



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0115043
(43) 공개일자 2023년08월02일

(51) 국제특허분류(Int. Cl.)

G06T 3/40 (2006.01) G06N 3/04 (2023.01)
G06N 3/08 (2023.01) G06T 7/40 (2017.01)
G06T 7/50 (2017.01)

(52) CPC특허분류

G06T 3/4053 (2013.01)
G06N 3/047 (2023.01)

(21) 출원번호 10-2022-0011541

(22) 출원일자 2022년01월26일

심사청구일자 없음

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

이진영

서울특별시 광진구 능동로 209

더 반 레

서울특별시 광진구 능동로 209

최용호

서울특별시 송파구 새말로 185, 101동 302호(문정동, 인택스빌아파트)

(74) 대리인

성병기

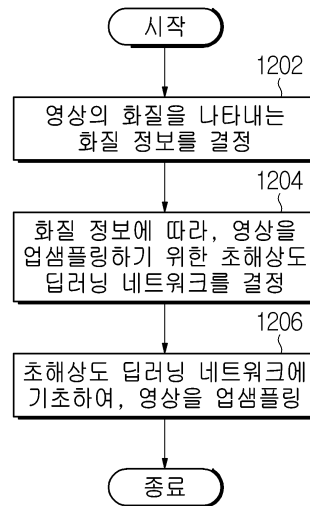
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **영상의 화질에 따라 초해상도 딥러닝 네트워크를 적용하는 비디오 처리 방법 및 비디오 처리 장치**

(57) 요약

본 개시에 있어서, 영상의 화질을 나타내는 화질 정보를 결정하는 단계, 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계, 및 초해상도 딥러닝 네트워크에 기초하여, 영상을 업샘플링하는 단계를 포함하는 영상 처리 방법이 제공된다.

대표도 - 도12



(52) CPC특허분류

G06N 3/08 (2023.01)
 G06T 7/40 (2013.01)
 G06T 7/50 (2017.01)
 G06T 2207/10016 (2013.01)
 G06T 2207/20081 (2013.01)
 G06T 2207/20084 (2013.01)
 G06T 2207/30168 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711130928
과제번호	2020M3F6A1109603
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	미래국방혁신기술개발(R&D)
연구과제명	전장 환경 인식 및 무기체계 성능을 지원하는 AI 카메라 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2021.02.01 ~ 2022.01.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711139507
과제번호	2021-0-02067-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발(R&D)
연구과제명	다목적 비디오 검색을 위한 차세대 인공신경망 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2021.07.01 ~ 2022.03.31

명세서

청구범위

청구항 1

영상의 화질을 나타내는 화질 정보를 결정하는 단계;

상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계; 및

상기 초해상도 딥러닝 네트워크에 기초하여, 상기 영상을 업샘플링하는 단계를 포함하는 영상 처리 방법.

청구항 2

제1항에 있어서,

상기 화질 정보는 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 결정되는 것을 특징으로 하는 영상 처리 방법.

청구항 3

제1항에 있어서,

상기 화질 정보는 화질 측정 딥러닝 네트워크에 결정되는 것을 특징으로 하는 영상 처리 방법.

청구항 4

제3항에 있어서,

상기 화질 정보는,

영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나를 상기 화질 측정 딥러닝 네트워크에 입력함으로써, 상기 화질 측정 딥러닝 네트워크로부터 출력된 것을 특징으로 하는 영상 처리 방법.

청구항 5

제3항에 있어서,

상기 화질 정보는,

상기 화질 측정 딥러닝 네트워크로부터 출력된 출력 벡터에 따라 결정되며, 상기 출력 벡터는 복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타내는 것을 특징으로 하는 영상 처리 방법.

청구항 6

제1항에 있어서,

상기 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정되는 것을 특징으로 하는 영상 처리 방법.

청구항 7

제1항에 있어서,

상기 영상 처리 방법은,

상기 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 단계를 더 포함하고,

상기 플래그 정보가 상기 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 영상의 화질을 나타내는 화질 정보를 결정하는 단계, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계, 및 상기 초해상도 딥러닝 네트워크에 기초하여, 상기 영상을 업샘플링하는 단계가 수행되는 것을 특징으로 하는 영상 처리 방법.

청구항 8

제7항에 있어서,

상기 플래그 정보는,

시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득되는 것을 특징으로 하는 영상 처리 방법.

청구항 9

제1항에 있어서,

상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵(feature map) 중 적어도 하나를 포함하고,

상기 화질 정보는 상기 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 결정되는 것을 특징으로 하는 영상 처리 방법.

청구항 10

제1항에 있어서,

상기 영상 처리 방법은,

상기 업샘플링된 영상에 대한 비디오 캡션을 수행하는 단계를 더 포함하는 것을 특징으로 하는 영상 처리 방법.

청구항 11

복수의 초해상도 딥러닝 네트워크; 및

영상의 화질을 나타내는 화질 정보를 결정하고, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 화질 결정 유닛을 포함하며,

상기 복수의 초해상도 딥러닝 네트워크 중 상기 화질 정보에 따라 선택된 초해상도 딥러닝 네트워크에 의하여 상기 영상이 업샘플링되는 것을 특징으로 하는 영상 처리 장치.

청구항 12

제11항에 있어서,

상기 화질 결정 유닛은 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 결정하는 것을 특징으로 하는 영상 처리 장치.

청구항 13

제11항에 있어서,

상기 화질 결정 유닛은 화질 측정 딥러닝 네트워크를 포함하고,

상기 화질 측정 딥러닝 네트워크에 의하여 상기 화질 정보가 결정되는 것을 특징으로 하는 영상 처리 장치.

청구항 14

제13항에 있어서,

상기 화질 측정 딥러닝 네트워크는,

영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 출력하는 것을 특징으로 하는 영상 처리 장치.

청구항 15

제13항에 있어서,

상기 화질 측정 딥러닝 네트워크는,

복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타내는 출력 벡터를 출력하고, 상기 출력 벡터에 따라 상기 화질 정보를 결정하는 것을 특징으로 하는 영상 처리 장치.

청구항 16

제11항에 있어서,

상기 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정되는 것을 특징으로 하는 영상 처리 장치.

청구항 17

제11항에 있어서,

상기 영상 처리 장치는,

상기 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 영상 복호화 유닛을 더 포함하고,

상기 플래그 정보가 상기 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 상기 화질 결정 유닛, 및 상기 복수의 초해상도 딥러닝 네트워크의 기능이 수행되고,

상기 플래그 정보는 시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득되는 것을 특징으로 하는 영상 처리 장치.

청구항 18

제11항에 있어서,

상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 포함하고,

상기 화질 결정 유닛은 상기 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 상기 화질 정보를 결정하는 것을 특징으로 하는 영상 처리 장치.

