



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0043912
(43) 공개일자 2022년04월05일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2021-0128887
(22) 출원일자 2021년09월29일
심사청구일자 없음
(30) 우선권주장
1020200127284 2020년09월29일 대한민국(KR)

(71) 출원인
현대자동차주식회사
서울특별시 서초구 현릉로 12 (양재동)
기아 주식회사
서울특별시 서초구 현릉로 12 (양재동)
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
강제원
서울특별시 마포구 서강로 53 서강쌍용예가아파트 104-904호
유채화
서울특별시 은평구 통일로 925, 409호
(뒷면에 계속)
(74) 대리인
이철희

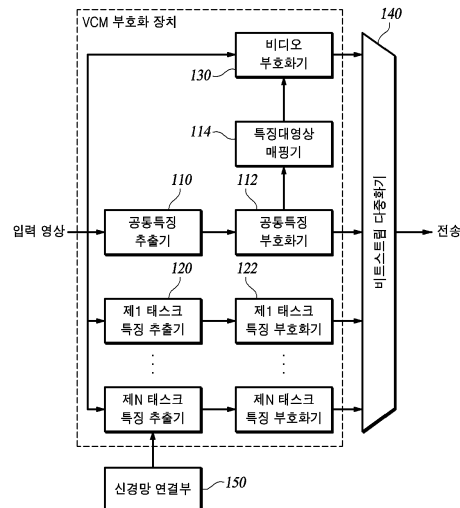
전체 청구항 수 : 총 26 항

(54) 발명의 명칭 머신 비전을 위한 다중 태스크 시스템에서의 딥러닝 기반 특징맵 코딩 장치 및 방법

(57) 요약

머신 비전을 위한 다중 태스크 시스템에서의 딥러닝 기반 특징맵 코딩 장치에 관한 개시로서, 본 실시예는, 머신 비전 및 인간의 시각 모두에 대해 비교적 용인할 수 있는 성능을 보장하기 위해, 원본 비디오가 함축하는 다중 태스크와 관련된 공통 특징맵을 생성하여 압축하되, 공통 특징맵을 이용하는 경우보다 개선된 성능이 필요 시, 태스크 특화 특징맵을 추가로 생성하여 압축하는 VCM 코딩 장치 및 방법을 제공한다.

대표도 - 도1



(72) 발명자

박승욱

경기도 용인시 수지구 태봉로 17, 403동 302호

임화평

경기도 화성시 남양읍 역골로 49번길 29 202호

명세서

청구범위

청구항 1

머신 비전을 위한 복호화 장치가 수행하는 복호화 방법에 있어서,

다중화 비트스트림을 획득하는 단계;

상기 다중화 비트스트림으로부터 제1 비트스트림을 획득하는 단계, 여기서, 제1 비트스트림은, 원본 영상이 함축하는 대표 태스크를 표현하는 공통 특징맵이 부호화되어 생성됨;

공통특징 복호화기를 이용하여 상기 제1 비트스트림으로부터 상기 공통 특징맵을 복호화하는 단계; 및

딥러닝 기반 영상복원 모델을 이용하여 상기 공통 특징맵으로부터 기본 영상을 생성하는 단계

를 포함하는 것을 특징으로 하는, 복호화 방법.

청구항 2

제1항에 있어서,

제2 비트스트림이 상기 다중화 비트스트림에 포함된 경우,

상기 다중화 비트스트림으로부터 상기 제2 비트스트림을 획득하는 단계, 여기서, 상기 제2 비트스트림은 상기 대표 태스크에 포함되는 적어도 하나의 개별 태스크를 표현하는 태스크 특화 특징맵이 부호화되어 생성됨; 및

태스크 특징 복호화기를 이용하여 상기 제2 비트스트림으로부터 상기 개별 태스크를 표현하는 태스크 특화 특징맵을 복호화하는 단계

를 더 포함하는 것을 특징으로 하는, 복호화 방법.

청구항 3

제2항에 있어서,

제3 비트스트림이 상기 다중화 비트스트림에 포함된 경우,

상기 다중화 비트스트림으로부터 상기 제3 비트스트림을 획득하는 단계, 여기서, 상기 제3 비트스트림은 적어도 하나의 잔여 태스크를 표현하는 태스크 특화 특징맵이 부호화되어 생성됨; 및

상기 태스크 특징 복호화기를 이용하여 상기 제3 비트스트림으로부터 상기 잔여 태스크를 표현하는 태스크 특화 특징맵을 복호화하는 단계

를 더 포함하는 것을 특징으로 하는, 복호화 방법.

청구항 4

제1항에 있어서,

제4 비트스트림이 상기 다중화 비트스트림에 포함된 경우,

상기 다중화 비트스트림으로부터 상기 제4 비트스트림을 획득하는 단계, 여기서, 상기 제4 비트스트림은 상기 원본 영상에서 상기 기본 영상을 감산하여 생성된 잔차 영상이 부호화되어 생성됨; 및

비디오 복호화기를 이용하여 상기 제4 비트스트림으로부터 상기 잔차 영상을 복호화한 후, 상기 잔차 영상과 상기 기본 영상을 가산하여 복원 영상을 생성하는 단계

를 더 포함하는 것을 특징으로 하는, 복호화 방법.

청구항 5

제2항에 있어서,

상기 공통특징 복호화기 및 상기 태스크 특징 복호화기는,

비디오 신호용 복호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현되는 것을 특징으로 하는, 복호화 방법.

청구항 6

제1항에 있어서,

상기 기본 영상을 생성하는 단계는,

딥러닝 기반 변환 모델을 이용하여 상기 공통 특징맵으로부터 변환 영상을 생성한 후, 상기 영상복원 모델을 이용하여 상기 변환 영상으로부터 상기 기본 영상을 생성하는 것을 특징으로 하는, 복호화 방법.

청구항 7

제6항에 있어서,

상기 변환 모델은 상기 원본 영상과 상기 변환 영상 간의 차이에 기반하는 손실함수를 이용하여 사전에 학습되고, 상기 영상복원 모델은 상기 원본 영상과 상기 기본 영상 간의 차이에 기반하는 손실함수를 이용하여 사전에 학습되는 것을 특징으로 하는, 복호화 방법.

청구항 8

머신 비전을 위한 부호화 장치가 수행하는 부호화 방법에 있어서,

원본 영상을 획득하는 단계;

딥러닝 기반 공통특징 추출 모델을 이용하여 상기 원본 영상으로부터 공통 특징맵을 추출하는 단계, 여기서, 상기 공통 특징맵은 상기 원본 영상이 함축하는 대표 태스크를 표현함;

공통특징 부호화기를 이용하여 상기 공통 특징맵을 부호화하여 제1 비트스트림을 생성하는 단계; 및

공통특징 복호화기를 이용하여 상기 제1 비트스트림으로부터 복원 공통 특징맵을 복호화한 후, 딥러닝 기반 영상복원 모델을 이용하여 상기 복원 공통 특징맵으로부터 기본 영상을 생성하는 단계

를 포함하는 것을 특징으로 하는, 부호화 방법.

청구항 9

제8항에 있어서,

상기 공통특징 추출 모델은 기본 신경망, 결정 신경망 및 변환 모델을 포함하고,

상기 공통 특징맵을 추출하는 단계는,

상기 기본 신경망을 이용하여 상기 원본 영상으로부터 상기 공통 특징맵을 추출하되, 상기 결정 신경망을 이용하여 상기 공통 특징맵을 기반으로 상기 대표 태스크의 분석 결과를 생성하고, 상기 변환 모델을 이용하여 상기 공통 특징맵으로부터 제1 변환 영상을 생성하는 것을 특징으로 하는, 부호화 방법.

청구항 10

제9항에 있어서,

상기 결정 신경망을 이용하여 상기 공통 특징맵을 기반으로 상기 대표 태스크를 분석하여 분석 결과를 생성하는 단계;

상기 분석 결과의 누적 신뢰도가 기설정된 임계치 미만인 경우, 딥러닝 기반 태스크 특징 추출 모델을 이용하여 상기 원본 영상으로부터 적어도 하나의 개별 태스크를 표현하는 태스크 특화 특징맵을 추출하는 단계, 여기서, 상기 적어도 하나의 개별 태스크는 상기 대표 태스크에 포함됨; 및

태스크 특징 부호화기를 이용하여 상기 개별 태스크를 표현하는 태스크 특화 특징맵을 부호화하여 제2 비트스트림을 생성하는 단계

를 더 포함하는 것을 특징으로 하는, 부호화 방법.

청구항 11

제10항에 있어서,

적어도 하나의 잔여 태스크가 존재하는 경우,

상기 태스크 특징 추출 모델을 이용하여 상기 원본 영상으로부터 상기 잔여 태스크를 표현하는 태스크 특화 특징맵을 추출하는 단계; 및

상기 태스크 특징 부호화기를 이용하여 상기 잔여 태스크를 표현하는 태스크 특화 특징맵을 부호화하여 제3 비트스트림을 생성하는 단계

를 더 포함하는 것을 특징으로 하는, 부호화 방법.

청구항 12

제11항에 있어서,

상기 원본 영상에서 상기 기본 영상을 감산하여 잔차 영상을 생성하는 단계; 및

비디오 부호화기를 이용하여 상기 잔차 영상을 부호화하여 제4 비트스트림을 생성하는 단계

를 포함하는 것을 특징으로 하는, 부호화 방법.

청구항 13

제12항에 있어서,

상기 제1 비트스트림, 상기 제2 비트스트림, 상기 제3 비트스트림, 및 상기 제4 비트스트림의 전부 또는 일부를 결합하여 다중화 비트스트림을 생성하는 단계를 더 포함하는 것을 특징으로 하는, 부호화 방법.

청구항 14

제9항에 있어서,

상기 기본 신경망은,

다중 태스크 딥러닝 모델로 구현되는 것을 특징으로 하는, 부호화 방법.

청구항 15

제9항에 있어서,

상기 공통특징 추출 모델은,

상기 분석 결과 및 대응되는 레이블 간의 차이에 기반하는 손실함수, 및 상기 제1 변환 영상과 상기 원본 영상 간의 차이에 기반하는 손실함수의 가중합을 이용하여 종단간(end-to-end)으로 학습되는 것을 특징으로 하는, 부호화 방법.

청구항 16

제8항에 있어서,

상기 공통특징 부호화기 및 상기 공통특징 복호화기는,

비디오 신호용 코덱(codec) 또는 딥러닝 기반 오토인코더를 이용하여 구현되되, 상기 오토인코더는 상기 공통 특징맵과 상기 복원 공통 특징맵 간의 차이에 기반하는 손실함수를 이용하여 학습되는 것을 특징으로 하는, 부호화 방법.

청구항 17

제10항에 있어서,

상기 태스크 특징 부호화기는,

비디오 신호용 부호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현되는 것을 특징으로 하는, 부호화 방법.

청구항 18

제11항에 있어서,

상기 태스크 특징 추출 모델은 태스크 신경망 및 결정 신경망을 포함하고,

상기 태스크 특화 특징맵을 추출하는 단계는,

상기 태스크 신경망을 이용하여 상기 원본 영상으로부터 상기 태스크 특화 특징맵을 추출하되, 상기 결정 신경망을 이용하여 상기 태스크 특화 특징맵을 기반으로 상기 개별 태스크 또는 상기 잔여 태스크의 분석 결과를 생성하는 것을 특징으로 하는, 부호화 방법.

청구항 19

제18항에 있어서,

상기 태스크 특징 추출 모델은,

상기 분석 결과 및 대응되는 레이블 간의 차이에 기반하는 손실함수 이용하여 학습되는 것을 특징으로 하는, 부호화 방법.

청구항 20

제12항에 있어서,

상기 비디오 부호화기는,

비디오 신호용 부호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현되는 것을 특징으로 하는, 부호화 방법.

청구항 21

제9항에 있어서,

상기 기본 영상을 생성하는 단계는,

상기 변환 모델을 이용하여 상기 복원 공통 특징맵으로부터 제2 변환 영상을 생성한 후, 상기 영상복원 모델을 이용하여 상기 제2 변환 영상으로부터 상기 기본 영상을 생성하는 것을 특징으로 하는, 부호화 방법.

청구항 22

제21항에 있어서,

상기 영상복원 모델은,

상기 원본 영상과 상기 기본 영상 간의 차이에 기반하는 손실함수를 이용하여 학습되는 것을 특징으로 하는, 부호화 방법.

청구항 23

다중화 비트스트림으로부터 제1 비트스트림을 획득하는 역다중화기, 여기서, 제1 비트스트림은, 원본 영상이 함축하는 대표 태스크를 표현하는 공통 특징맵이 부호화되어 생성됨;

상기 제1 비트스트림으로부터 상기 공통 특징맵을 복호화하는 공통특징 복호화기; 및

딥러닝 기반 영상복원 모델을 이용하여 상기 공통 특징맵으로부터 기본 영상을 생성하는 특징대영상 매핑기를 포함하는 것을 특징으로 하는, 머신 비전을 위한 복호화 장치.

청구항 24

제23항에 있어서,

제2 비트스트림이 상기 다중화 비트스트림에 포함된 경우, 상기 역다중화기가 상기 다중화 비트스트림으로부터 생성하는 제2 비트스트림을 획득하여, 상기 제2 비트스트림으로부터 상기 대표 태스크에 포함되는 적어도 하나의 개별 태스크를 표현하는 태스크 특화 특징맵을 복호화하는 적어도 하나의 태스크 특징 복호화기를 더 포함하되,

상기 제2 비트스트림은 상기 개별 태스크를 표현하는 태스크 특화 특징맵이 부호화된 것을 특징으로 하는, 머신 비전을 위한 복호화 장치.

청구항 25

제24항에 있어서,

상기 태스크 특징 복호화기는,

제3 비트스트림이 상기 다중화 비트스트림에 포함된 경우, 상기 역다중화기가 상기 다중화 비트스트림으로부터 생성하는 제3 비트스트림을 획득하여, 상기 제3 비트스트림으로부터 적어도 하나의 잔여 태스크를 표현하는 태스크 특화 특징맵을 복호화하되, 상기 제3 비트스트림은 상기 잔여 태스크를 표현하는 태스크 특화 특징맵이 부호화된 것을 특징으로 하는, 머신 비전을 위한 복호화 장치.

청구항 26

제23항에 있어서,

제4 비트스트림이 상기 다중화 비트스트림에 포함된 경우, 상기 역다중화기가 상기 다중화 비트스트림으로부터 생성하는 제4 비트스트림을 획득하여, 상기 제4 비트스트림으로부터 잔차 영상을 복호화한 후, 상기 잔차 영상과 상기 기본 영상을 가산하여 복원 영상을 생성하는 비디오 복호화기를 더 포함하되,

상기 제4 비트스트림은 상기 원본 영상에서 상기 기본 영상을 감산하여 생성된 잔차 영상이 부호화된 것을 특징으로 하는, 머신 비전을 위한 복호화 장치.

