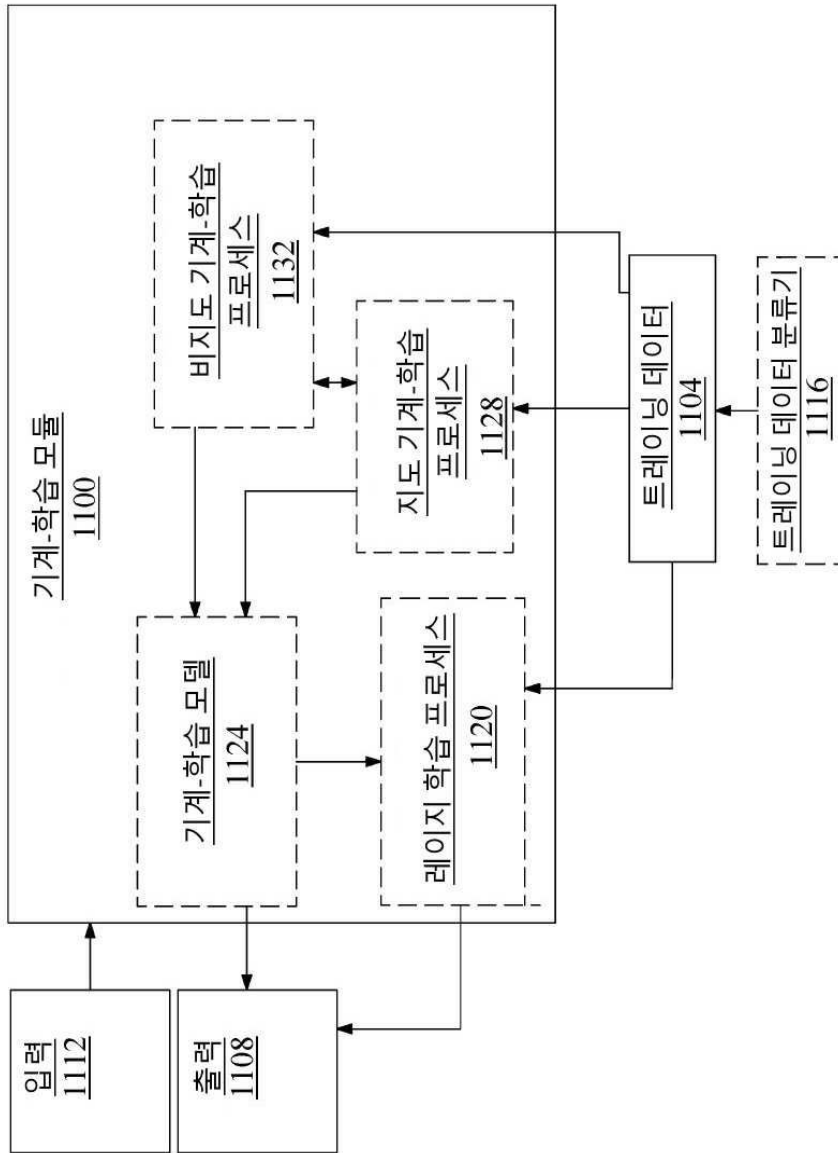
도면9



도면10



도면11



도면12