청구항 19

제11항에 있어서,

상기 영상 처리 장치는,

상기 업샘플링된 영상에 대한 비디오 캡션을 수행하는 비디오 캡션 유닛을 더 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 20

영상 처리 방법의 각 단계가 구현된 컴퓨터 프로그램이 저장된 컴퓨터로 판독가능한 기록 매체에 있어서,

상기 영상 처리 방법은,

영상의 화질을 나타내는 화질 정보를 결정하는 단계,

상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계, 및

상기 초해상도 딥러닝 네트워크에 기초하여, 상기 영상을 업샘플링하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 판독가능한 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 비디오 처리 방법에 관한 발명으로, 보다 구체적으로 영상의 화질에 따라 초해상도 딥러닝 네트워크를 적용하는 비디오 처리 방법에 관한 발명이다.

배경 기술

[0002] 영상은 전송 효율성을 고려하여, 부호화기에서 부호화되어 복호화기로 전송된 후, 복호화기에서 복호화된다. 부호화 과정에서 예측, 변화, 양자화 등을 통해 영상 내의 데이터를 효율적으로 줄여서 전송한다. 예측 과정에서는 영상 내의 연관성 혹은 영상 간의 연관성을 이용하여 원영상에 가까운 예측영상이 생성된다. 그리고 원영상과 예측영상 간의 잔차 신호가 생성되고, 상기 잔차 신호는 시간 영역에서 주파수 영역으로 변환된다. 그리고 변환된 잔차 신호에 대하여 양자화가 수행된다. 이 때 양자화 파라미터 (Quantization Parameter, QP)를 이용하여 양자화의 정도, 즉 압축의 정도가 결정된다. 일반적으로 양자화 파라미터가 높을수록 화질 열화가 증가한다. 네트워크 상에서 대부분의 영상들은 부호화되어 전송되므로, 영상 내 압축 에러가 존재할 수 있다. 그러므로 네트워크 및 저장 환경에 따라서 영상의 화질이 크게 열화될 수 있다.

[0003] 비디오 캡션 (Video caption)은 비디오 콘텐츠를 설명하기 위해 자연어를 수집하는 작업이다. 비디오 캡션은 최근 몇 년 동안 딥 러닝 기술 덕분에 눈에 띄는 발전이 있었다. 비디오 캡션에 대한 대부분의 기술은 주로 딥 러닝 네트워크 아키텍처에 중점을 두었다. 하지만, 비디오 품질 및 해상도가 비디오 캡션 성능에 미치는 영향이 매우 강력함에도 불구하고, 비디오 품질과 해상도는 일반적으로 비디오 캡션의 논의에서 중요하게 고려되지 않았다.

[0004] 따라서, 본 개시에서 비디오 캡션에 대한 비디오 품질과 해상도의 영향을 설명한다. 비디오 품질은 일반적으로 압축 데이터의 양자화로 인해 생성되는 압축 아티팩트에 따라 달라질 수 있다. 또한, 해상도가 높은 동영상은 다운샘플링 후 압축되는 경우가 있어 화질이 심하게 저하될 수 있다. 그리고 그러한 화질 저하는 비디오 캡션에 대한 에러를 발생시킬 수 있다.

[0005] 그러므로, 본 효율적인 비디오 캡션을 위한 초해상도(Super Resolution, SR) 기반의 품질 향상 방법 및 장치를 제안한다. 먼저, 상기 방법 및 장치에 따르면, 품질 분류 작업을 수행하여 각 프레임의 품질을 분석한다. 다음으로 초해상도가 구현된 딥러닝 네트워크를 이용하여, 품질과 해상도 측면에서 프레임이 향상된다. 마지막으로

비디오 캡션 네트워크는 향상된 프레임을 사용하여 정확한 캡션을 생성한다.

발명의 내용

해결하려는 과제

[0006] 물체 검출 (Object Detection) 및 비디오 캡션 (Video Caption) 등의 머신 비전 임무 수행 시 화질 열화에 의하여 에러가 발생할 수 있다. 이를 해결하기 위하여, 본 개시는 영상의 화질을 평가하고 분류하여, 결정된 화질에 최적화된 초해상도 혹은 초고화질 복원 기술을 적용하는 방법 및 장치를 제공한다. 그리고 본 개시는 강화된 화질의 영상을 이용하여 물체 검출 및 비디오 캡션 등의 머신 비전 임무를 수행하는 방법 및 장치를 제공한다.

과제의 해결 수단

[0007] 본 개시에서, 영상의 화질을 나타내는 화질 정보를 결정하는 단계, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계, 및 상기 초해상도 딥러닝 네트워크에 기초하여, 상기 영상을 업샘플링하는 단계를 포함하는 영상 처리 방법이 제공된다.

[0008] 일 실시예에 따르면, 상기 화질 정보는 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 결정될 수 있다.

[0009] 일 실시예에 따르면, 상기 화질 정보는 화질 측정 딥러닝 네트워크에 결정될 수 있다.

[0010] 일 실시예에 따르면, 상기 화질 정보는, 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나를 상기 화질 측정 딥러닝 네트워크에 입력함으로써, 상기 화질 측정 딥러닝 네트워크로부터 출력될 수 있다.

[0011] 일 실시예에 따르면, 상기 화질 정보는, 상기 화질 측정 딥러닝 네트워크로부터 출력된 출력 벡터에 따라 결정되며, 상기 출력 벡터는 복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타낼 수 있다.

[0012] 일 실시예에 따르면, 상기 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정될 수 있다.

[0013] 일 실시예에 따르면, 상기 영상 처리 방법은, 상기 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 단계를 더 포함하고, 상기 플래그 정보가 상기 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 영상의 화질을 나타내는 화질 정보를 결정하는 단계, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 단계, 및 상기 초해상도 딥러닝 네트워크에 기초하여, 상기 영상을 업샘플링하는 단계가 수행될 수 있다.

[0014] 일 실시예에 따르면, 상기 플래그 정보는, 시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득되는 것을 특징으로 할 수 있다.

[0015] 일 실시예에 따르면, 상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵(feature map) 중 적어도 하나를 포함하고, 상기 화질 정보는 상기 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 결정될 수 있다.

[0016] 일 실시예에 따르면, 상기 영상 처리 방법은, 상기 업샘플링된 영상에 대한 비디오 캡션을 수행하는 단계를 더 포함할 수 있다.

[0017] 본 개시에서, 복수의 초해상도 딥러닝 네트워크, 및 영상의 화질을 나타내는 화질 정보를 결정하고, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 화질 결정 유닛을 포함하며, 상기 복수의 초해상도 딥러닝 네트워크 중 상기 화질 정보에 따라 선택된 초해상도 딥러닝 네트워크에 의하여 상기 영상이 업샘플링되는 것을 특징으로 하는 영상 처리 장치가 제공된다.

[0018] 일 실시예에 따르면, 상기 화질 결정 유닛은 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 결정할 수 있다.

[0019] 일 실시예에 따르면, 상기 화질 결정 유닛은 화질 측정 딥러닝 네트워크를 포함하고, 상기 화질 측정 딥러닝 네트워크에 의하여 상기 화질 정보가 결정될 수 있다.