발명의 설명

기술 분야

[0001] 본 개시는 머신 비전을 위한 다중 태스크 시스템에서의 딥러닝 기반 특징맵 코딩 장치 및 방법에 관한 것이다.

배경 기술

[0002] 이하에 기술되는 내용은 단순히 본 발명과 관련되는 배경 정보만을 제공할 뿐 종래기술을 구성하는 것이 아니다.

[0003] 딥러닝 기술의 발전과 컴퓨팅 파워의 발전과 함께 머신 비전 어플리케이션이 폭발적으로 증가함에 따라 비디오 트래픽(video traffic)의 대부분이 기계에 의해 이용될 것이라는 인식이 확산되고 있다. 머신 투 머신 애플리케이션은 향후 인터넷 비디오 트래픽의 가장 큰 사용량을 차지할 것으로 예상된다. 따라서, 기계가 사용하는 비디오 데이터의 정보를 최적화하는 방안은 비디오 처리 기술의 혁신 및 신규 솔루션의 상용화 측면에서 핵심 요소가 될 수 있다.

[0004] 기존의 비디오 코딩 방식은 특정 비트 전송률 제약 조건 하에서 최고의 영상 또는 영상 품질을 목표로 하므로, 인간의 시각에 최적화되어 있다. 반면, 머신 비전을 위한 코딩의 경우, 복원 영상/비디오가 반드시 높은 시각적 성능을 가질 필요는 없다. 지연 시간(latency)과 규모(scale)에 대한 엄격한 제한사항이 있는 커넥티드 차량, IoT(Internet of Things) 장치, 초대형 비디오 감시 네트워크, 스마트 시티, 품질 검사 등의 분야가 출현함에 따라 머신 비전을 위한 새로운 패러다임이 등장하여 머신 비전을 목표로 하는 신규 이미지/비디오 코딩 방법의 필요성이 대두되었다.

[0005] 이에 따라, 표준화 기구인 MPEG(Moving Picture Expert Group)에서는 머신 비전을 위한 표준화의 필요성이 논의되고, 머신 비전을 위한 압축 코딩, 및 휴먼-머신 하이브리드 시야를 위한 압축 코딩을 제공하는 차세대 비디오 코덱으로서 VCM (Video Coding for Machines)이 제안되었다.

- [0006] VCM 코덱의 구조에는 다양한 변형이 있을 수 있으나, 기본적인 VCM 코덱의 구조는 도 16에 예시된 바와 같다. 센서의 출력인 비디오가 입력되었을 때, VCM 부호화기는 머신 비전을 위한 정보로서 특징을 추출(feature extraction)하고, 이를 필요에 맞게 변형(feature conversion)한 후, 부호화(feature encoding)한다. 또한, VCM 부호화기는 입력 영상 또는 비디오를 부호화할 때 부호화된 특징을 참조할 수 있다. 최종적으로, VCM 부호화기는 머신 비전을 위한 특징과 입력 영상(또는, 잔차 영상)을 부호화하여 비트스트림을 생성한다. VCM 부호화기는 특징 및 비디오를 부호화하여 생성되는 각각의 비트스트림을 다중화(multiplexing)하여 함께 전송한다.
- [0007] VCM 복호화기는 전송된 비트스트림으로부터 특징 비트스트림과 비디오 비트스트림으로 역다중화(demultiplexing)한 후, 특징과 비디오를 각각 복호화한다. 이때, VCM 복호화기는, 비디오의 복호화 시, 복원된 특징을 참조할 수 있다. 복원된 특징은 머신 비전과 인간의 시각에 동시에 사용될 수 있다.
- [0008] 한편, 자율주행(self-driving) 시스템은 VCM 기술의 활용 사례 중 대표적인 다중 태스크 시스템이다. 이때 기계에 의해 수행되는 다중 태스크는 다중 객체 탐지(multiple object detection), 객체 분할(object segmentation), 객체(예컨대, 차선(line)) 추적(tracking), 활동 인식(action recognition 또는 action localization), 이벤트 예측(event prediction) 등을 포함한다. 일반적으로, 카메라, 적외선, 라이다(LiDAR), 레이더, 초음파 등의 센서들로부터 취득한 비디오를 이용하여, 전술한 태스크별로 단일 태스크 딥러닝 모델을 학습시킨 후, 학습된 단일 태스크 모델들을 이용하여 기계는 해당되는 태스크 각각을 수행할 수 있다.
- [0009] 태스크별 단일 태스크 모델을 학습시키고, 학습 모델의 특징맵을 머신 비전을 위한 정보로서 압축 및 전송할 경우, 태스크의 개수가 증가할수록 학습이 필요한 모델의 수, 및 전송해야 하는 정보의 양도 비례해서 증가한다는 문제가 있다. 따라서, VCM 기술을 활용함에 있어서, 부호화 효율을 향상시키고 비용을 절감하기 위해, 다중 태스크 시스템에 적합한 딥러닝 모델 및 그에 대응되는 학습방법을 고려하는 것이 필요하다.

선행기술문헌

비특허문헌

- [0010] (비특허문헌 0001) Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3712-3722).
- (비특허문헌 0002) Ruder, S. (2017). An overview of multi-task learning in deep neural networks. ARXIV:1706.05098.

발명의 내용

해결하려는 과제

- [0011] 본 개시는, 머신 비전 및 인간의 시각 모두에 대해 비교적 용인할 수 있는 성능을 보장하기 위해, 원본 비디오가 함축하는 다중 태스크와 관련된 공통 특징맵을 생성하여 압축하되, 공통 특징맵을 이용하는 경우보다 개선된 성능이 필요 시, 태스크 특화 특징맵을 추가로 생성하여 압축하는 VCM 코딩 장치 및 방법을 제공하는 데 목적이 있다.

과제의 해결 수단

- [0012] 본 개시의 실시예에 따르면, 머신 비전을 위한 복호화 장치가 수행하는 복호화 방법에 있어서, 다중화 비트스트림을 획득하는 단계; 상기 다중화 비트스트림으로부터 제1 비트스트림을 획득하는 단계, 여기서, 제1 비트스트림은, 원본 영상이 함축하는 대표 태스크를 표현하는 공통 특징맵이 부호화되어 생성됨; 공통특징 복호화기를 이용하여 상기 제1 비트스트림으로부터 상기 공통 특징맵을 복호화하는 단계; 및 딥러닝 기반 영상복원 모델을 이용하여 상기 공통 특징맵으로부터 기본 영상을 생성하는 단계를 포함하는 것을 특징으로 하는, 복호화 방법을 제공한다.
- [0013] 본 개시의 다른 실시예에 따르면, 머신 비전을 위한 부호화 장치가 수행하는 부호화 방법에 있어서, 원본 영상을 획득하는 단계; 딥러닝 기반 공통특징 추출 모델을 이용하여 상기 원본 영상으로부터 공통 특징맵을 추출하는 단계, 여기서, 상기 공통 특징맵은 상기 원본 영상이 함축하는 대표 태스크를 표현함; 공통특징 부호화기를

이용하여 상기 공통 특징맵을 부호화하여 제1 비트스트림을 생성하는 단계; 및 공통특징 복호화기를 이용하여 상기 제1 비트스트림으로부터 복원 공통 특징맵을 복호화한 후, 딥러닝 기반 영상복원 모델을 이용하여 상기 복원 공통 특징맵으로부터 기본 영상을 생성하는 단계를 포함하는 것을 특징으로 하는, 부호화 방법을 제공한다.

[0014] 본 개시의 다른 실시예에 따르면, 다중화 비트스트림으로부터 제1 비트스트림을 획득하는 역다중화기, 여기서, 제1 비트스트림은, 원본 영상이 함축하는 대표 태스크를 표현하는 공통 특징맵이 부호화되어 생성됨; 상기 제1 비트스트림으로부터 상기 공통 특징맵을 복호화하는 공통특징 복호화기; 및 딥러닝 기반 영상복원 모델을 이용하여 상기 공통 특징맵으로부터 기본 영상을 생성하는 특징대영상 매핑기를 포함하는 것을 특징으로 하는, 머신 비전을 위한 복호화 장치를 제공한다.

발명의 효과

[0015] 이상에서 설명한 바와 같이 본 실시예에 따르면, 원본 비디오가 함축하는 다중 태스크와 관련된 공통 특징맵을 생성하는 VCM 코딩 장치 및 방법을 제공함으로써, 머신 비전 및 인간의 시각 모두에 대해 비교적 용인할 수 있는 성능을 보장하고, 원본 비디오를 전송하는 것에 비해 적은 비용으로 전송하는 것이 가능해지는 효과가 있다.

[0016] 또한 본 실시예에 따르면, 원본 비디오가 포함하는 다중 태스크에 대한 공통 특징맵을 생성하여 압축하되, 태스크 특화 특징맵을 추가로 생성하여 압축하는 VCM 코딩 장치 및 방법을 제공함으로써, 머신 비전 또는 인간의 시각 관점에서 공통 특징맵을 수신했을 때보다 개선된 성능을 보장하는 것이 가능해지는 효과가 있다.

[0017] 또한 본 실시예에 따르면, 원본 비디오가 포함하는 다중 태스크에 대한 공통 특징맵을 생성하여 압축하되, 태스크 특화 특징맵을 추가로 생성하여 압축하는 VCM 코딩 장치 및 방법을 제공함으로써, VCM 코딩 장치가 수행하는 태스크의 개수에 제한을 없애고, 태스크가 추가되거나 삭제되더라도 VCM 코딩 장치의 구조를 유지하는 것이 가능해지는 효과가 있다.

도면의 간단한 설명

[0018] 도 1은 본 개시에 일 실시예에 따른 VCM 부호화 장치를 개념적으로 나타내는 블록도이다.

도 2는 본 개시의 일 실시예에 따른 공통특징 추출기(110)를 개념적으로 나타내는 블록도이다.

도 3a 및 도 3b는 본 개시의 일 실시예에 따른 다중 태스크 모델을 개념적으로 나타내는 예시도이다.

도 4는 본 개시의 일 실시예에 따른 딥러닝 기반 변환 모델을 나타내는 예시도이다.

도 5는 본 개시의 일 실시예에 따른, 공통 특징맵의 부호화 및 복호화를 수행하는 오토인코더를 나타내는 예시도이다.

도 6은 본 개시의 일 실시예에 따른 딥러닝 기반 영상복원 모델을 나타내는 예시도이다.

도 7은 본 개시의 일 실시예에 따른 태스크 특징 추출기(120)를 개념적으로 나타낸 블록도이다.

도 8은 본 개시의 일 실시예에 따른, 태스크 특화 특징맵의 부호화 및 복호화를 수행하는 오토인코더를 나타내는 예시도이다.

도 9는 본 개시의 일 실시예에 따른, 잔차 영상의 부호화 및 복호화를 수행하는 오토인코더를 나타내는 예시도이다.

도 10은 본 개시에 일 실시예에 따른 VCM 복호화 장치를 개념적으로 나타내는 예시도이다.

도 11은 본 개시의 다른 실시예에 따른 VCM 부호화 장치를 개념적으로 나타내는 블록도이다.

도 12는 본 개시의 다른 실시예에 따른 VCM 복호화 장치를 개념적으로 나타내는 블록도이다.

도 13은 본 개시의 또다른 실시예에 따른 VCM 코덱을 개념적으로 나타내는 블록도이다.

도 14는 본 개시의 일 실시예에 따른 VCM 부호화 방법을 나타내는 순서도이다.

도 15는 본 개시의 일 실시예에 따른 VCM 복호화 방법을 나타내는 순서도이다.

도 16은 본 개시의 일 실시예에 따른 VCM 코덱을 개념적으로 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0019] 이하, 본 발명의 실시예들을 예시적인 도면을 참조하여 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 실시예들을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 실시예들의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [0020] 본 실시예는 머신 비전을 위한 다중 태스크 시스템에서의 딥러닝 기반 특징맵 코딩 장치 및 방법에 관한 내용을 개시한다. 보다 자세하게는, 머신 비전 및 인간의 시각 모두에 대해 비교적 용인할 수 있는 성능을 보장하기 위해, 원본 비디오가 함축하는 다중 태스크와 관련된 공통 특징맵을 생성하여 압축하되, 공통 특징맵을 이용하는 경우보다 개선된 성능이 필요 시, 태스크 특화 특징맵을 추가로 생성하여 압축하는 VCM(Video Coding for Machines) 코딩 장치 및 방법을 제공한다.
- [0021] 여기서, VCM 코딩 장치 또는, VCM 코덱은 VCM 부호화 장치와 VCM 복호화 장치를 포함한다.
- [0022] 이하의 설명에서, 머신 비전을 위한 다중 태스크들로부터 특징맵을 추출, 부호화, 및 전송하는 장치 및 방법을 VCM 부호화 장치 및 방법으로 나타내고, 수신된 비트스트림으로부터 특징맵을 복호화하는 장치 및 방법을 VCM 복호화 장치 및 방법으로 나타낸다. 따라서, 본 개시에 따른 VCM 부호화 장치 및 VCM 부호화 장치는 다중 태스크를 수행하는 다중 태스크 시스템의 일 예일 수 있다.
- [0023] 한편, 인간의 시각에 최적화되도록 비디오 신호를 부호화 및 복호화하는 기존의 코덱을 비디오 부호화기 및 비디오 복호화기로 나타낸다.
- [0024] 이하의 설명에서, VCM 부호화 장치 및 복호화 장치가 처리하는 태스크의 개수를 N (여기서, N 은 자연수)으로 나타낸다. 이때, 전체 태스크는 태스크 간 유사도에 따라 S (여기서, S 는 자연수) 개의 부분 태스크 집합 $T = \{T_1, T_2, \dots, T_S\}$ 로 분할된다고 가정한다. $T_1, T_2, \dots, T_S$ 가 서로 소(즉, disjoint)임을 가정하면, $n(T_1) + n(T_2) + \dots + n(T_S) = N$ 을 만족한다.
- [0025] 이때, $T_1, T_2, \dots, T_S$ 중 최대 개수의 원소를 갖는 집합을 T^* 로 나타내고, 대표 태스크 집합으로 정의한다. 여기서, 대표 태스크 집합 T^* 의 원소의 개수는 $M (=n(T^*))$, M 은 자연수)로 나타낸다. 대표 태스크 집합에 포함된 개별(individual) 태스크들을 통칭하여 대표 태스크(representative task)로 정의한다. 대표 태스크 집합의 여집합($T - T^*$)을 잔여 태스크 집합으로 정의한다. 잔여 태스크 집합에 포함되는 태스크들 각각을 잔여 태스크(residue task)라고 정의한다. 따라서, 잔여 태스크의 개수는 $N - M$ 개이다. 전체 태스크와 대표 태스크가 상호 일치하는 경우, 잔여 태스크는 존재하지 않을 수 있다.
- [0026] 한편, 본 실시예에 있어서, 전체 태스크에 대해 대표 태스크의 집합은 하나 또는 그 이상이 존재할 수 있다.
- [0027] 대표 태스크 집합에 포함된 개별 태스크들의 분석에 공통적으로 많이 활용하는 특징맵을 공통 특징맵(common feature map)으로 나타낸다. VCM 부호화 장치 또는 VCM 복호화 장치는 공통 특징맵을 이용하여 대표 태스크 집합에 포함된 개별 태스크에 대한 분석을 수행할 수 있다. 하나의 대표 태스크 집합에 대하여 공통 특징맵은 하나 또는 그 이상이 존재할 수 있다.
- [0028] VCM 부호화 장치 또는 VCM 복호화 장치는, 필요 시, 개별 태스크 각각의 태스크 특화 특징을 이용하여 더 우수한 태스크 분석 결과, 즉, 머신 비전 성능을 제공할 수 있다.
- [0029] 반면, 잔여 태스크의 분석에는 공통 특징맵을 사용하지 않기 때문에, 잔여 태스크 각각의 태스크 특화 특징맵(task-specific feature map)을 이용하여 태스크 분석이 수행될 수 있다. 잔여 태스크의 개수에 따라 태스크 특화 특징맵과 관련된 과정은 선형적으로 증가한다. 따라서, 잔여 태스크 집합의 크기가 작을수록, 즉 잔여 태스크의 개수가 적을수록 압축 효율 및 소요 시간 측면에서 유리하다.
- [0030] 전체 태스크를 부분 태스크의 집합으로 분할하기 위해 이용되는 태스크 간 유사도는 두 태스크 간의 이전성(transferability)을 나타내는 유사도 행렬(affinity matrix)로부터 측정될 수 있다. 여기서, 두 태스크 간의 이전성은, 소스 태스크(source task)를 대상으로 학습된 신경망 모델의 특징의 표현(feature representation)을 타겟 태스크(target task)에 그대로 적용하여 사후적으로 학습했을 때와, 타겟 태스크를 단독으로 학습했을 때 간의 성능 향상 수준을 나타낸다(비특허문헌 1 참조).
- [0031] 한편, 태스크의 개수 N 이 1인 경우, VCM 부호화 장치 및 VCM 복호화 장치는 도 16에 예시된 바와 동일한 단일

태스크 시스템일 수 있다.

- [0032] 전술한 바와 같은, 전체 태스크를 대표 태스크 및 잔여 태스크로 분할하는 과정은 VCM 부호화 및 VCM 복호화 이전에 수행될 수 있다.
- [0033] 도 1은 본 개시에 일 실시예에 따른 VCM 부호화 장치를 개념적으로 나타내는 블록도이다.
- [0034] VCM 부호화 장치는 센서의 출력에 해당하는 비디오 데이터, 또는 입력 영상을 획득한다. VCM 부호화 장치는 입력 영상으로부터 대표 태스크의 공통 특징맵을 추출하고, 이를 부호화한다. VCM 부호화 장치는 필요 시, 대표 태스크에 포함되는 개별 태스크들의 특화 특징맵을 추출하고, 이를 부호화한다. VCM 부호화 장치는, 잔여 태스크들의 태스크 특화 특징맵을 추출하고, 이를 부호화한다. VCM 부호화 장치는 필요 시, 공통 특징맵을 이용하여 기본 영상을 생성하고, 비디오 데이터로부터 기본 영상을 감산하여 잔차 영상을 생성한 후, 이를 부호화한다. VCM 부호화 장치는 공통 특징맵, 개별 태스크들의 태스크 특화 특징맵, 잔여 태스크들의 태스크 특화 특징맵, 및 잔차 영상을 부호화한 비트스트림을 다중화(multiplexing)한 후, 이를 VCM 복호화 장치로 전송한다.
- [0035] VCM 부호화 장치는, 도 1에 예시된 바와 같이, 공통특징 추출기(110), 공통특징 부호화기(112), 특징대영상 매핑기(feature to image mapper, 114), N 개의 태스크 특징 추출기(120), N 개의 태스크 특징 부호화기(122), 비디오 부호화기(130), 다중화기(multiplexer, 140), 및 신경망 연결부(interface unit for Neural Network, 150)의 전부 또는 일부를 포함한다. 여기서, 본 실시예에 따른 VCM 부호화 장치에 포함되는 구성요소가 반드시 이에 한정되는 것은 아니다. 예컨대, VCM 부호화 장치가 포함하는 다수의 딥러닝 모델을 트레이닝하기 위해, 외부의 트레이닝부와 연동되는 형태로 구현될 수 있다.
- [0036] 도 2는 본 개시의 일 실시예에 따른 공통특징 추출기(110)를 개념적으로 나타내는 블록도이다.
- [0037] 공통특징 추출기(110)는, 딥러닝 기반으로, 입력 영상으로부터 대표 태스크의 공통 특징맵을 생성하고, 대표 태스크에 포함되는 개별 태스크들의 분석 결과를 생성한다. 공통특징 추출기(110)는 딥러닝 기반으로 공통 특징맵으로부터 변환 영상을 생성한다. 공통특징 추출기(110)는 기본 신경망(basic neural network, 202), M 개의 결정(decision) 신경망(204), 및 특징구조 변환기(feature-structure transformer, 206)의 전부 또는 일부를 포함한다.
- [0038] 이하의 설명에서, 입력 영상과 원본 영상은 동일한 의미로 사용된다.
- [0039] 기본 신경망(202)은 입력 영상으로부터 공통 특징맵 f_c 를 생성한다.
- [0040] 기본 신경망(202)은 다중 태스크 학습이 수행된 딥러닝 모델로서, 도 3a 및 도 3b에 예시된 바와 같은 다중 태스크 딥러닝 모델(비특허문헌 2 참조)로 구현될 수 있다. 이때, 기본 신경망(202)은 영상 처리에 적합한 CNN(Convolutional Neural Network) 기반 딥러닝 모델로 구현될 수 있다.
- [0041] 일반적으로 하나 이상의 목적을 달성하기 위해 딥러닝 기반 학습 모델 및 학습 메트릭(metric)을 구성하여 모델을 학습하는 것을 다중 태스크 학습(Multi-task Learning: MTL)이라고 한다. 단일 태스크를 위해 특화된 시스템에 비하여 복수의 목적을 달성하기 위해 설계된 딥러닝 기반 다중 태스크 시스템은 성능이 열화될 수 있다. 다중 태스크 학습은 각 단일 태스크를 위해 학습한 표현(representation)을 공유함으로써, 학습 모델이 다중 태스크에 적응하도록 일반화하는 학습 방법이다. 다중 태스크 학습은 joint learning, learning to learn, 또는 learning with auxiliary tasks로도 명칭되며, 하나보다 많은 태스크에 대한 성능을 최적화하는 것을 목표로 한다.
- [0042] 딥러닝 기반 다중 태스크 학습은, 은닉층(hidden layer)에 포함되는 파라미터의 공유에 따라, 하드 파라미터 공유(hard parameter sharing) 또는 소프트(soft) 파라미터 공유라는 두 가지 방법을 이용한다. 전자는, 도 3a의 예시와 같이, 단일 태스크 특화 출력층(output layer)을 제외하고 모든 태스크 간에 은닉층을 공유하는 방식으로, 학습 과정에서 과적합(overfitting)을 감소시킬 수 있다. 반면, 후자는, 도 3b의 예시와 같이, 각각의 태스크가 자신의 파라미터를 포함하는 모델을 이용하되, 모델의 파라미터 간 거리를 정규화(regularization)한다.
- [0043] 기본 신경망(202)이 생성하는 공통 특징맵은 각각의 태스크에 공통적으로 특화된 특징이고, 다중 태스크 학습에 기반하여 태스크 간 공유될 수 있는 가장 대표적인 정보를 포함할 수 있다.
- [0044] M 개의 결정 신경망(204)은, 공통 특징맵을 기반으로 개별 태스크와 관련된 분석 결과로서 M 개의 출력 $y_1, y_2, \dots, y_M$ 을 생성한다. 이 분석 결과는 추후 개별 태스크의 태스크 특화 특징맵의 검출 및 부호화의 수행 여부를

결정하는데 이용될 수 있다.

[0045] M 개의 결정 신경망(204) 각각은 전연결 레이어(fully-connected layer) 및 활성화 함수(activation function)로 구현될 수 있다.

[0046] 특징구조 변환기(206)는 딥러닝 기반 변환 모델(transformation model)을 이용하여 공통 특징맵으로부터 변환 영상 x_{trans} 을 생성한다. 딥러닝 기반 변환 모델의 일 실시예는 도 4에 예시된 바와 같다. 도 4의 예시에서, $D(a,b,c)$ 는 디콘볼루션 레이어(deconvolution layer)를 나타내고, $C(a,b,c)$ 는 콘볼루션 레이어(convolution layer)를 나타낸다. 또한, a, b, 및 c는 각각 콘볼루션/디콘볼루션 필터의 사이즈, 필터의 개수, 및 스트라이드(stride)를 나타낸다. 변환 영상 x_{trans} 는 원본 영상 x와 같은 크기일 필요는 없으며, 설정된 크기에서 원본 영상의 시각 정보를 잘 표현한 저해상도의 영상일 수 있다.

[0047] 트레이닝부는 공통특징 추출기(110)를 종단 간(end-to-end)으로 트레이닝할 수 있다. 이때, 수학적 식 1에 나타낸 바와 같이, 손실함수는 대표 태스크에 포함되는 개별 태스크들의 손실함수와 인간의 시각을 위한 영상 복원 손실함수를 가중합하여 정의된다.

수학적 식 1

$$L_{common} = \sum_{i=1}^M \alpha_i L_i + \beta L_I$$

[0048]

[0049] 여기서 L_i 는 전체 M 개의 태스크 중 i 번째 태스크를 위한 손실이고, α_i 는 학습 중 i 번째 태스크에 대한 손실의 영향을 조절하는 파라미터이다. 영상 복원 손실 L_I 는 MSE(Mean Square Error) 손실, SATD(Sum of Absolute Transformed Difference) 손실 등과 같은 영상 복원에 보편적으로 사용되는 손실일 수 있다. β 는 영상 복원 손실의 영향을 조절하는 파라미터이다.

[0050] 한편, 수학적 식 1에 나타낸 손실함수에 기초하여 학습된 공통특징 추출기(110)는 개별 태스크와 영상 복원에 특화된 신경망들에 비해 모든 태스크를 포괄하는 공통적인 특징맵을 생성하므로, 개별 태스크를 위한 최상의 특징맵을 생성하지는 못한다. 하지만, 공통 특징맵은 대표 태스크를 위한 기본적인 성능을 제공할 수 있으며, 이러한 기본 성능을 바탕으로 개별 태스크에 대해 확장성을 보장할 수 있다.

[0051] 예를 들어, 기계가 영상 분류(image classification)와 객체 인식(object detection)을 동시에 수행하는 경우, 공통특징 추출기(110)의 기본 신경망(202)은 각 태스크에 대한 예측값을 결과로 내는 부분이 분기(branch)되는 구조이거나 또는 결과 로짓(logit)의 일정 부분이 개별 태스크에 대한 예측값을 나타내는 등의 구조를 가질 수 있다. 이때, 태스크 손실 함수 L_i 는, 영상 분류에 대해서는 레이블과 영상 분류를 위한 기본 신경망의 예측 y_i 간의 크로스 엔트로피(cross-entropy) 손실이고, 객체 인식에 대해서는 실제 객체의 위치와 신경망이 해당 객체의 위치로 예측한 위치 간의 회귀(regression) 손실일 수 있다. 또한, 특징구조 변환기(206) 내의 변환 모델에 의해 공통 특징맵으로부터 복원되는 변환 영상 x_{trans} 와 대응되는 원본 영상 간의 손실이, 인간의 시각을 위한 영상 복원 손실함수 L_I 이다.

[0052] 공통특징 부호화기(112)는 딥러닝 기반으로 공통 특징맵 f_c 를 부호화하여 비트스트림을 생성한다. 이하, 공통 특징맵이 부호화된 비트스트림을 제1 비트스트림으로 나타낸다.

[0053] 공통 특징맵 f_c 를 포함하여, 일반적인 딥러닝 모델의 특징맵은 (W,H,C)의 크기를 갖는다. 채널의 크기 C를 시간으로 대응시킴으로써, 공통 특징맵은 W×H(Width×Height) 크기의 프레임 C 개를 포함하는 비디오로 가정될 수 있다. 따라서, HEVC, VVC 등과 같은 기존의 비디오 코덱을 사용하여 공통특징 부호화기(112)는 공통 특징맵을 부호화할 수 있다. 또는, 딥러닝 기반 오토인코더(auto-encoder)를 사용하여 공통특징 부호화기(112)는 공통 특징맵을 부호화할 수 있다.

[0054] 딥러닝 기반 오토인코더를 이용하는 경우, 트레이닝부는, 도 5에 예시된 바와 같이, 공통특징 부호화기(1012)까지 포함하여 공통특징 부호화기(112)를 학습시킬 수 있다. 영상 복호화 장치에 포함되는 공통특징 복호화기

(1012)는 추후 설명하기로 한다. 이때, 손실함수는 수학적 식 2에 나타난 바와 같이 정의된다.

수학적 식 2

$$L_{b_c} = L_2(f_{c,raw}, f_{c,rec}) + \lambda L_1(b_c)$$

[0055]

[0056]

여기서, $L_2(\cdot)$ 는 L2 손실, $L_1(\cdot)$ 는 L1 손실을 나타낸다. 또한, 첫 번째 항은 전송된 공통 특징맵 $f_{c,raw}$ 과 복원된 공통 특징맵 $f_{c,rec}$ 간의 차이를 감소시키기 위한 손실을 나타내고, 두 번째 항은 공통 특징맵의 전송 비트수 b_c 를 감소시키기 위한 손실을 나타낸다. λ 는 두 가지 손실에 대해, 학습 중 영향을 조절하기 위한 파라미터이다.