- [0020] 일 실시예에 따르면, 상기 화질 측정 딥러닝 네트워크는, 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 출력할 수 있다.
- [0021] 일 실시예에 따르면, 상기 화질 측정 딥러닝 네트워크는, 복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타내는 출력 벡터를 출력하고, 상기 출력 벡터에 따라 상기 화질 정보를 결정할 수 있다.
- [0022] 일 실시예에 따르면, 상기 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정될 수 있다.
- [0023] 일 실시예에 따르면, 상기 영상 처리 장치는, 상기 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 영상 복호화 유닛을 더 포함하고, 상기 플래그 정보가 상기 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 상기 화질 결정 유닛, 및 상기 복수의 초해상도 딥러닝 네트워크의 기능이 수행될 수 있다.
- [0024] 일 실시예에 따르면, 상기 플래그 정보는 시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득될 수 있다.
- [0025] 일 실시예에 따르면, 상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 포함하고, 상기 화질 결정 유닛은 상기 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 상기 화질 정보를 결정할 수 있다.
- [0026] 일 실시예에 따르면, 상기 영상 처리 장치는, 상기 업샘플링된 영상에 대한 비디오 캡션을 수행하는 비디오 캡션 유닛을 더 포함할 수 있다.
- [0027] 본 개시에 있어서, 상기의 영상 처리 방법의 각 단계가 구체화된 컴퓨터 프로그램이 저장된 컴퓨터로 판독가능한 기록 매체가 제공된다.

발명의 효과

- [0028] 본 개시에 따라, 영상의 화질을 나타내는 화질 정보를 결정하고, 화질 정보에 대응되는 초해상도 딥러닝 네트워크에 기초하여 영상을 업샘플링함으로써, 영상의 해상도 증가하고, 영상의 아티팩트가 감소할 수 있다.

도면의 간단한 설명

- [0029] 도 1은 본 발명의 일 실시예에 따른 영상 부호화 장치를 나타낸 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 영상 복호화 장치를 나타낸 블록도이다.
- 도 3은 일 실시예에 따른 비디오 캡션 방법의 프레임워크의 일 실시 예를 도시한다.
- 도 4는 화질 측정 딥러닝 네트워크의 일 실시 예를 나타낸다.
- 도 5는 초해상도 딥러닝 네트워크의 일 실시 예를 도시한다.
- 도 6은 비디오 프레임의 시퀀스를 입력으로 사용하고 단어 시퀀스를 출력으로 생성하는 인코더-디코더 구조의 일 실시 예를 도시한다.
- 도 7은 VCM을 구현하기 위한 파이프 라인의 제1 실시 예를 나타낸다.
- 도 8은 VCM을 구현하기 위한 파이프 라인의 제2 실시 예를 나타낸다.
- 도 9는 복수의 특징 맵을 포함하는 특징 프레임을 구성하는 방법의 일 실시 예를 설명한다.
- 도 10은 특징 프레임에 포함된 특징 맵을 재정렬하기 위한, 공간 우선 순위 할당 방법 및 시간 우선 순위 할당 방법을 설명한다.
- 도 11은 초해상도(SR) 기술에 따른 특징 맵 압축 방법의 일 실시 예를 도시한다.
- 도 12는 영상 처리 방법의 일 실시예를 도시한다.
- 도 13은 영상 처리 장치의 일 실시예를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0030] 본 개시는 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다. 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다. 후술하는 예시적 실시예들에 대한 상세한 설명은, 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 실시예를 실시할 수 있기에 충분하도록 상세히 설명된다. 다양한 실시예들은 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 실시예의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 예시적 실시예들의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다.
- [0031] 본 발명에서 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0032] 본 발명의 어떤 구성 요소가 다른 구성 요소에 “연결되어” 있다거나 “접속되어” 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있으나, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 “직접 연결되어” 있다거나 “직접 접속되어” 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0033] 본 발명의 실시예에 나타나는 구성부들은 서로 다른 특징적인 기능들을 나타내기 위해 독립적으로 도시되는 것으로, 각 구성부들이 분리된 하드웨어나 하나의 소프트웨어 구성단위로 이루어짐을 의미하지 않는다. 즉, 각 구성부는 설명의 편의상 각각의 구성부로 나열하여 포함한 것으로 각 구성부 중 적어도 두 개의 구성부가 합쳐져 하나의 구성부로 이루어지거나, 하나의 구성부가 복수 개의 구성부로 나뉘어져 기능을 수행할 수 있고 이러한 각 구성부의 통합된 실시예 및 분리된 실시예도 본 발명의 본질에서 벗어나지 않는 한 본 발명의 권리범위에 포함된다.
- [0034] 본 발명에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 발명에서, “포함하다” 또는 “가지다” 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다. 즉, 본 발명에서 특정 구성을 “포함” 한다고 기술하는 내용은 해당 구성 이외의 구성을 배제하는 것이 아니며, 추가적인 구성이 본 발명의 실시 또는 본 발명의 기술적 사상의 범위에 포함될 수 있음을 의미한다.
- [0035] 본 발명의 일부의 구성 요소는 본 발명에서 본질적인 기능을 수행하는 필수적인 구성 요소는 아니고 단지 성능을 향상시키기 위한 선택적 구성 요소일 수 있다. 본 발명은 단지 성능 향상을 위해 사용되는 구성 요소를 제외한 본 발명의 본질을 구현하는데 필수적인 구성부만을 포함하여 구현될 수 있고, 단지 성능 향상을 위해 사용되는 선택적 구성 요소를 제외한 필수 구성 요소만을 포함한 구조도 본 발명의 권리범위에 포함된다.
- [0036] 이하, 도면을 참조하여 본 발명의 실시 형태에 대하여 구체적으로 설명한다. 본 명세서의 실시예를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 명세서의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략하고, 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [0038] 도 1은 본 발명의 일실시예에 따른 영상 부호화 장치를 나타낸 블록도이다.
- [0039] 도 1을 참조하면, 영상 부호화 장치(100)는 픽처 분할부(110), 예측부(120, 125), 변환부(130), 양자화부(135), 재정렬부(160), 엔트로피 부호화부(165), 역양자화부(140), 역변환부(145), 필터부(150) 및 메모리(155)를 포함할 수 있다.