[0057]

도 5의 예시에서, GDN(Generalized Divisive Normalization)은 비선형 영상 변환(nonlinear image transform)의 학습 과정에서 이용되는 비선형 활성화 함수(activation function)를 나타낸다. IGDN은 GDN의 역작용(inverse operation)을 수행한다. 예컨대, GDN FC 64는 활성화 함수로서 GDN을 이용하는 64 채널의 전연결 레이어를 나타내고, IGDN FC 128은 활성화 함수로서 IGDN을 이용하는 128 채널의 전연결 레이어를 나타낸다.

[0058]

오토인코더는 입력을 출력으로 복사하는 딥러닝 모델이다. 간단한 딥러닝 모델처럼 보이나, 모델에 다양한 제약(constraint)을 설정함으로써 복잡한 모델을 생성할 수 있다. 예컨대, 은닉층의 크기를 입력층(input layer)의 크기보다 작게 하여 데이터를 압축, 즉 차원을 축소할 수 있다. 또는, 입력 데이터에 잡음(noise)을 추가하여 원본 입력을 복원할 수 있도록 딥러닝 모델을 학습시킬 수 있다. 이러한 제약들은 오토인코더가 단순히 입력을 바로 출력으로 복사하는 것을 방지하고, 데이터를 효율적으로 표현(representation)하는 방법을 학습하도록 한다.

[0059]

도 5에 예시된 바와 같이, 오토인코더는 항상 부호화기(encoder)와 복호화기(decoder), 두 부분으로 구성된다. 본 개시는, 부호화기의 출력 데이터의 크기를 입력 데이터보다 작게 설정함으로써, 입력 데이터를 압축하여 비트스트림을 생성할 수 있다.

[0060]

이상의 설명에서, 공통특징 추출기(110)가 공통 특징맵으로부터 변환 영상 x_{trans} 를 생성하나, 반드시 이에 한정하는 것은 아니다. 추후 설명될 VCM 복호화 장치는, 사전에 공유된 변환 모델을 이용하여 복원된 공통 특징맵으로부터 변환 영상 x_{trans} 를 생성한다. 본 개시에 따른 다른 예에 있어서, VCM 부호화 장치는 공통특징 복호화기를 추가로 포함하여, 공통특징 부호화기(112)가 생성한 제1 비트스트림으로부터 공통 특징맵을 복호화할 수 있다. VCM 부호화 장치는, 특징구조 변환기(206) 내의 변환 모델을 이용하여 복호화된 공통 특징맵 f_{rec} 로부터 변환 영상 x_{trans} 를 생성할 수 있다. 여기서, 변환 모델은, 수학적 식 1에 나타난 바와 같은 손실함수를 기반으로 사전에 트레이닝된 공통특징 추출기(110)의 구성요소 중 하나이다.

[0061]

특징대영상 매핑기(114)는, 공통특징 추출기(110)가 생성한 변환 영상 x_{trans} 로부터 원본 영상 x 와 유사한 기본 영상 x_{base} 를 생성한다. 기본 영상은 추후 VCM 복호화 장치에서 복원된 공통 특징맵을 기반으로 생성할 수 있는 가장 최소 화질의 영상이다. 따라서, 인간의 시각을 위해 더 개선된 화질을 요구된다면 VCM 부호화 장치는 비디오 부호화기(130)를 이용하여 기본 영상보다 더 개선된 화질을 갖는 영상을 제공할 수 있다.

[0062]

특징대영상 매핑기(114)는 변환 영상 x_{trans} 를 디콘볼루션 레이어로 구성된 딥러닝 기반 영상복원(image restoration) 모델에 입력하여 기본 영상 x_{base} 를 생성한다. 여기서, 영상복원 모델은 원본 영상 x 와 같은 영상을 출력하도록 학습된 모델일 수 있다.

[0063]

한편, 영상복원 모델은, 도 6에 예시된 바와 같이, 피라미드 구조를 가질 수 있다. 영상복원 모델 내의 레이어 중 일정 간격으로 또는 임의로 설정한 레이어 인덱스 k 에서 출력된 특징맵과 그 특징맵과 동일한 크기로 업샘플링된 $x_{k,trans}$ 를 가산하여 중간 복원 영상 $x_{k,base}$ 를 생성한다. 영상복원 모델의 학습 과정에서는, $x_{k,base}$ 와 그와 동일한 크기로 다운샘플링된 원본 영상 x_k 간의 손실을 추가적으로 감소시킴으로써, 영상복원 모델의 최종단의 출력인 기본 영상 x_{base} 가 원본 영상 x 의 고주파 영역의 시각적 정보를 더 잘 복원할 수 있도록 한다.

[0064]

트레이닝부는 이러한 피라미드 구조의 모델을, 수학적 식 3에 나타난 바와 같은 손실함수를 이용하여 트레이닝시킬

수 있다.

수학식 3

$$L_{base} = \sum_{k=1}^K L_I(x_{k,base}, x_k)$$

[0065]

[0066]

여기서, 복원 손실 $L_I(\cdot)$ 는 MSE 손실, SATD 손실 등과 같은 영상 복원에 보편적으로 사용되는 손실일 수 있다. K (여기서, K 는 자연수)는 영상복원 모델에 설정된 피라미드의 개수이다. 피라미드 구조를 사용하지 않고 최종 단의 출력과 원본 영상 간의 손실만을 고려하는 영상복원 모델은, $K=1$ 이다. $K>1$ 인 경우, $x_{k,base}$ 는 k 번째 피라미드에서 출력되는 영상이고 x_k 은 그와 동일한 크기로 다운샘플링된 원본 영상이다.

[0067]

한편, VCM 부호화 장치는, 수행하는 태스크가 대표 태스크이고 그 태스크에 대한 더 개선된 성능이 필요한 경우, 태스크 특징 추출기(120)를 이용하여 개별 태스크의 태스크 특화 특징맵을 추출한다. 여기서, 대표 태스크에 대한 더 개선된 성능이 필요한 경우는, 예컨대, M 개의 결정 신경망(204)에 의해 생성된 분석 결과의 누적 신뢰도가 기설정된 임계치 미만이어서, 대표 태스크에 포함되는 개별 태스크별 분석 결과가 만족스럽지 않은 경우이다. 이때, 개별 태스크별 분석 결과에 따라, VCM 부호화 장치는 M 개의 개별 태스크 전체 또는 일부에 대해 태스크 특징 추출기(120)를 적용할 수 있다.

[0068]

잔여 태스크가 존재하여, VCM 부호화 장치가 수행하는 태스크가 잔여 태스크인 경우, 태스크 특징 추출기(120)를 이용하여 잔여 태스크의 태스크 특화 특징맵을 추출한다.

[0069]

VCM 부호화 장치는 M 개의 개별 태스크를 취급하기 위해 M 개의 태스크 특징 추출기(120)를 포함할 수 있다. VCM 부호화 장치는 잔여 태스크를 수행하기 위해 $N-M$ 개의 태스크 특징 추출기(120)를 포함할 수 있다. 각 태스크 특징 추출기(120)에 포함된 구성요소들의 파라미터는 독립적으로 존재한다.

[0070]

태스크 특징 추출기(120)는 딥러닝 기반으로, 입력 영상으로부터 대표 태스크에 포함되는 개별 태스크, 또는 잔여 태스크의 태스크 특화 특징맵을 생성하고, 개별 태스크 또는 잔여 태스크의 분석 결과를 생성한다.

[0071]

도 7은 본 개시의 일 실시예에 따른 태스크 특징 추출기(120)를 개념적으로 나타낸 블록도이다.

[0072]

태스크 특징 추출기(120)는 태스크 신경망(task neural network, 702), 및 결정 신경망(704)을 포함한다.

[0073]

태스크 신경망(702)은, 개별 태스크 또는 잔여 태스크의 학습이 수행된 딥러닝 모델이다.

[0074]

결정 신경망(704)은 개별 태스크 또는 잔여 태스크에 대한 분석 결과 y 를 생성한다. 결정 신경망(704)은 전연결 레이어 및 활성화 함수로 구현될 수 있다.

[0075]

트레이닝부는 태스크 특징 추출기(120)를 트레이닝시키기 위해, 수학식 4에 나타낸 바와 같은 손실함수를 이용할 수 있다.

수학식 4

$$L_{task} = L_T(y) + \lambda L_1(f_t)$$

[0076]

[0077]

여기서, 첫 번째 항의 $L_T(\cdot)$ 는 태스크에 대해 보편적으로 적용하는 손실이고, 두 번째 항의 $L_1(\cdot)$ 은 태스크 특화 특징맵의 전송 비트수를 감소시키기 위한 손실을 나타낸다. λ 는 두 가지 손실에 대해, 학습 중 영향을 조절하기 위한 파라미터이다.

[0078]

VCM 부호화 장치는 M 개의 개별 태스크를 수행하기 위해 M 개의 태스크 특징 부호화기(122)를 포함할 수 있다. VCM 부호화 장치는 잔여 태스크를 수행하기 위해 $N-M$ 개의 태스크 특징 부호화기(122)를 포함할 수 있다. 각 태스크 특징 부호화기(122)가 딥러닝 기반으로 구현되는 경우, 각 태스크 특징 부호화기(122)에 포함된 구성요소

들의 파라미터는 독립적으로 존재한다

[0079] 태스크 특징 부호화기(122)는 딥러닝 기반으로 태스크 특화 특징맵을 부호화하여 비트스트림을 생성한다. 이하, 개별 태스크의 태스크 특화 특징맵이 부호화된 비트스트림을 제2 비트스트림으로 나타낸다. 또한, 잔여 태스크의 태스크 특화 특징맵이 부호화된 비트스트림을 제3 비트스트림으로 나타낸다.

[0080] 공통 특징맵과 유사하게, 태스크 특화 특징맵도 (W,H,C)의 크기를 가지므로, 태스크 특화 특징맵은 $W \times H$ 크기의 프레임 C 개를 포함하는 비디오로 가정될 수 있다. 따라서 HEVC, VVC 등과 같은 기존의 비디오 코덱을 사용하여 태스크 특징 부호화기(122)는 태스크 특화 특징맵을 부호화할 수 있다. 또는, 딥러닝 기반 오토인코더를 사용하여 태스크 특징 부호화기(122)는 태스크 특화 특징맵을 부호화할 수 있다.

[0081] 딥러닝 기반 오토인코더를 이용하는 경우, 트레이닝부는, 도 8에 예시된 바와 같이, 태스크 특징 복호화기(1022)까지 포함하여 태스크 특징 부호화기(122)를 트레이닝시킬 수 있다. 영상 복호화 장치에 포함되는 태스크 특징 복호화기(1022)는 추후 설명하기로 한다. 이때, 손실함수는 수학적식 5에 나타난 바와 같이 정의된다.

수학적식 5

$$L_{b_t} = L_2(f_{t,raw}, f_{t,rec}) + \lambda L_1(b_t)$$

[0083] 여기서, $L_2(\cdot)$ 는 L2 손실, $L_1(\cdot)$ 는 L1 손실을 나타낸다. 또한, 첫 번째 항은 전송된 태스크 특화 특징맵 $f_{t,raw}$ 과 복원된 태스크 특화 특징맵 $f_{t,rec}$ 간의 차이를 감소시키기 위한 손실을 나타내고, 두 번째 항은 태스크 특화 특징맵의 전송 비트수 b_t 를 감소시키기 위한 손실을 나타낸다. λ 는 두 가지 손실에 대해, 학습 중 영향을 조절하기 위한 파라미터이다.

[0084] 인간의 시각을 만족시키기 위해 특징대영상 매핑기(114)가 생성하는 기본 영상보다 더 개선된 영상이 필요한 경우, VCM 부호화 장치는 비디오 부호화기(130)를 이용하여 개선된 복원 영상(reconstructed image)의 생성에 필요한 잔차 영상(residual image)을 부호화하여 비트스트림을 생성할 수 있다. 이하, 잔차 영상이 부호화된 비트스트림을 제4 비트스트림으로 명칭한다.

[0085] 잔차 영상은 입력 영상으로부터 기본 영상을 감산하여 생성되는 텍스처(texture)이다. 따라서, 비디오 부호화기(130)는 텍스처(texture) 부호화기로도 명칭될 수 있다.

[0086] 비디오 부호화기(130)는, HEVC, VVC 등과 같은 기존의 비디오 코덱을 이용하여 구현될 수 있다. 또는, 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.

[0087] 딥러닝 기반 오토인코더를 이용하는 경우, 트레이닝부는, 도 9에 예시된 바와 같이, 비디오 복호화기(1030)까지 포함하여 비디오 부호화기(130)를 학습시킬 수 있다. 영상 복호화 장치에 포함되는 비디오 복호화기(1030)는 추후 설명하기로 한다. 이때, 손실함수는 수학적식 6에 나타난 바와 같이 정의된다.

수학적식 6

$$L_{b_{res}} = L_I(x_{res,raw}, x_{res,rec}) + \lambda L_1(b_{res})$$

[0089] 여기서, 복원 손실 $L_I(\cdot)$ 는 MSE 손실, SATD 손실 등과 같은 영상 복원에 보편적으로 사용되는 손실, 또는 L2 손실일 수 있다. $L_1(\cdot)$ 는 L1 손실을 나타낸다. 또한, 첫 번째 항은 잔차 영상과 복원된 잔차 영상 간의 차이를 감소시키기 위한 손실을 나타내고, 두 번째 항은 잔차 영상의 전송 비트수 b_{res} 를 감소시키기 위한 손실을 나타낸다. λ 는 두 가지 손실에 대해, 학습 중 영향을 조절하기 위한 파라미터이다.

[0090] 다중화기(140)는, 공통특징 부호화기(112)에 생성되는 제1 비트스트림, N 개의 태스크 특징 부호화기(122)에 생성되는 제2 비트스트림과 제3 비트스트림, 및 비디오 부호화기(130)에 의해 생성되는 제4 비트스트림의 전부 또는 일부를 다중화하여 다중화 비트스트림을 생성한 후, VCM 복호화 장치로 전송한다.

- [0091] 이때, VCM 부호화 장치는 제2 비트스트림, 제3 비트스트림 및 제4 비트스트림 각각이 존재함을 나타내는 플래그들을 VCM 복호화 장치로 전송할 수 있다.
- [0092] 신경망 연결부(150)는, VCM 부호화 장치에서 이용하는 딥러닝 모델들의 정보(예컨대, 파라미터)를 저장하는 모듈이다. 이 모듈은 트레이닝부에 의해 트레이닝된 딥러닝 모델들의 파라미터를 저장하되, VCM 부호화 장치의 구성요소일 필요는 없다.
- [0093] 도 10은 본 개시에 일 실시예에 따른 VCM 복호화 장치를 개념적으로 나타내는 예시도이다.
- [0094] VCM 복호화 장치는 다중화 비트스트림을 획득하여, 공통 특징맵, 개별 태스크와 잔여 태스크의 태스크 특화 특징맵, 및 잔차 영상에 해당하는 비트스트림을 획득한다. VCM 복호화 장치는 비트스트림으로부터 공통 특징맵을 복호화한다. VCM 복호화 장치는 필요 시, 대표 태스크에 포함되는 개별 태스크들의 태스크 특화 특징맵을 복호화한다. VCM 복호화 장치는, 잔여 태스크들의 태스크 특화 특징맵을 복호화한다. VCM 복호화 장치는 필요 시, 복원된 공통 특징맵을 이용하여 기본 영상을 생성하고, 비트스트림으로부터 잔차 영상을 복호화한 후, 기본 영상을 가산하여 비디오 복원 영상을 생성한다. VCM 복호화 장치는, 도 10에 예시된 바와 같이, 공통특징 복호화기(1012), 특징대영상 매핑기(1014), N 개의 태스크 특징 복호화기(1022), 비디오 복호화기(1030), 역다중화기(demultiplexer, 1040), 및 신경망 연결부(1050)의 전부 또는 일부를 포함한다.
- [0095] 역다중화기(1040)는 다중화 비트스트림으로부터 공통특징 복호화기(1012)에 의해 이용되는 제1 비트스트림, N 개의 태스크 특징 복호화기(1022)에 이용되는 제2 비트스트림과 제3 비트스트림, 및 비디오 복호화기(1030)에 의해 이용되는 제4 비트스트림을 역다중화한다.
- [0096] 이때, VCM 복호화 장치는 제2 비트스트림, 제3 비트스트림 및 제4 비트스트림 각각이 존재함을 나타내는 플래그들을 이용하여 다중화 비트스트림을 역다중화할 수 있다.
- [0097] 공통특징 복호화기(1012)는 제1 비트스트림으로부터 공통 특징맵을 복호화한다. 공통특징 복호화기(1012)는 기존의 비디오 코덱을 사용하여 공통 특징맵을 복호화할 수 있다. 또는, 딥러닝 기반 오토인코더를 사용하여 공통특징 복호화기(1012)는 공통 특징맵을 복호화할 수 있다.
- [0098] 딥러닝 기반 오토인코더를 이용하는 경우, 도 5에 예시된 바와 같이, 공통특징 부호화기(112)까지 포함하여 공통특징 복호화기(1012)가 사전에 트레이닝될 수 있다. 손실함수는, 전술한 바와 같은, 수학적 2에 나타난 바와 같이 정의되므로, 자세한 설명은 생략한다.
- [0099] 특징대영상 매핑기(1014)는, 인간의 시각을 위해 복원된 영상이 필요한 경우, 복호화된 공동 특징맵 f_{rec} 로부터 변환 영상 x_{trans} 을 생성한 후, 변환 영상 x_{trans} 으로부터 원본 영상 x 와 유사한 기본 영상 x_{base} 을 생성한다. 기본 영상은 VCM 복호화 장치에서 제공할 수 있는 가장 최소 화질의 영상이다.
- [0100] VCM 복호화 장치는, 도 4에 예시된 바와 같은, 특징구조 변환기(206)를 포함하여, 특징구조 변환기(206) 내의 딥러닝 기반 변환 모델을 이용하여 복원된 공동 특징맵으로부터 변환 영상 x_{trans} 을 생성할 수 있다. VCM 복호화 장치에서 특징구조 변환기(206)는 특징대영상 매핑기(1014)의 일부로 포함될 수도 있다.
- [0101] 특징대영상 매핑기(1014)는 딥러닝 기반 영상복원 모델을 이용하여 변환 영상 x_{trans} 으로부터 원본 영상 x 와 유사한 기본 영상 x_{base} 을 생성한다. 한편, 영상복원 모델 모델은, 도 6에 예시된 바와 같이, 피라미드 구조를 가질 수 있다. 또한, 피라미드 구조의 모델은, 수학적 3에 나타내 바와 같은 손실함수를 이용하여 사전에 트레이닝된다.
- [0102] 한편, VCM 복호화 장치는, 취급하는 태스크가 대표 태스크이고 그에 대한 더 개선된 성능이 필요하여 개별 태스크의 태스크 특화 특징맵이 전송된 경우, 태스크 특징 복호화기(1022)를 이용하여 제2 비트스트림으로부터 개별 태스크의 태스크 특화 특징맵을 복호화한다. 이때, 개별 태스크에 대한 분석 결과에 따라, M 개의 개별 태스크 전체 또는 일부에 대해 태스크 특징 복호화기(1022)를 적용할 수 있다.
- [0103] 잔여 태스크의 태스크 특화 특징맵이 전송된 경우, VCM 복호화 장치는 태스크 특징 복호화기(1022)를 이용하여 제3 비트스트림으로부터 잔여 태스크의 태스크 특화 특징맵을 복호화한다.
- [0104] VCM 복호화 장치는 M 개의 개별 태스크를 수행하기 위해 M 개의 태스크 특징 복호화기(1022)를 포함할 수 있다. VCM 복호화 장치는 잔여 태스크를 수행하기 위해 N-M 개의 태스크 특징 복호화기(1022)를 포함할 수 있다. 각 태스크 특징 복호화기(1022)가 딥러닝 기반으로 구현되는 경우, 각 태스크 특징 복호화기(1022)에 포함된 구성

요소들의 파라미터는 독립적으로 존재한다.

- [0105] 태스크 특징 복호화기(1022)는 기존의 비디오 코덱을 사용하여 태스크 특화 특징맵을 복호화할 수 있다. 또는, 딥러닝 기반 오토인코더를 사용하여 태스크 특화 특징맵을 복호화할 수 있다.
- [0106] 딥러닝 기반 오토인코더를 이용하는 경우, 도 8에 예시된 바와 같이, 태스크 특징 부호화기(122)까지 포함하여 태스크 특징 복호화기(1022)가 사전에 트레이닝될 수 있다. 손실함수는, 전술한 바와 같은, 수학적 식 5에 나타낸 바와 같이 정의되므로, 자세한 설명은 생략한다.
- [0107] 한편, 도 10에 점선 박스로 예시된 머신 비전 부분에서는, 복호화된 공통 특징맵 및 태스크 특화 특징맵을 이용하여, 대표 태스크에 포함된 개별 태스크, 또는 잔여 태스크에 대한 분석이 수행될 수 있다.
- [0108] 인간의 시각을 만족시키기 위한 더 개선된 영상을 제공하기 위해 잔차 영상이 전송된 경우, VCM 복호화 장치는 비디오 복호화기(1030)를 이용하여 제4 비트스트림으로부터 잔차 영상을 복호화할 수 있다. 추가적으로, 비디오 복호화기(1030)는 잔차 영상과 기본 영상을 가산하여 복원 영상을 생성할 수 있다. 예컨대, 도 10에 점선 박스로 예시된 인간 시각 부분에서는, 필요에 따라 기본 영상 또는 복원 영상이 선택적으로 이용될 수 있다.
- [0109] 비디오 부호화기(130)가 텍스처 부호화기로 명칭되듯이, 비디오 복호화기(1030)도 텍스처 복호화기로 명칭될 수 있다.
- [0110] 비디오 복호화기(1030)는 전술한 바와 같이, 기존의 비디오 코덱을 사용하여 잔차 영상을 복호화할 수 있다. 또는, 딥러닝 기반 오토인코더를 사용하여 잔차 영상을 복호화할 수 있다.
- [0111] 딥러닝 기반 오토인코더를 이용하는 경우, 도 8에 예시된 바와 같이, 비디오 부호화기(130)까지 포함하여 비디오 복호화기(1030)가 사전에 트레이닝될 수 있다. 손실함수는, 전술한 바와 같은, 수학적 식 6에 나타낸 바와 같이 정의되므로, 자세한 설명은 생략한다.
- [0112] 신경망 연결부(1050)는 VCM 복호화 장치에서 이용하는 딥러닝 모델들의 정보(예컨대, 파라미터)를 저장하는 모듈이다. 이 모듈은 트레이닝부에 의해 트레이닝된 딥러닝 모델들의 파라미터를 저장하되, VCM 복호화 장치의 구성요소일 필요는 없다.
- [0113] 도 1 및 도 10의 도시는 본 실시예에 따른 예시적인 구성의 하나이며, VCM 코딩 장치, 즉, 다중 태스크 시스템이 수행하는 태스크, 및 머신 비전과 인간의 시각 측면에서 기계 및 사용자의 요구 성능 수준에 따라 다른 구성요소 또는 구성요소 간의 다른 연결을 포함함으로써, 그 구성이 유동적으로 변경될 수 있다.
- [0114] 또한, 초기에 설정된 다중 태스크 시스템의 구조에서 다중 태스크 시스템이 수행하는 태스크가 추가 또는 삭제되거나 머신 비전과 인간의 시각 측면에서 기계 및 사용자의 요구 성능 수준이 달라질 경우, 구성요소가 확장적으로 추가되거나 삭제될 수 있다
- [0115] 도 1 및 도 10의 도시에서는 VCM 코딩 장치라는 측면에서 다중 태스크 시스템이 예시되었으나, 대안적으로 다중 태스크 시스템은 계층 구조로 기술될 수 있다. 다중 태스크 시스템은 태스크 측면에서 대표 태스크를 수행하는 공통 특징 계층(common feature layer), 개별 태스크 또는 잔여 태스크를 수행하는 태스크 특화 특징 계층(task-specific feature layer), 및 영상을 처리하는 영상 복원 계층(image reconstruction layer)를 포함한다.
- [0116] 공통 특징 계층은 입력 영상으로부터 대표 태스크의 공통 특징맵을 추출하고, 이를 부호화 및 복호화하는 계층이다. 공통 특징 계층은 공통특징 추출기(110), 공통특징 부호화기(112), 공통특징 복호화기(1012), 및 특징대 영상 매핑기(114, 1014)를 포함한다. 공통 특징 계층의 각 구성요소의 동작은 이미 기술되었으므로 더 이상의 설명은 생략한다.
- [0117] 공통 특징 계층은 다중 태스크 시스템에서 우선적이고 반드시 설정되어 수행되는 계층으로서, 공통 특징맵을 이용하여 대표 태스크들에 대한 최소한의 성능을 보장하고 기본 영상을 이용하여 인간의 시각 측면에서의 최소 화질을 보장한다. 다른 두 계층은, 부호화기에 포함된 구성요소들이 미리 생성된 상태에서 기계와 사용자가 필요로 할 때에만 선택적으로 그 계층과 관련된 정보를 압축 및 전송한다.
- [0118] 태스크 특화 특징 계층은 입력 영상으로부터 개별 태스크 및 잔여 태스크의 태스크 특화 특징맵을 추출하고, 이를 부호화 및 복호화하는 계층이다. 태스크 특화 특징 계층은 태스크 특징 추출기(120), 태스크 특징 부호화기(122), 및 태스크 특징 복호화기(1022)를 포함한다. 태스크 특화 특징 계층의 각 구성요소의 동작은 이미 기술되었으므로 더 이상의 설명은 생략한다.

- [0119] 태스크 특화 특징 계층은, 기계가 대표 태스크에 대하여 보장된 최소 성능보다 개선된 성능을 필요로 하거나 잔여 태스크에 대한 분석이 필요할 때 정보를 전송한다.
- [0120] 영상 복원 계층은 공통 특징맵을 기반으로 입력 영상의 잔차 영상으로부터 복원 영상을 생성한다. 영상 복원 계층은 비디오 부호화기(130) 및 비디오 복호화기(1030)를 포함한다. 영상 복원 계층의 각 구성요소의 동작은 이미 기술되었으므로 더 이상의 설명은 생략한다.
- [0121] 영상 복원 계층은, 사용자가 기본 영상이 제공하는 최소 화질보다 개선된 화질의 복원 영상을 요구할 때 정보를 전송한다.
- [0122] 이상의 설명에서, 하나의 대표 태스크 및 잔여 태스크를 가정하여, 다중 태스크 시스템, 즉, VCM 코딩 장치가 기술되었다. 본 개시에 따른 다른 실시예에 있어서, 하나의 대표 태스크가 주태스크(main task) 및 부태스크(sub-task)를 포함하는 경우, 다중 태스크 시스템은 주태스크 및 부태스크를 수행하도록 변형될 수 있다.
- [0123] 예컨대, 대표 태스크 집합의 구성 태스크 중 나머지 태스크와 가장 밀접한 정보를 공유하고 나머지 태스크들을 그의 부태스크로 설정해도 될만큼 특징적인 태스크가 존재하는 경우, 이러한 특징적인 태스크를 주태스크로 정의하고 나머지 태스크들을 부태스크들로 정의한다. 이때, 잔여 태스크는 존재하지 않는 것으로 설정한다.
- [0124] 이하, 도 11 및 도 12의 예시를 이용하여, 하나의 주태스크와 N 개의 부태스크를 수행하기 위한 VCM 부호화 장치 및 VCM 복호화 장치를 설명한다.
- [0125] 도 11은 본 개시의 다른 실시예에 따른 VCM 부호화 장치를 개념적으로 나타내는 블록도이다.
- [0126] 도 11에 예시된 VCM 부호화 장치는 주태스크의 수행을 위한 구성요소로서 주태스크 특징 추출기(1110) 및 주태스크 특징 부호화기(1112)를 포함하고, 부태스크의 수행을 위한 구성요소로서 N 개의 부태스크 특징 추출기(1120) 및 N 개의 부태스크 특징 부호화기(1122)를 포함한다. VCM 부호화 장치의 나머지 구성요소는 도 1의 예시와 동일하다.
- [0127] 도 12는 본 개시의 다른 실시예에 따른 VCM 복호화 장치를 개념적으로 나타내는 블록도이다.
- [0128] 도 12에 예시된 VCM 복호화 장치는 주태스크의 수행을 위해 주태스크 특징 복호화기(1212)를 포함하고, 부태스크 수행을 위해 N 개의 부태스크 특징 복호화기(1222)를 포함한다. VCM 복호화 장치의 나머지 구성요소는 도 10의 예시와 동일하다.
- [0129] 도 11 및 도 12의 예시에서 VCM 부호화 장치 및 VCM 복호화 장치가 주태스크를 수행하는 것은, 도 1 및 도 10에 예시된 바와 같이, 공통 특징 계층에서 대표 태스크를 수행하는 것과 유사하다. 따라서, VCM 부호화 장치 및 VCM 복호화 장치는, 특징대영상 매핑기(114, 1014)를 이용하여, 주태스크 특징 복호화기(1212)가 생성하는 주태스크 특화 특징맵으로부터 기본 영상을 생성할 수 있다.
- [0130] 도 11 및 도 12의 예시에서 VCM 부호화 장치 및 VCM 복호화 장치가 부태스크를 수행하는 것은, 도 1 및 도 10에 예시된 바와 같이, 태스크 특화 특징 계층에서 개별 태스크 또는 잔여 태스크를 수행하는 것과 유사하다.
- [0131] 도 11 및 도 12의 예시에서, 주태스크를 수행하는 구성요소와 부태스크를 수행하는 구성요소는 동일한 구조를 가질 수 있다. 하지만, 부태스크 특징 부호화기(1122)는 주태스크 특화 특징맵을 참조 프레임으로 이용하여, 부태스크 특화 특징맵의 잔차 프레임을 생성한 후, 잔차 프레임을 전송할 수 있다.
- [0132] 한편, 전술한 바와 같이, 태스크 간 유사도에 따라 태스크를 분류 시, 구성 태스크의 수에 큰 차이가 없는 복수의 대표 태스크 집합이 설정될 수 있다. 본 개시에 따른 또다른 실시예에 있어서, 다중 태스크 시스템은 복수의 대표 태스크 부그룹(subgroup)을 이용하여 복수의 대표 태스크를 수행할 수 있다. 이때, 각 부그룹은 공통 특징맵 및 태스크 특화 특징맵을 처리하는 구성요소들, 즉, 공통 특징 계층 및 태스크 특화 특징 계층을 포함한 채로 상호 독립적으로 동작하고, 부그룹 간에 공유되는 정보는 존재하지 않는다.
- [0133] 도 13은 본 개시의 또다른 실시예에 따른 VCM 코덱을 개념적으로 나타내는 블록도이다.
- [0134] 도 13에 예시된 바는, VCM 부호화 코덱이 2 개의 대표 태스크를 수행하는 예시를 나타낸다. 여기서, VCM 부호화 장치 및 VCM 복호화 장치 각각의 대표 태스크 부그룹(1302)은, 도 1 및 도 10에 예시된 바와 같은, 공통 특징맵 및 N 개의 태스크 특화 특징맵을 처리하는 구성요소를 모두 포함한다.
- [0135] 도 13의 예시에서, VCM 코덱은 제1 대표 태스크 부그룹이 생성하는 공통 특징맵을 이용하여 영상 복원을 수행하나, 반드시 이에 한정하지 않는다. VCM 코덱은 영상 복원을 위해, 부호화 장치 내에 포함된 부그룹이 생성한