- [0040] 도 1에 나타난 각 구성부들은 영상 부호화 장치에서 서로 다른 특징적인 기능들을 나타내기 위해 독립적으로 도시한 것으로, 각 구성부들이 분리된 하드웨어나 하나의 소프트웨어 구성단위로 이루어짐을 의미하지 않는다. 즉, 각 구성부는 설명의 편의상 각각의 구성부로 나열하여 포함한 것으로 각 구성부 중 적어도 두 개의 구성부가 합쳐져 하나의 구성부로 이루어지거나, 하나의 구성부가 복수개의 구성부로 나뉘어져 기능을 수행할 수 있고 이러한 각 구성부의 통합된 실시예 및 분리된 실시예도 본 발명의 본질에서 벗어나지 않는 한 본 발명의 권리범위에 포함된다.
- [0041] 또한, 일부의 구성 요소는 본 발명에서 본질적인 기능을 수행하는 필수적인 구성 요소는 아니고 단지 성능을 향상시키기 위한 선택적 구성 요소일 수 있다. 본 발명은 단지 성능 향상을 위해 사용되는 구성 요소를 제외한 본 발명의 본질을 구현하는데 필수적인 구성부만을 포함하여 구현될 수 있고, 단지 성능 향상을 위해 사용되는 선택적 구성 요소를 제외한 필수 구성 요소만을 포함한 구조도 본 발명의 권리범위에 포함된다.
- [0042] 픽처 분할부(110)는 입력된 픽처를 적어도 하나의 처리 단위로 분할할 수 있다. 이때, 처리 단위는 예측 단위(Prediction Unit: PU)일 수도 있고, 변환 단위(Transform Unit: TU)일 수도 있으며, 부호화 단위(Coding Unit: CU)일 수도 있다. 픽처 분할부(110)에서는 하나의 픽처에 대해 복수의 부호화 단위, 예측 단위 및 변환 단위의 조합으로 분할하고 소정의 기준(예를 들어, 비용 함수)으로 하나의 부호화 단위, 예측 단위 및 변환 단위 조합을 선택하여 픽처를 부호화 할 수 있다.
- [0043] 예를 들어, 하나의 픽처는 복수개의 부호화 단위로 분할될 수 있다. 픽처에서 부호화 단위를 분할하기 위해서는 쿼드 트리 구조(Quad Tree Structure)와 같은 계층적인 트리 구조를 사용할 수 있는데 하나의 영상 또는 최대 크기 부호화 단위(largest coding unit)를 루트로 하여 다른 부호화 단위로 분할되는 부호화 유닛은 분할된 부호화 단위의 개수만큼의 자식 노드를 가지고 분할될 수 있다. 일정한 제한에 따라 더 이상 분할되지 않는 부호화 단위는 리프 노드가 된다. 즉, 하나의 코딩 유닛에 대하여 정방형 분할만이 가능하다고 가정하는 경우, 하나의 부호화 단위는 최대 4개의 다른 부호화 단위로 분할될 수 있다.
- [0044] 이하, 본 발명의 실시예에서는 부호화 단위는 부호화를 수행하는 단위의 의미로 사용할 수도 있고, 복호화를 수행하는 단위의 의미로 사용할 수도 있다.
- [0045] 예측 단위는 하나의 부호화 단위 내에서 동일한 크기의 적어도 하나의 정사각형 또는 직사각형 등의 형태를 가지고 분할된 것일 수도 있고, 하나의 부호화 단위 내에서 분할된 예측 단위 중 어느 하나의 예측 단위가 다른 하나의 예측 단위와 상이한 형태 및/또는 크기를 가지도록 분할된 것일 수도 있다.
- [0046] 부호화 단위를 기초로 인트라 예측을 수행하는 예측 단위를 생성시 최소 부호화 단위가 아닌 경우, 복수의 예측 단위 NxN으로 분할하지 않고 인트라 예측을 수행할 수 있다.
- [0047] 예측부(120, 125)는 인터 예측을 수행하는 인터 예측부(120)와 인트라 예측을 수행하는 인트라 예측부(125)를 포함할 수 있다. 예측 단위에 대해 인터 예측을 사용할 것인지 또는 인트라 예측을 수행할 것인지를 결정하고, 각 예측 방법에 따른 구체적인 정보(예컨대, 인트라 예측 모드, 모션 벡터, 참조 픽처 등)를 결정할 수 있다. 이때, 예측이 수행되는 처리 단위와 예측 방법 및 구체적인 내용이 정해지는 처리 단위는 다를 수 있다. 예컨대, 예측의 방법과 예측 모드 등은 예측 단위로 결정되고, 예측의 수행은 변환 단위로 수행될 수도 있다. 생성된 예측 블록과 원본 블록 사이의 잔차 값(잔차 블록)은 변환부(130)로 입력될 수 있다. 또한, 예측을 위해 사용한 예측 모드 정보, 모션 벡터 정보 등은 잔차 값과 함께 엔트로피 부호화부(165)에서 부호화되어 복호화기에 전달될 수 있다. 특정한 부호화 모드를 사용할 경우, 예측부(120, 125)를 통해 예측 블록을 생성하지 않고, 원본 블록을 그대로 부호화하여 복호화부에 전송하는 것도 가능하다.
- [0048] 인터 예측부(120)는 현재 픽처의 이전 픽처 또는 이후 픽처 중 적어도 하나의 픽처의 정보를 기초로 예측 단위를 예측할 수도 있고, 경우에 따라서는 현재 픽처 내의 부호화가 완료된 일부 영역의 정보를 기초로 예측 단위를 예측할 수도 있다. 인터 예측부(120)는 참조 픽처 보간부, 모션 예측부, 움직임 보상부를 포함할 수 있다.
- [0049] 참조 픽처 보간부에서는 메모리(155)로부터 참조 픽처 정보를 제공받고 참조 픽처에서 정수 화소 이하의 화소 정보를 생성할 수 있다. 휘도 화소의 경우, 1/4 화소 단위로 정수 화소 이하의 화소 정보를 생성하기 위해 필터 계수를 달리하는 DCT 기반의 8탭 보간 필터(DCT-based Interpolation Filter)가 사용될 수 있다. 색차 신호의 경우 1/8 화소 단위로 정수 화소 이하의 화소 정보를 생성하기 위해 필터 계수를 달리하는 DCT 기반의 4탭 보간 필터(DCT-based Interpolation Filter)가 사용될 수 있다.
- [0050] 모션 예측부는 참조 픽처 보간부에 의해 보간된 참조 픽처를 기초로 모션 예측을 수행할 수 있다. 모션 벡터를

산출하기 위한 방법으로 FBMA(Full search-based Block Matching Algorithm), TSS(Three Step Search), NTS(New Three-Step Search Algorithm) 등 다양한 방법이 사용될 수 있다. 모션 벡터는 보간된 화소를 기초로 1/2 또는 1/4 화소 단위의 모션 벡터 값을 가질 수 있다. 모션 예측부에서는 모션 예측 방법을 다르게 하여 현재 예측 단위를 예측할 수 있다. 모션 예측 방법으로 스킵(Skip) 방법, 머지(Merge) 방법, AMVP(Advanced Motion Vector Prediction) 방법, 인트라 블록 카피(Intra Block Copy) 방법 등 다양한 방법이 사용될 수 있다.

[0051] 인트라 예측부(125)는 현재 픽처 내의 화소 정보인 현재 블록 주변의 참조 픽셀 정보를 기초로 예측 단위를 생성할 수 있다. 현재 예측 단위의 주변 블록이 인트라 예측을 수행한 블록이어서, 참조 픽셀이 인트라 예측을 수행한 픽셀일 경우, 인트라 예측을 수행한 블록에 포함되는 참조 픽셀을 주변의 인트라 예측을 수행한 블록의 참조 픽셀 정보로 대체하여 사용할 수 있다. 즉, 참조 픽셀이 가용하지 않는 경우, 가용하지 않은 참조 픽셀 정보를 가용한 참조 픽셀 중 적어도 하나의 참조 픽셀로 대체하여 사용할 수 있다.

[0052] 인트라 예측에서 예측 모드는 참조 픽셀 정보를 예측 방향에 따라 사용하는 방향성 예측 모드와 예측을 수행시 방향성 정보를 사용하지 않는 비방향성 모드를 가질 수 있다. 상기 방향성 예측 모드의 개수는 HEVC 표준에 정의된 33개와 같거나 그 이상일 수 있으며, 예를 들어 60 내지 70 범위 내의 개수로 확장될 수 있다. 휘도 정보를 예측하기 위한 모드와 색차 정보를 예측하기 위한 모드가 상이할 수 있고, 색차 정보를 예측하기 위해 휘도 정보를 예측하기 위해 사용된 인트라 예측 모드 정보 또는 예측된 휘도 신호 정보를 활용할 수 있다.

[0053] 인트라 예측을 수행할 때 예측 단위의 크기와 변환 단위의 크기가 동일할 경우, 예측 단위의 좌측에 존재하는 픽셀, 좌측 상단에 존재하는 픽셀, 상단에 존재하는 픽셀을 기초로 예측 단위에 대한 인트라 예측을 수행할 수 있다. 그러나 인트라 예측을 수행할 때 예측 단위의 크기와 변환 단위의 크기가 상이할 경우, 변환 단위를 기초로 한 참조 픽셀을 이용하여 인트라 예측을 수행할 수 있다. 또한, 최소 부호화 단위에 대해서만 $N \times N$ 분할을 사용하는 인트라 예측을 사용할 수 있다.

[0054] 인트라 예측 방법은 예측 모드에 따라 참조 화소에 AIS(Adaptive Intra Smoothing) 필터를 적용한 후 예측 블록을 생성할 수 있다. 참조 화소에 적용되는 AIS 필터의 종류는 상이할 수 있다. 인트라 예측 방법을 수행하기 위해 현재 예측 단위의 인트라 예측 모드는 현재 예측 단위의 주변에 존재하는 예측 단위의 인트라 예측 모드로부터 예측할 수 있다. 주변 예측 단위로부터 예측된 모드 정보를 이용하여 현재 예측 단위의 예측 모드를 예측하는 경우, 현재 예측 단위와 주변 예측 단위의 인트라 예측 모드가 동일하면 소정의 플래그 정보를 이용하여 현재 예측 단위와 주변 예측 단위의 예측 모드가 동일하다는 정보를 전송할 수 있고, 만약 현재 예측 단위와 주변 예측 단위의 예측 모드가 상이하면 엔트로피 부호화를 수행하여 현재 블록의 예측 모드 정보를 부호화할 수 있다.

[0055] 또한, 예측부(120, 125)에서 생성된 예측 단위를 기초로 예측을 수행한 예측 단위와 예측 단위의 원본 블록과 차이 값인 잔차 값(Residual) 정보를 포함하는 잔차 블록이 생성될 수 있다. 생성된 잔차 블록은 변환부(130)로 입력될 수 있다.

[0056] 변환부(130)에서는 원본 블록과 예측부(120, 125)를 통해 생성된 예측 단위의 잔차 값(residual)정보를 포함한 잔차 블록을 DCT(Discrete Cosine Transform), DST(Discrete Sine Transform), KLT와 같은 변환 방법을 사용하여 변환시킬 수 있다. 잔차 블록을 변환하기 위해 DCT를 적용할지, DST를 적용할지 또는 KLT를 적용할지는 잔차 블록을 생성하기 위해 사용된 예측 단위의 인트라 예측 모드 정보를 기초로 결정할 수 있다.

[0057] 양자화부(135)는 변환부(130)에서 주파수 영역으로 변환된 값들을 양자화할 수 있다. 블록에 따라 또는 영상의 중요도에 따라 양자화 계수는 변할 수 있다. 양자화부(135)에서 산출된 값은 역양자화부(140)와 재정렬부(160)에 제공될 수 있다.

[0058] 재정렬부(160)는 양자화된 잔차 값에 대해 계수 값의 재정렬을 수행할 수 있다.

[0059] 재정렬부(160)는 계수 스캐닝(Coefficient Scanning) 방법을 통해 2차원의 블록 형태 계수를 1차원의 벡터 형태로 변경할 수 있다. 예를 들어, 재정렬부(160)에서는 지그-재그 스캔(Zig-Zag Scan)방법을 이용하여 DC 계수부터 고주파수 영역의 계수까지 스캔하여 1차원 벡터 형태로 변경시킬 수 있다. 변환 단위의 크기 및 인트라 예측 모드에 따라 지그-재그 스캔 대신 2차원의 블록 형태 계수를 열 방향으로 스캔하는 수직 스캔, 2차원의 블록 형태 계수를 행 방향으로 스캔하는 수평 스캔이 사용될 수도 있다. 즉, 변환 단위의 크기 및 인트라 예측 모드에 따라 지그-재그 스캔, 수직 방향 스캔 및 수평 방향 스캔 중 어떠한 스캔 방법이 사용될지 여부를 결정할 수 있다.