어느 공통 특징맵이든 이용할 수 있다. 또한, VCM 코덱은, 부호화 장치 내에 포함된 부그룹들의 전부 또는 일부에 의해 생성되는 공통 특징맵들을 이용하여 영상 복원을 수행할 수 있다.

- [0136] 도 14는 본 개시의 일 실시예에 따른 VCM 부호화 방법을 나타내는 순서도이다.
- [0137] VCM 부호화 장치는 원본 영상을 획득한다(S1400).
- [0138] VCM 부호화 장치는 딥러닝 기반 공통특징 추출 모델을 이용하여 원본 영상으로부터 공통 특징맵을 추출한다(S1402). 여기서, 공통 특징맵은 원본 영상이 함축하는 대표 태스크를 표현한다. 전술한 공통특징 추출기(110)는 딥러닝 기반 공통특징 추출 모델을 나타낸다.
- [0139] 공통특징 추출기(110)는 기본 신경망(202), 결정 신경망(204), 및 특징구조 변환기(206)에 해당하는 변환 모델을 포함한다. 공통특징 추출기(110)는 기본 신경망을 이용하여 원본 영상으로부터 공통 특징맵을 추출하되, 결정 신경망을 이용하여 공통 특징맵을 기반으로 대표 태스크의 분석 결과를 생성하고, 변환 모델을 이용하여 공통 특징맵으로부터 변환 영상을 생성한다.
- [0140] VCM 부호화 장치는 공통특징 부호화기(112)를 이용하여 공통 특징맵을 부호화하여 제1 비트스트림을 생성한다(S1404).
- [0141] VCM 부호화 장치는 결정 신경망(204)을 이용하여 공통 특징맵을 기반으로 대표 태스크의 분석 결과를 생성한다(S1406).
- [0142] VCM 부호화 장치는 분석 결과의 누적 신뢰도가 기설정된 임계치 미만인지를 확인한다(S1408).
- [0143] 분석 결과의 누적 신뢰도가 기설정된 임계치 미만인 경우, VCM 부호화 장치는 제2 비트스트림을 생성할 수 있다.
- [0144] VCM 부호화 장치는 딥러닝 기반 태스크 특징 추출 모델을 이용하여 원본 영상으로부터 적어도 하나의 개별 태스크를 표현하는 태스크 특화 특징맵을 추출한다(S1410). 여기서, 적어도 하나의 개별 태스크는 대표 태스크에 포함된다. 전술한 태스크 특징 추출기(120)는 딥러닝 기반 태스크 특징 추출 모델을 나타낸다.
- [0145] VCM 부호화 장치는 태스크 특징 부호화기(122)를 이용하여 개별 태스크를 표현하는 태스크 특화 특징맵을 부호화하여 제2 비트스트림을 생성한다(S1412). 태스크 특징 부호화기는, 비디오 신호용 부호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0146] 분석 결과의 누적 신뢰도가 기설정된 임계치 이상인 경우, VCM 부호화 장치는 적어도 하나의 잔여 태스크가 존재하는지를 확인한다(S1414).
- [0147] 적어도 하나의 잔여 태스크가 존재하는 경우, VCM 부호화 장치는 제3 비트스트림을 생성할 수 있다.
- [0148] VCM 부호화 장치는 태스크 특징 추출기(120)를 이용하여 원본 영상으로부터 잔여 태스크를 표현하는 태스크 특화 특징맵을 추출한다(S1416).
- [0149] 태스크 특징 추출기(120)는 태스크 신경망(702) 및 결정 신경망(704)를 포함한다. 태스크 특징 추출기(120)는 태스크 신경망(702)을 이용하여 원본 영상으로부터 태스크 특화 특징맵을 추출하되, 결정 신경망(704)을 이용하여 태스크 특화 특징맵을 기반으로 개별 태스크 또는 잔여 태스크의 분석 결과를 생성할 수 있다.
- [0150] VCM 부호화 장치는 태스크 특징 부호화기(122)를 이용하여 잔여 태스크를 표현하는 태스크 특화 특징맵을 부호화하여 제2 비트스트림을 생성한다(S1418).
- [0151] 잔여 태스크가 존재하지 않으나, 인간의 시각 측면에서 개선된 영상이 필요한 경우를 고려하여, VCM 부호화 장치는 제4 비트스트림을 생성할 수 있다.
- [0152] VCM 부호화 장치는 공통특징 복호화기(1012)를 이용하여 제1 비트스트림으로부터 복원 공통 특징맵을 복호화한다(S1420).
- [0153] 여기서, 전술한 공통특징 부호화기 및 공통특징 복호화기는, 비디오 신호용 코덱(codec) 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0154] VCM 부호화 장치는 딥러닝 기반 영상복원 모델을 이용하여 복원 공통 특징맵으로부터 기본 영상을 생성한다(S1422). 즉, VCM 부호화 장치는 변환 모델을 이용하여 복원 공통 특징맵으로부터 변환 영상을 생성한 후, 영상 복원 모델을 이용하여 변환 영상으로부터 기본 영상을 생성할 수 있다. 전술한 특징대영상 매핑기(114)는 딥러

닝 기반 영상복원 모델을 나타낸다.

- [0155] VCM 부호화 장치는 비디오 부호화기(130)를 이용하여 원본 영상에서 기본 영상을 감산하여 잔차 영상을 생성한 후, 잔차 영상을 부호화하여 제4 비트스트림을 생성한다(S1424).
- [0156] 비디오 부호화기(130)는 비디오 신호용 부호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0157] VCM 부호화 장치는 제1 비트스트림, 제2 비트스트림, 제3 비트스트림, 제4 비트스트림의 일부를 결합하여 다중화 비트스트림을 생성한다(S1426). VCM 부호화 장치는 다중화 비트스트림을 VCM 복호화 장치로 전송한다. 이때, VCM 부호화 장치는 제2 비트스트림, 제3 비트스트림 및 제4 비트스트림 각각이 존재함을 나타내는 플래그들을 VCM 복호화 장치로 전송할 수 있다.
- [0158] 도 15는 본 개시의 일 실시예에 따른 VCM 복호화 방법을 나타내는 순서도이다.
- [0159] VCM 복호화 장치는, VCM 부호화 장치로부터 전송된 다중화 비트스트림을 획득한다(S1500). 이때, VCM 복호화 장치는 제2 비트스트림, 제3 비트스트림 및 제4 비트스트림 각각이 존재함을 나타내는 플래그들을 이용하여 다중화 비트스트림을 역다중화할 수 있다.
- [0160] VCM 복호화 장치는 다중화 비트스트림으로부터 제1 비트스트림을 획득한다(S1502). 여기서, 제1 비트스트림은, 대표 태스크를 표현하는 공통 특징맵이 부호화된 비트스트림이다.
- [0161] VCM 복호화 장치는 공통특징 복호화기(1012)를 이용하여 제1 비트스트림으로부터 공통 특징맵을 복호화한다(S1504). 여기서, 공통특징 복호화기(1012)는 기존의 비디오 신호용 복호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0162] VCM 부호화 장치가 제2 비트스트림을 전송한 경우, 다중화 비트스트림은 제2 비트스트림을 포함할 수 있다. 여기서, 제2 비트스트림은 대표 태스크에 포함되는 적어도 하나의 개별 태스크를 표현하는 태스크 특화 특징맵이 부호화된 비트스트림이다.
- [0163] VCM 복호화 장치는 다중화 비트스트림에 제2 비트스트림이 존재하는지를 확인한다(S1506).
- [0164] 제2 비트스트림이 존재하는 경우, VCM 복호화 장치는 다중화 비트스트림으로부터 제2 비트스트림을 획득한다(S1508).
- [0165] VCM 복호화 장치는 태스크 특징 복호화기(1022)를 이용하여 제2 비트스트림으로부터 개별 태스크를 표현하는 태스크 특화 특징맵을 복호화한다(S1510). 여기서, 태스크 특징 복호화기(1022)는 기존의 비디오 신호용 복호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0166] 제2 비트스트림이 존재하는 않으나, VCM 부호화 장치가 제3 비트스트림을 전송한 경우, 다중화 비트스트림은 제3 비트스트림을 포함할 수 있다. 여기서, 제3 비트스트림은 적어도 하나의 잔여 태스크를 표현하는 태스크 특화 특징맵이 부호화된 비트스트림이다.
- [0167] VCM 복호화 장치는 다중화 비트스트림에 제3 비트스트림이 존재하는지를 확인한다(S1512).
- [0168] 제3 비트스트림이 존재하는 경우, VCM 복호화 장치는 다중화 비트스트림으로부터 제3 비트스트림을 획득한다(S1514).
- [0169] VCM 복호화 장치는 태스크 특징 복호화기(1022)를 이용하여 제3 비트스트림으로부터 잔여 태스크를 표현하는 태스크 특화 특징맵을 복호화한다(S1516).
- [0170] 제3 비트스트림이 존재하는 않으나, 인간의 시각 측면에서 영상이 필요한 경우, VCM 복호화 장치는 딥러닝 기반 영상복원 모델을 이용하여 공통 특징맵으로부터 기본 영상을 생성한다(S1518). 즉, VCM 복호화 장치는 딥러닝 기반 변환 모델을 이용하여 공통 특징맵으로부터 변환 영상을 생성한 후, 영상복원 모델을 이용하여 변환 영상으로부터 기본 영상을 생성할 수 있다. 전술한 특징대영상 매핑기(114)는 딥러닝 기반 영상복원 모델을 나타낸다.
- [0171] 인간의 시각 측면에서 기본 영상보다 더 개선된 영상을 제공하기 위해, VCM 부호화 장치가 제4 비트스트림을 전송한 경우, 다중화 비트스트림은 제4 비트스트림을 포함할 수 있다. 여기서, 제4 비트스트림은 원본 영상에서 기본 영상을 감산하여 생성된 잔차 영상이 부호화된 비트스트림이다.
- [0172] VCM 복호화 장치는 다중화 비트스트림에 제4 비트스트림이 존재하는지를 확인한다(S1520).

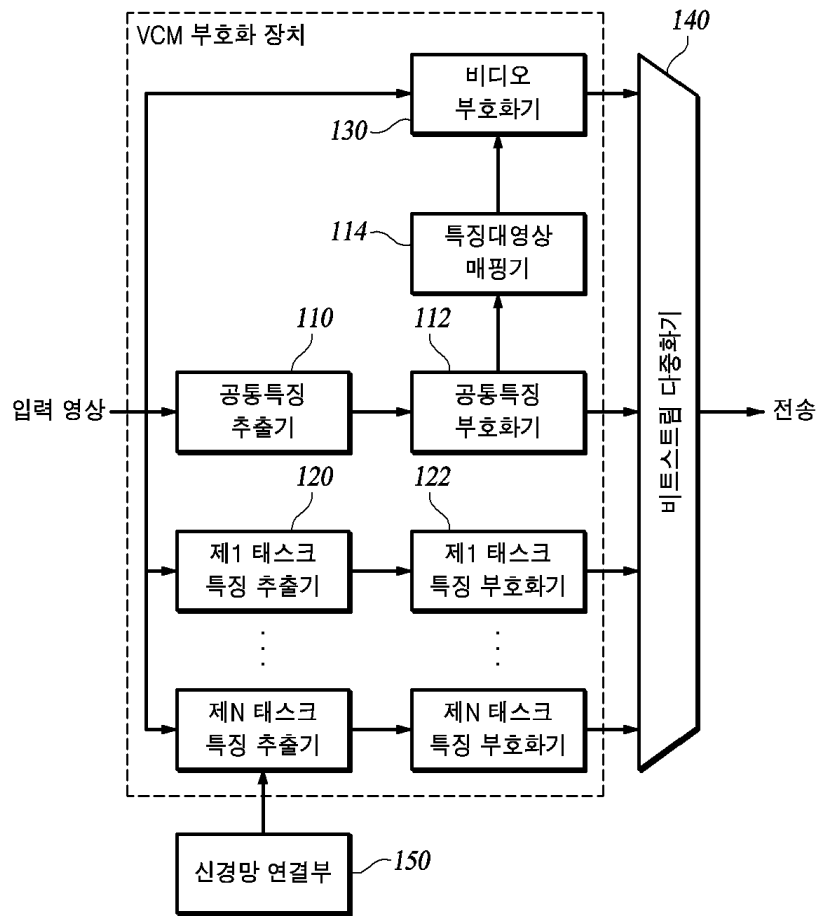
- [0173] VCM 복호화 장치는 다중화 비트스트림으로부터 제4 비트스트림을 획득한다(S1522).
- [0174] VCM 복호화 장치는 비디오 복호화기(1030)를 이용하여 제4 비트스트림으로부터 잔차 영상을 복호화한 후, 잔차 영상과 기본 영상을 가산하여 복원 영상을 생성한다(S1524). 여기서, 비디오 복호화기(1030)를 기존의 비디오 신호용 복호화기 또는 딥러닝 기반 오토인코더를 이용하여 구현될 수 있다.
- [0175] 본 실시예에 따른 각 순서도에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 순서도에 기재된 과정을 변경하여 실행하거나 하나 이상의 과정을 병렬적으로 실행하는 것이 적용 가능할 것이므로, 순서도는 시계열적인 순서로 한정되는 것은 아니다.
- [0176] 이상의 설명에서 예시적인 실시예들은 많은 다른 방식으로 구현될 수 있다는 것을 이해해야 한다. 하나 이상의 예시들에서 설명된 기능들 혹은 방법들은 하드웨어, 소프트웨어, 펌웨어 또는 이들의 임의의 조합으로 구현될 수 있다. 본 명세서에서 설명된 기능적 컴포넌트들은 그들의 구현 독립성을 특히 더 강조하기 위해 "...부(unit)" 로 라벨링되었음을 이해해야 한다.
- [0177] 한편, 본 실시예에서 설명된 다양한 기능들 혹은 방법들은 하나 이상의 프로세서에 의해 관독되고 실행될 수 있는 비일시적 기록매체에 저장된 명령어들로 구현될 수도 있다. 비일시적 기록매체는, 예를 들어, 컴퓨터 시스템에 의하여 관독가능한 형태로 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 예를 들어, 비일시적 기록매체는 EPROM(erasable programmable read only memory), 플래시 드라이브, 광학 드라이브, 자기 하드 드라이브, 솔리드 스테이트 드라이브(SSD)와 같은 저장매체를 포함한다.
- [0178] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

- [0179] 110: 공통특징 추출기 112: 공통특징 부호화기
- 114: 특징대영상 매핑기
- 120: 태스크 특징 추출기 122: 태스크 특징 부호화기
- 130: 비디오 부호화기 140: 다중화기
- 150: 신경망 연결부
- 1012: 공통특징 복호화기 1014: 특징대영상 매핑기
- 1022: 태스크 특징 복호화기
- 1030: 비디오 부호화기 1040: 다중화기
- 1050: 신경망 연결부

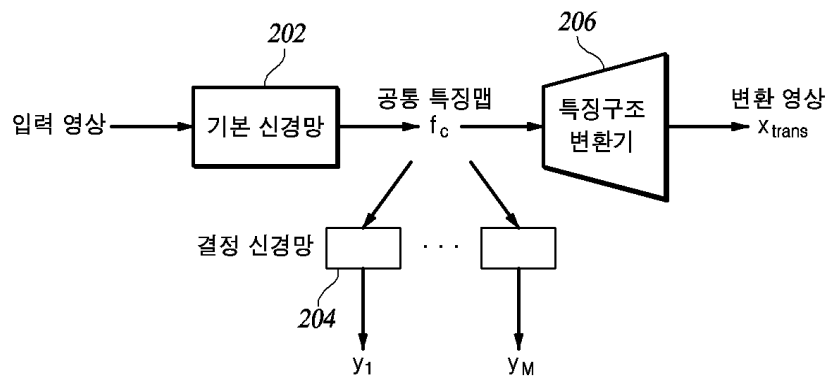
도면

도면1

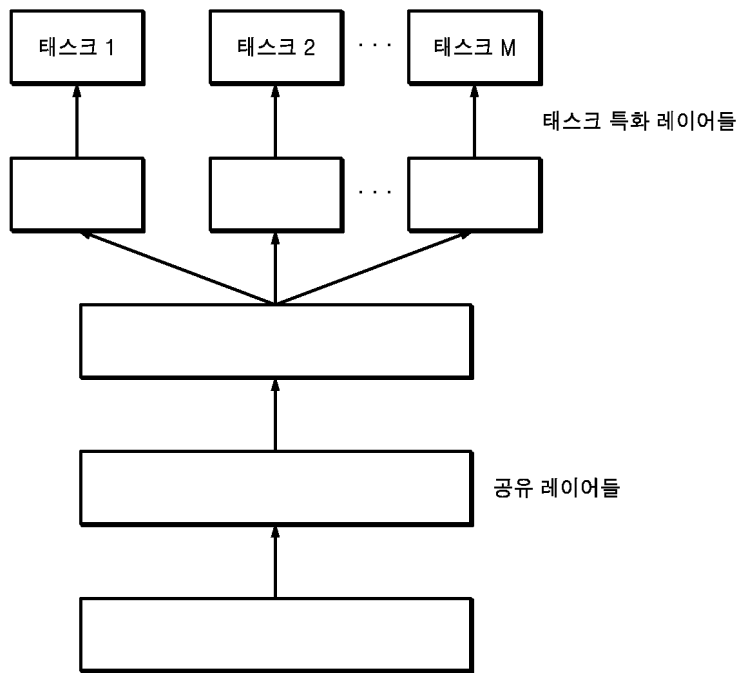


도면2

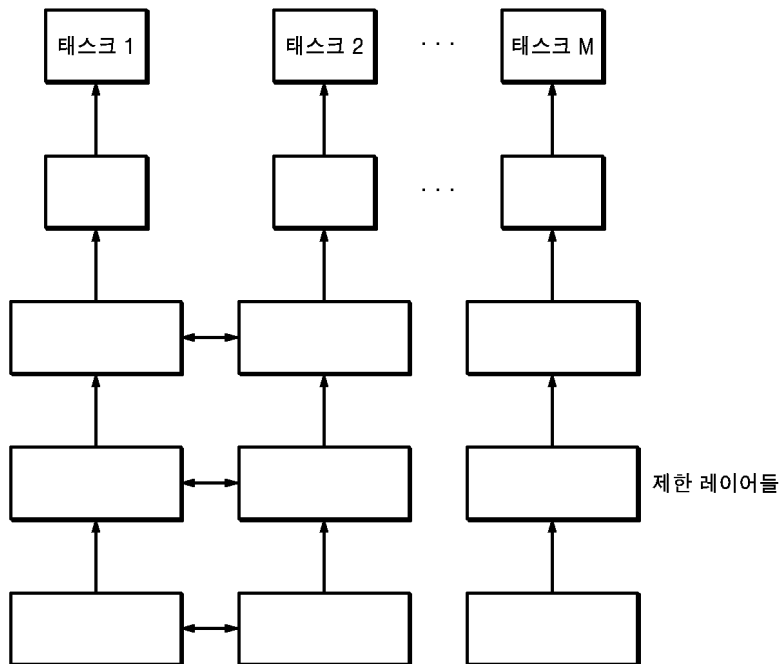
110



도면3a

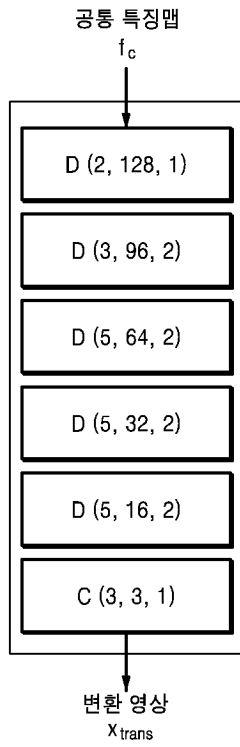


도면3b

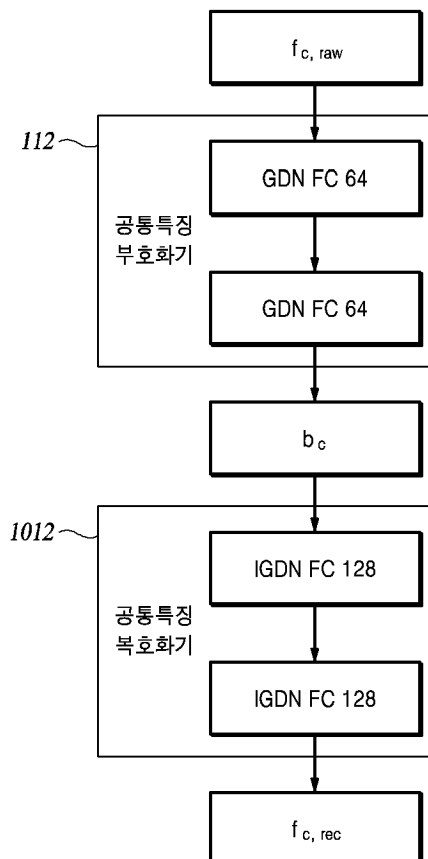


도면4