- [0060] 엔트로피 부호화부(165)는 재정렬부(160)에 의해 산출된 값들을 기초로 엔트로피 부호화를 수행할 수 있다. 엔트로피 부호화는 예를 들어, 지수 골롬(Exponential Golomb), CAVLC(Context-Adaptive Variable Length Coding), CABAC(Context-Adaptive Binary Arithmetic Coding)과 같은 다양한 부호화 방법을 사용할 수 있다.
- [0061] 엔트로피 부호화부(165)는 재정렬부(160) 및 예측부(120, 125)로부터 부호화 단위의 잔차 값 계수 정보 및 블록 타입 정보, 예측 모드 정보, 분할 단위 정보, 예측 단위 정보 및 전송 단위 정보, 모션 벡터 정보, 참조 프레임 정보, 블록의 보간 정보, 필터링 정보 등 다양한 정보를 부호화할 수 있다.
- [0062] 엔트로피 부호화부(165)에서는 재정렬부(160)에서 입력된 부호화 단위의 계수 값을 엔트로피 부호화할 수 있다.
- [0063] 역양자화부(140) 및 역변환부(145)에서는 양자화부(135)에서 양자화된 값들을 역양자화하고 변환부(130)에서 변환된 값들을 역변환한다. 역양자화부(140) 및 역변환부(145)에서 생성된 잔차 값(Residual)은 예측부(120, 125)에 포함된 움직임 추정부, 움직임 보상부 및 인트라 예측부를 통해서 예측된 예측 단위와 합쳐져 복원 블록(Reconstructed Block)을 생성할 수 있다.
- [0064] 필터부(150)는 디블록킹 필터, 오프셋 보정부, ALF(Adaptive Loop Filter)중 적어도 하나를 포함할 수 있다.
- [0065] 디블록킹 필터는 복원된 픽처에서 블록 간의 경계로 인해 생긴 블록 왜곡을 제거할 수 있다. 디블록킹을 수행할지 여부를 판단하기 위해 블록에 포함된 몇 개의 열 또는 행에 포함된 픽셀을 기초로 현재 블록에 디블록킹 필터 적용할지 여부를 판단할 수 있다. 블록에 디블록킹 필터를 적용하는 경우 필요한 디블록킹 필터링 강도에 따라 강한 필터(Strong Filter) 또는 약한 필터(Weak Filter)를 적용할 수 있다. 또한 디블록킹 필터를 적용함에 있어 수직 필터링 및 수평 필터링 수행시 수평 방향 필터링 및 수직 방향 필터링이 병행 처리되도록 할 수 있다.
- [0066] 오프셋 보정부는 디블록킹을 수행한 영상에 대해 픽셀 단위로 원본 영상과의 오프셋을 보정할 수 있다. 특정 픽처에 대한 오프셋 보정을 수행하기 위해 영상에 포함된 픽셀을 일정한 수의 영역으로 구분한 후 오프셋을 수행할 영역을 결정하고 해당 영역에 오프셋을 적용하는 방법 또는 각 픽셀의 에지 정보를 고려하여 오프셋을 적용하는 방법을 사용할 수 있다.
- [0067] ALF(Adaptive Loop Filtering)는 필터링한 복원 영상과 원래의 영상을 비교한 값을 기초로 수행될 수 있다. 영상에 포함된 픽셀을 소정의 그룹으로 나눈 후 해당 그룹에 적용될 하나의 필터를 결정하여 그룹마다 차별적으로 필터링을 수행할 수 있다. ALF를 적용할지 여부에 관련된 정보는 휘도 신호는 부호화 단위(Coding Unit, CU) 별로 전송될 수 있고, 각각의 블록에 따라 적용될 ALF 필터의 모양 및 필터 계수는 달라질 수 있다. 또한, 적용 대상 블록의 특성에 상관없이 동일한 형태(고정된 형태)의 ALF 필터가 적용될 수도 있다.
- [0068] 메모리(155)는 필터부(150)를 통해 산출된 복원 블록 또는 픽처를 저장할 수 있고, 저장된 복원 블록 또는 픽처는 인트라 예측을 수행 시 예측부(120, 125)에 제공될 수 있다.
- [0070] 도 2는 본 발명의 일실시예에 따른 영상 복호화 장치를 나타낸 블록도이다.
- [0071] 도 2를 참조하면, 영상 복호화기(200)는 엔트로피 복호화부(210), 재정렬부(215), 역양자화부(220), 역변환부(225), 예측부(230, 235), 필터부(240), 메모리(245)가 포함될 수 있다.
- [0072] 영상 부호화기에서 영상 비트스트림이 입력된 경우, 입력된 비트스트림은 영상 부호화기와 반대의 절차로 복호화될 수 있다.
- [0073] 엔트로피 복호화부(210)는 영상 부호화기의 엔트로피 부호화부에서 엔트로피 부호화를 수행한 것과 반대의 절차로 엔트로피 복호화를 수행할 수 있다. 예를 들어, 영상 부호화기에서 수행된 방법에 대응하여 지수 골롬(Exponential Golomb), CAVLC(Context-Adaptive Variable Length Coding), CABAC(Context-Adaptive Binary Arithmetic Coding)과 같은 다양한 방법이 적용될 수 있다.
- [0074] 엔트로피 복호화부(210)에서는 부호화기에서 수행된 인트라 예측 및 인트라 예측에 관련된 정보를 복호화할 수 있다.
- [0075] 재정렬부(215)는 엔트로피 복호화부(210)에서 엔트로피 복호화된 비트스트림을 부호화부에서 재정렬한 방법을 기초로 재정렬을 수행할 수 있다. 1차원 벡터 형태로 표현된 계수들을 다시 2차원의 블록 형태의 계수로 복원하여 재정렬할 수 있다. 재정렬부(215)에서는 부호화부에서 수행된 계수 스캐닝에 관련된 정보를 제공받고 해당

부호화부에서 수행된 스캐닝 순서에 기초하여 역으로 스캐닝하는 방법을 통해 재정렬을 수행할 수 있다.