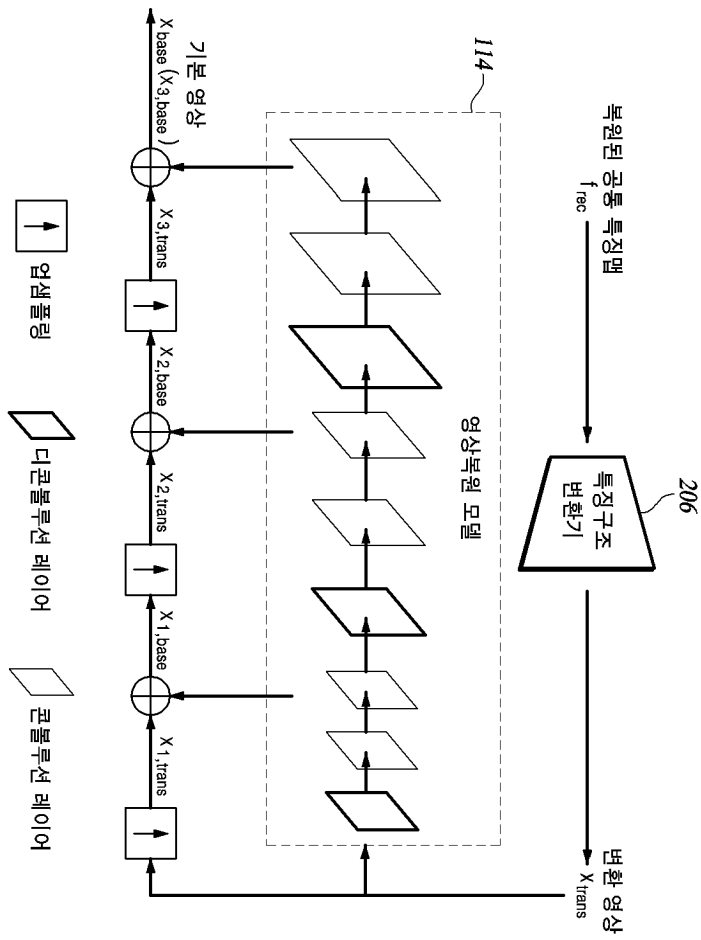
206



도면5

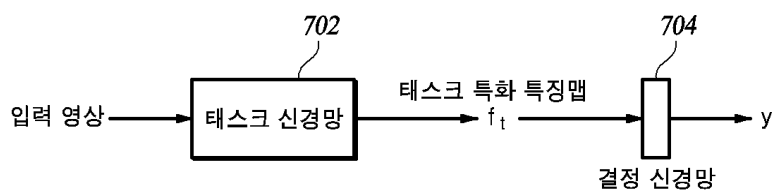


도면6

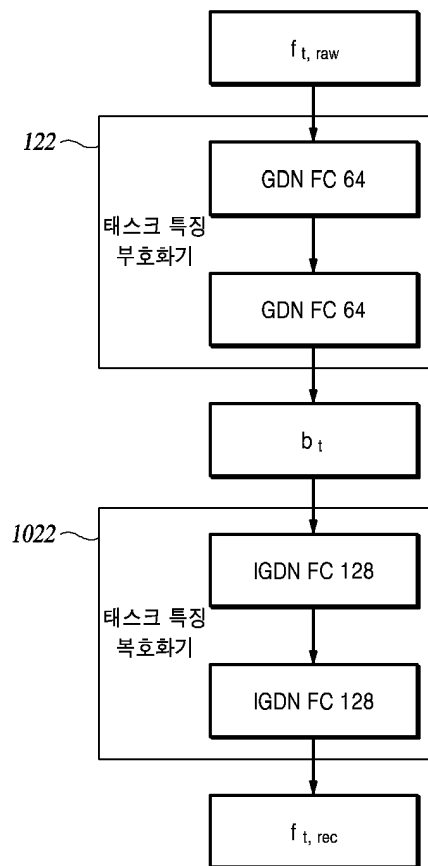


도면7

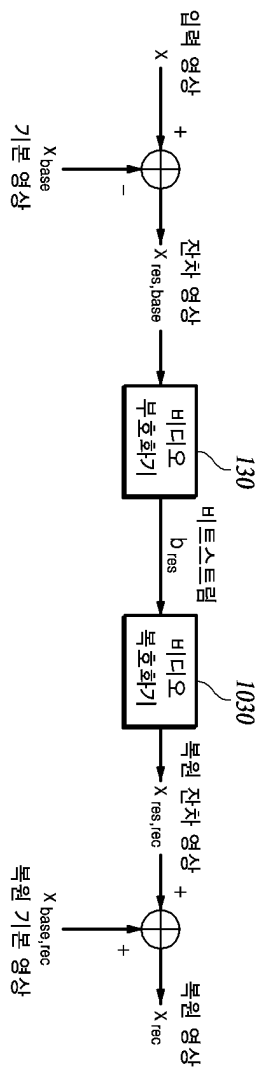
120



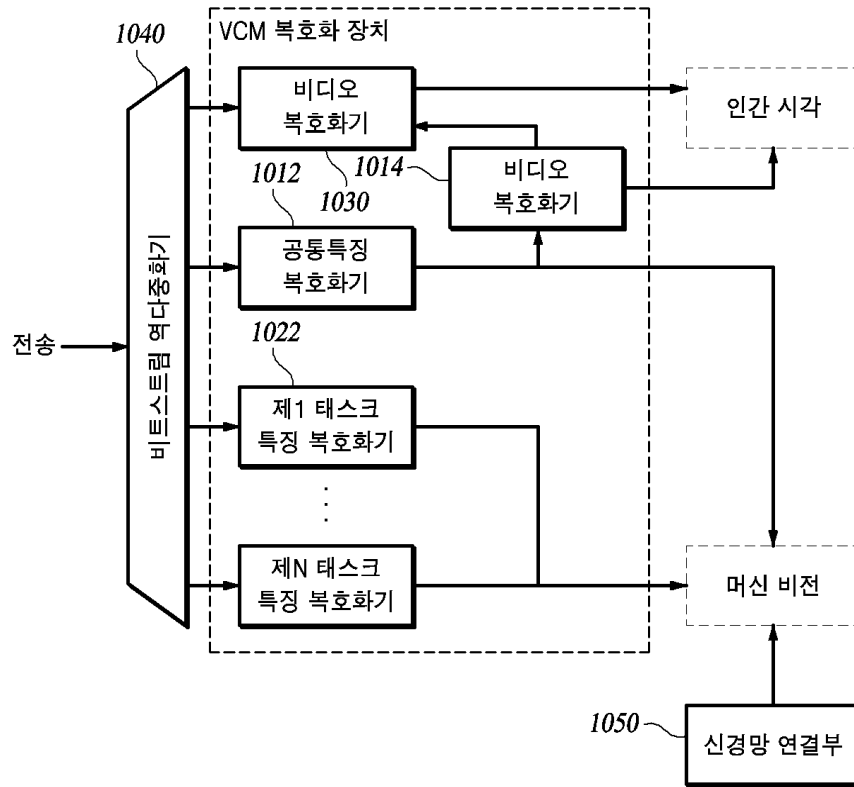
도면8



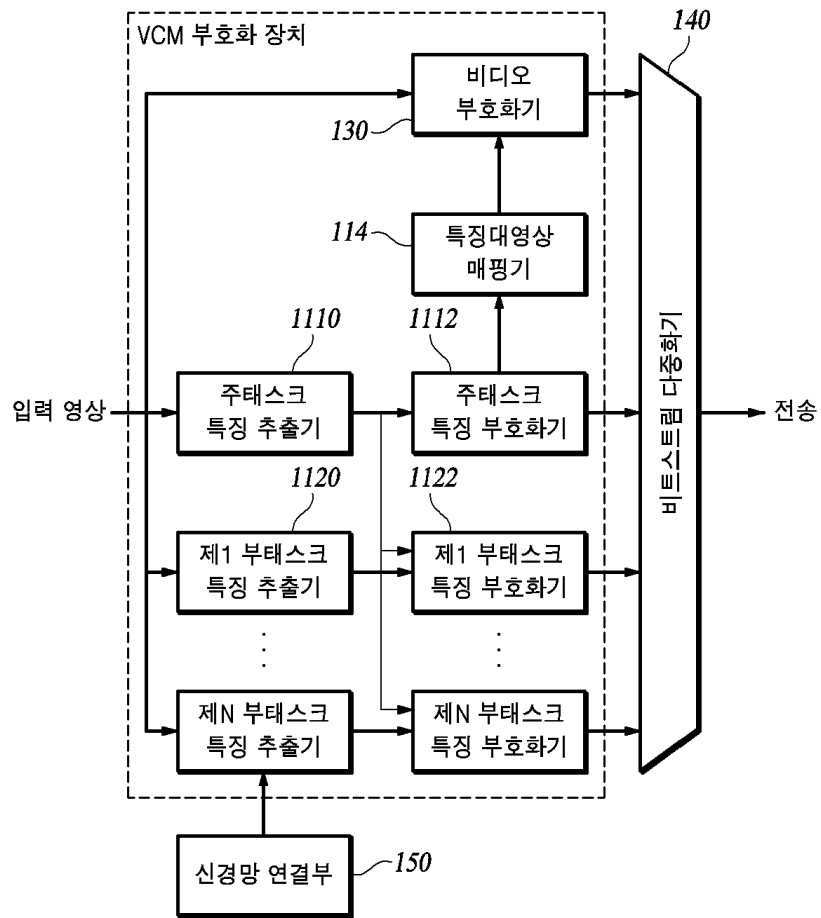
도면9



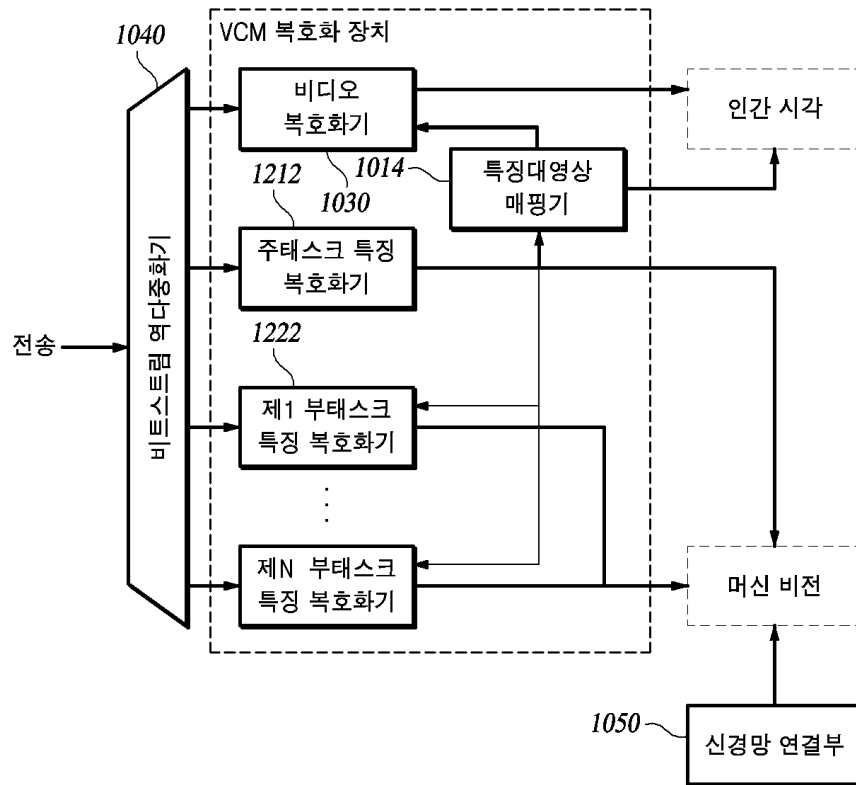
도면10



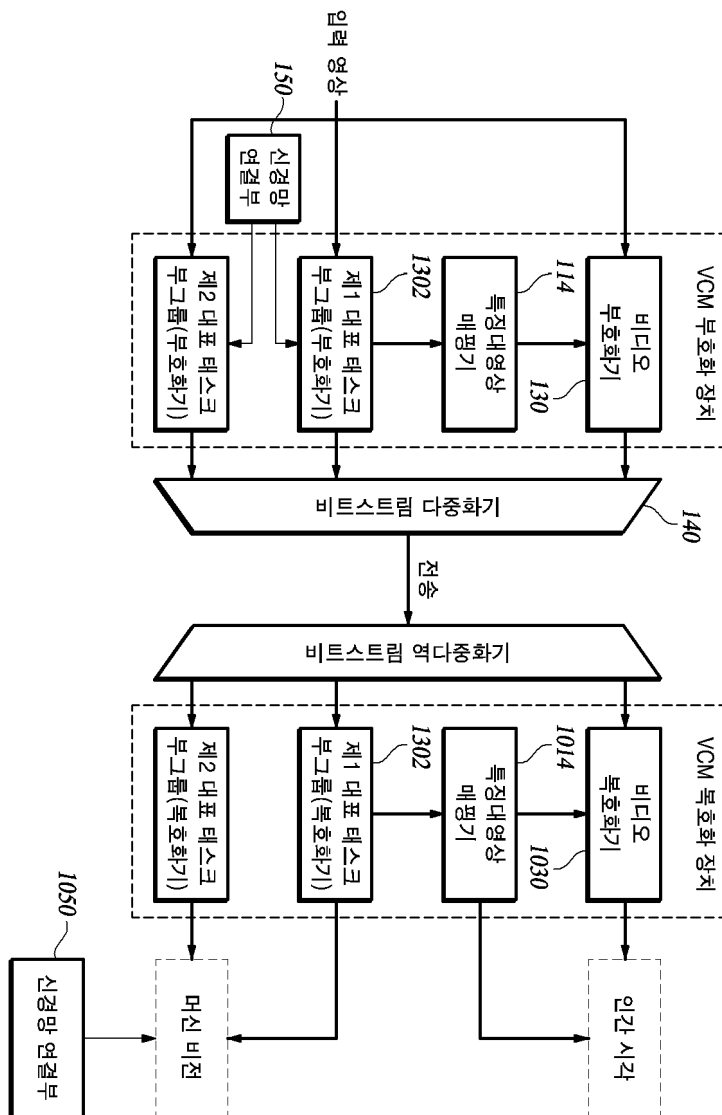
도면11



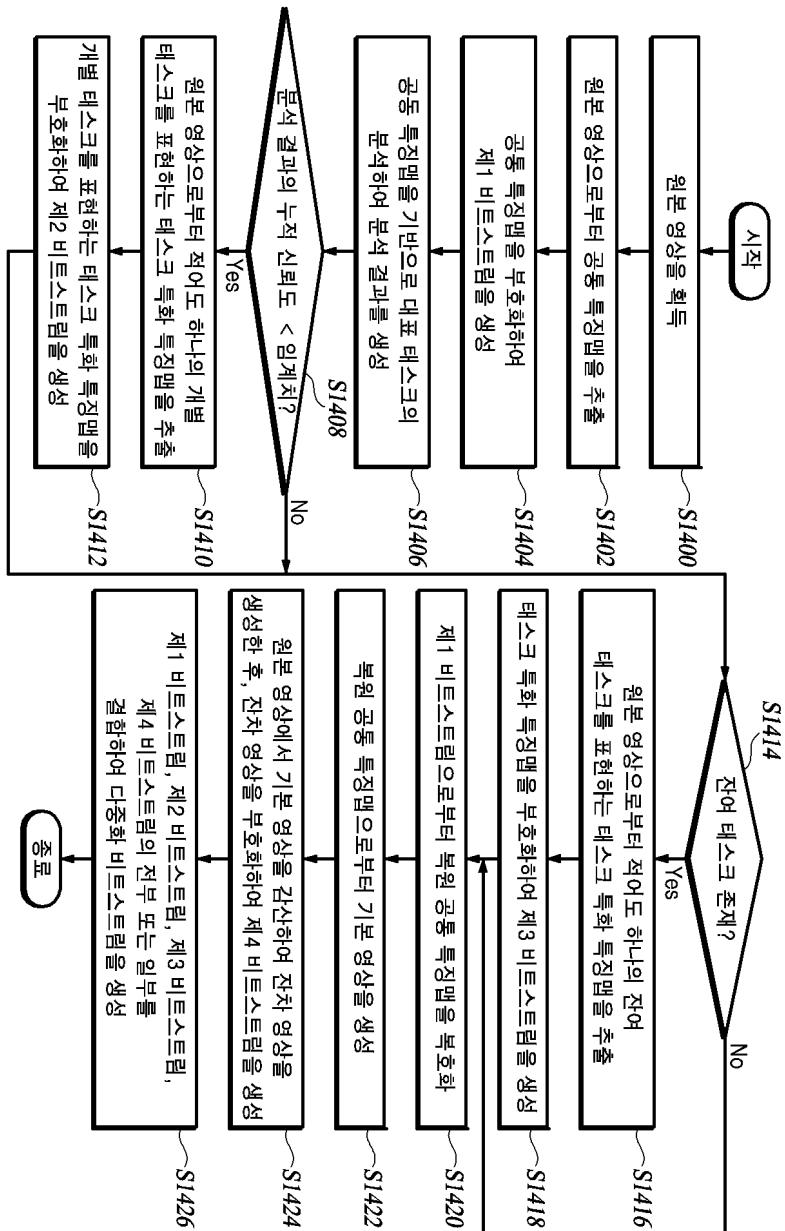
도면12



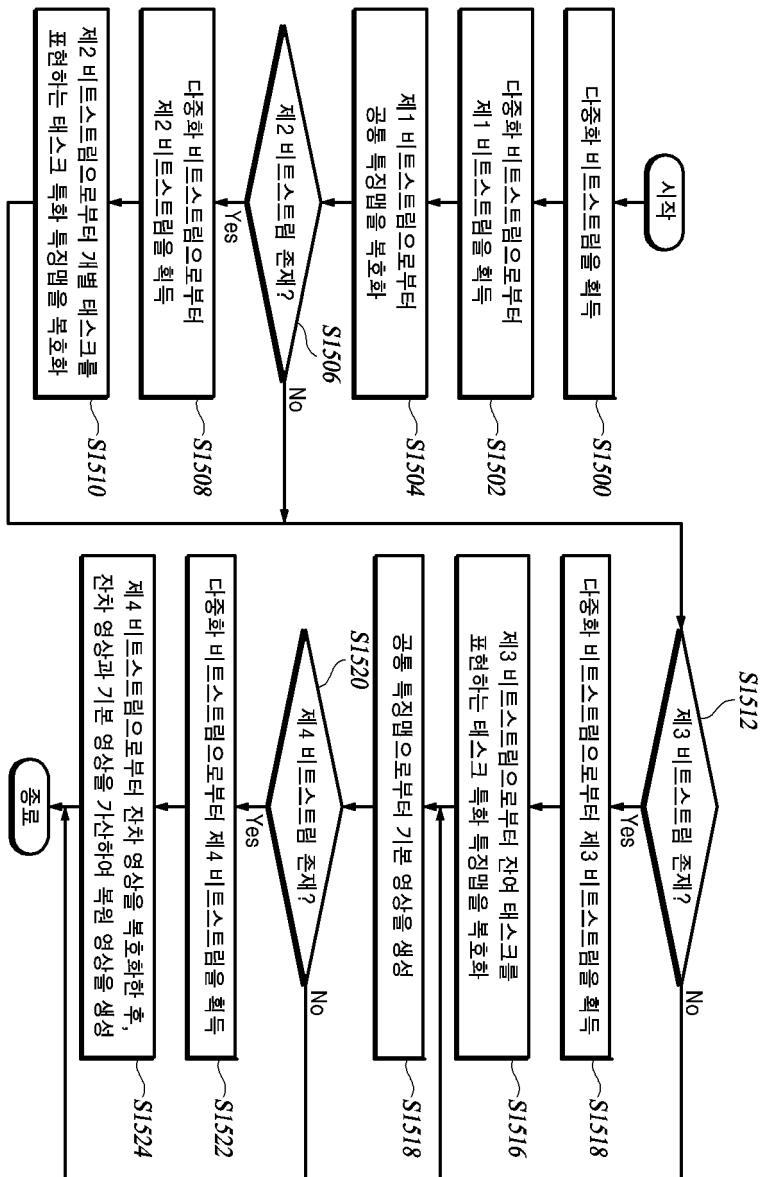
도면13



도면14



도면15



도면16