- [0076] 역양자화부(220)는 부호화기에서 제공된 양자화 파라미터와 재정렬된 블록의 계수 값을 기초로 역양자화를 수행할 수 있다.
- [0077] 역변환부(225)는 영상 부호화기에서 수행한 양자화 결과에 대해 변환부에서 수행한 변환 즉, DCT, DST, 및 KLT에 대해 역변환 즉, 역 DCT, 역 DST 및 역 KLT를 수행할 수 있다. 역변환은 영상 부호화기에서 결정된 전송 단위를 기초로 수행될 수 있다. 영상 복호화기의 역변환부(225)에서는 예측 방법, 현재 블록의 크기 및 예측 방향 등 복수의 정보에 따라 변환 기법(예를 들어, DCT, DST, KLT)이 선택적으로 수행될 수 있다.
- [0078] 예측부(230, 235)는 엔트로피 복호화부(210)에서 제공된 예측 블록 생성 관련 정보와 메모리(245)에서 제공된 이전에 복호화된 블록 또는 픽처 정보를 기초로 예측 블록을 생성할 수 있다.
- [0079] 전술한 바와 같이 영상 부호화기에서의 동작과 동일하게 인트라 예측을 수행시 예측 단위의 크기와 변환 단위의 크기가 동일할 경우, 예측 단위의 좌측에 존재하는 픽셀, 좌측 상단에 존재하는 픽셀, 상단에 존재하는 픽셀을 기초로 예측 단위에 대한 인트라 예측을 수행하지만, 인트라 예측을 수행시 예측 단위의 크기와 변환 단위의 크기가 상이할 경우, 변환 단위를 기초로 한 참조 픽셀을 이용하여 인트라 예측을 수행할 수 있다. 또한, 최소 부호화 단위에 대해서만 $N \times N$ 분할을 사용하는 인트라 예측을 사용할 수도 있다.
- [0080] 예측부(230, 235)는 예측 단위 판별부, 인터 예측부 및 인트라 예측부를 포함할 수 있다. 예측 단위 판별부는 엔트로피 복호화부(210)에서 입력되는 예측 단위 정보, 인트라 예측 방법의 예측 모드 정보, 인터 예측 방법의 모션 예측 관련 정보 등 다양한 정보를 입력 받고 현재 부호화 단위에서 예측 단위를 구분하고, 예측 단위가 인터 예측을 수행하는지 아니면 인트라 예측을 수행하는지 여부를 판별할 수 있다. 인터 예측부(230)는 영상 부호화기에서 제공된 현재 예측 단위의 인터 예측에 필요한 정보를 이용해 현재 예측 단위가 포함된 현재 픽처의 이전 픽처 또는 이후 픽처 중 적어도 하나의 픽처에 포함된 정보를 기초로 현재 예측 단위에 대한 인터 예측을 수행할 수 있다. 또는, 현재 예측 단위가 포함된 현재 픽처 내에서 기-복원된 일부 영역의 정보를 기초로 인터 예측을 수행할 수도 있다.
- [0081] 인터 예측을 수행하기 위해 부호화 단위를 기준으로 해당 부호화 단위에 포함된 예측 단위의 모션 예측 방법이 스킵 모드(Skip Mode), 머지 모드(Merge 모드), AMVP 모드(AMVP Mode), 인트라 블록 카피 모드 중 어떠한 방법인지 여부를 판단할 수 있다.
- [0082] 인트라 예측부(235)는 현재 픽처 내의 화소 정보를 기초로 예측 블록을 생성할 수 있다. 예측 단위가 인트라 예측을 수행한 예측 단위인 경우, 영상 부호화기에서 제공된 예측 단위의 인트라 예측 모드 정보를 기초로 인트라 예측을 수행할 수 있다. 인트라 예측부(235)에는 AIS(Adaptive Intra Smoothing) 필터, 참조 화소 보간부, DC 필터를 포함할 수 있다. AIS 필터는 현재 블록의 참조 화소에 필터링을 수행하는 부분으로써 현재 예측 단위의 예측 모드에 따라 필터의 적용 여부를 결정하여 적용할 수 있다. 영상 부호화기에서 제공된 예측 단위의 예측 모드 및 AIS 필터 정보를 이용하여 현재 블록의 참조 화소에 AIS 필터링을 수행할 수 있다. 현재 블록의 예측 모드가 AIS 필터링을 수행하지 않는 모드일 경우, AIS 필터는 적용되지 않을 수 있다.
- [0083] 참조 화소 보간부는 예측 단위의 예측 모드가 참조 화소를 보간한 화소 값을 기초로 인트라 예측을 수행하는 예측 단위일 경우, 참조 화소를 보간하여 정수 값 이하의 화소 단위의 참조 화소를 생성할 수 있다. 현재 예측 단위의 예측 모드가 참조 화소를 보간하지 않고 예측 블록을 생성하는 예측 모드일 경우 참조 화소는 보간되지 않을 수 있다. DC 필터는 현재 블록의 예측 모드가 DC 모드일 경우 필터링을 통해서 예측 블록을 생성할 수 있다.
- [0084] 복원된 블록 또는 픽처는 필터부(240)로 제공될 수 있다. 필터부(240)는 디블록킹 필터, 오프셋 보정부, ALF를 포함할 수 있다.
- [0085] 영상 부호화기로부터 해당 블록 또는 픽처에 디블록킹 필터를 적용하였는지 여부에 대한 정보 및 디블록킹 필터를 적용하였을 경우, 강한 필터를 적용하였는지 또는 약한 필터를 적용하였는지에 대한 정보를 제공받을 수 있다. 영상 복호화기의 디블록킹 필터에서는 영상 부호화기에서 제공된 디블록킹 필터 관련 정보를 제공받고 영상 복호화기에서 해당 블록에 대한 디블록킹 필터링을 수행할 수 있다.
- [0086] 오프셋 보정부는 부호화시 영상에 적용된 오프셋 보정의 종류 및 오프셋 값 정보 등을 기초로 복원된 영상에 오프셋 보정을 수행할 수 있다.
- [0087] ALF는 부호화기로부터 제공된 ALF 적용 여부 정보, ALF 계수 정보 등을 기초로 부호화 단위에 적용될 수 있다.

이러한 ALF 정보는 특정한 파라미터 셋에 포함되어 제공될 수 있다.

- [0088] 메모리(245)는 복원된 픽처 또는 블록을 저장하여 참조 픽처 또는 참조 블록으로 사용할 수 있도록 할 수 있고 또한 복원된 픽처를 출력부로 제공할 수 있다.
- [0089] 전술한 바와 같이 이하, 본 발명의 실시예에서는 설명의 편의상 코딩 유닛(Coding Unit)을 부호화 단위라는 용어로 사용하지만, 부호화 뿐만 아니라 복호화를 수행하는 단위가 될 수도 있다.
- [0091] 비디오 캡션은 비디오 클립을 자연어로 자동 요약하여 시각과 언어 사이의 격차를 줄이기 위해 연구되고 있다. 비디오 검색, 휴머노이드 상호 작용 및 시각 장애인 지원과 같은 광범위한 응용 프로그램에 비디오 캡션이 사용된다. 일반적으로 비디오 캡션은 인코딩 단계의 시각적 텍스트 추출과 디코딩 단계의 텍스트 생성이라는 두 가지 주요 단계로 구성된다.
- [0092] 시각적 특징 추출은 컨볼루션 신경망(Convolution Neural Network, CNN)으로 수행된다. 그리고 CNN 특징(feature)은 장단기 기억(Long Short-Term Memory, LSTM) 또는 게이트 반복 단위(Gated Recurrent Unit, GRU)와 관련된 캡션을 생성하는 데 사용된다. UniVL(Unified Video and Language Pre-Training Model), COOT(Cooperative Hierarchical Transformer), 또는 TRL(teacher-recommended learning)이 포함된 ORG(Object Relational Graph)과 같은 딥 러닝 네트워크 아키텍처의 개선에 따라, 비디오 캡션 성능이 더욱 향상될 수 있다. 또한, 비디오 품질의 향상에 따라, 비디오 캡션 성능이 더욱 향상될 수 있다.
- [0093] 컴퓨터 비전(Computer Vision, CV) 성능은 비디오 품질에 영향을 받는다. 비디오 부호화에 의하여 제한된 대역폭과 스토리지에 기초하여 비디오의 데이터 크기가 감소한다. 예를 들어, AVC, HEVC 및 VCC와 같은 비디오 압축 표준은 비디오 내의 모든 프레임을 손실 방식으로 압축하는 양자화 프로세스를 수행한다. 상기 양자화 프로세스에서 양자화 파라미터는 압축이 수행되는 정도를 나타낸다. 높은 양자화 파라미터에 의하여 낮은 양자화 파라미터보다 데이터가 더 압축될 수 있다. 그러나 이러한 손실 압축은 블로킹 및 링잉 아티팩트와 같은 다양한 오류를 발생한다.
- [0094] 최근 영상의 해상도가 높아짐에 따라 데이터 크기를 크게 줄이기 위한 영상 압축에 전처리(preprocessing)로 다운샘플링(down-sampling)이 자주 적용된다. 이 경우 부호화 뿐만 아니라, 다운샘플링으로 인해 품질 저하가 발생할 수 있다. 따라서 본 개시에서 부호화 및 다운샘플링된 비디오에 초해상도(Super Resolution, SR)를 적용하여 효율적인 비디오 캡션을 위한 품질 향상 방법이 제안된다.
- [0095] 상기 SR의 목표는 저해상도 이미지를 고해상도 이미지로 향상시키는 것이다. SR은 감시 및 보안, 의료 이미징, 비디오 동작 인식, 이미지 처리, 이미지 분류, 물체 감지와 같은 다양한 실제 응용 프로그램에 적용될 수 있다. 또한 SR은 얼굴 정렬, 얼굴 파싱, 얼굴 인식, 얼굴 재구성 및 감정 감지에 적용될 수 있다. 본 개시는 화질이 알려지지 않은 저해상도 이미지를 비디오 캡션의 입력으로 하여 압축과 다운샘플링이 비디오 캡션에 미치는 영향을 평가하는 방법을 제공한다. 또한, 본 개시는 효율적인 비디오 캡션을 위해 SR을 이용한 화질 향상 방법을 제안한다.
- [0097] 도 3은 일 실시 예에 따른 비디오 캡션 방법의 프레임워크(300)의 일 실시 예를 도시한다.
- [0098] 품질 분류 네트워크(304)는 비디오 프레임(302)을 저, 중, 고품질 그룹으로 분류한다. 그리고 분류된 비디오 프레임(302)은 저, 중, 고품질 SR 네트워크들(306, 308, 310) 중 하나에 기초하여 품질과 해상도가 향상될 수 있다. 이 때, 상기 SR 네트워크들(306, 308, 310)은 각 품질 그룹에 속하는 이미지로 훈련되어 있다. 마지막으로 향상된 프레임은 비디오 캡션 네트워크(312)에서 사용된다.
- [0099] 블로킹 및 링잉 아티팩트와 같은 다양한 오류는 비디오 압축 및 통신 시스템으로 인해 이미지 품질에 영향을 미친다. 이러한 오류에 의해 왜곡된 저화질 이미지를 효율적으로 분류하기 위하여 여러 딥러닝 기법이 적용될 수 있다. 예를 들어, CNN을 사용하여 가우스 노이즈 및 압축 아티팩트와 같은 이미지 왜곡이 분류될 수 있다. 또한, 왜곡 유형을 분석한 후, 왜곡된 이미지의 품질을 향상시키기 위한 노이즈 제거 방법이 사용될 수 있다. 또는 평균 의견 점수(mean opinion score, MOS)를 사용하여 이미지를 5가지 클래스로 분류하는 DeepBIQ 네트워크가 사용될 수 있다.
- [0100] 구체적으로, 상기 SR 네트워크들(306, 308, 310)은 심층 잔차 네트워크(deep residual network, ResNet), 초

심층 컨볼루션 네트워크 (very deep convolutional network, VGG), GoogLeNet, Inception-v4 및 컨볼루션 3D 네트워크 (convolutional 3D network, C3D)와 같은 사전 훈련된 CNN이다. 그리고 상기 SR 네트워크들 (306, 308, 310)의 출력은 비디오 캡션 네트워크(312)의 입력으로 사용된다.

[0101] 딥 러닝은 SR 성능을 크게 향상시킬 수 있다. 예를 들어, CNN을 이용하여 SRCNN(SR Convolutional Neural Network)에 의하여 SR 성능이 크게 향상될 수 있다. 글로벌 잔차 학습(Global Residual Learning)은 SRCNN보다 정확도가 높은 20개의 Layer로 구성된 VDSR(very deep convolutional network)에서 적용될 수 있다. 그리고 로컬 잔차 학습(Local Residual Learning)은 SR 심층 잔차 네트워크(SRResNet)에 적용될 수 있다. 상기의 학습 기법은 잔차 연결(residual connection)로 인한 그라디언트 소실 문제를 완화할 수 있다. 많은 단순화된 잔차 모듈들을 스택킹(stack)함으로써, EDSR(Enhanced Deep SR Network) 및 MDSR(Multi-Scale Deep SR Model)의 초해상도화 효율이 향상된다. 저레벨 특징(feature)과 고레벨 특징 간의 신호 전파를 향상시키기 위해, 두 계층 간의 밀집 연결(dense connection)이 RDN(Residual Dense Network)과 ESRGAN(Enhanced SR Generative Adversarial Network)에서 사용된다.

[0102] 성능 향상을 위해 HAN(Holistic Attention Network), RBAN(Residual Block Attention Network), SAN(Second-order Attention Network), PASSRnet (Parallax-Attention Stereo SR Network), RCAN (residual channel attention network)와 같은 많은 네트워크에서 Attention 메커니즘이 사용될 수 있다. 상기 Attention 메커니즘은, 이미지의 모든 지역을 동등하게 고려하기 보다는, 중요한 부분만 선별적으로 집중하여 가치 있는 정보를 찾고 관련 없는 정보는 무시함으로써, SR의 성능을 향상시킨다. Attention 메커니즘을 수행하는 Attention 모듈은 채널 Attention, 공간 Attention 및 픽셀 Attention 로 분류될 수 있다.

[0103] RNN(Recurrent Neural Network)으로 인하여, 비디오 캡션 성능이 크게 개선될 수 있다. 특히, RNN은 많은 캡션 네트워크에서 사용되는 인코더-디코더 아키텍처에서 시퀀스 학습에 매우 유용하다. 일반적으로 인코더는 입력 비디오 내의 프레임에서 시각적 특징을 추출하고 디코더는 출력으로 문장을 생성한다. 인코딩 프로세스에서, 평균 풀링, 시간 인코딩 및 LSTM과 같은 입력에서 시각적 콘텐츠를 효율적으로 나타내기 위해 다양한 변환이 고려될 수 있다. 디코더는 LSTM 또는 GRU를 언어 모델로 사용할 수 있다. 모델 정확도를 향상시키기 위해 beam search, various modalities, attention mechanism 및 semantic attributes 과 같은 많은 기술이 함께 사용될 수 있다.

[0104] S2VT(sequence-to-sequence based video to text network)가 비디오 캡션에 사용될 수 있다. S2VT는 매개변수 공유를 위해 인코더와 디코더 모두에서 단일 LSTM을 사용한다. LSTM 매개변수는 역전파를 기반으로 예측된 단어의 로그 가능성을 최대화하기 위해 계속 업데이트된다. 디코더의 입력은 인코더의 출력인 은닉 표현 벡터이다. 디코더에 입력되는 상기 은닉 표현 벡터와 디코더에서 이전에 복호화된 단어에 기초하여, 다음 단어가 생성될 수 있다. 이러한 구조에 기초하여 입력 영상 내 각 프레임의 시간적 의미를 학습한다.

[0106] 도 4는 화질 측정 딥러닝 네트워크(400)의 일 실시 예를 나타낸다.

[0107] 압축 아티팩트는 일반적으로 이미지 품질을 많이 저하시킨다. 따라서 높은 압축률에 의해 발생하는 심각한 아티팩트가 비디오 캡션에 부정적인 방향으로 크게 영향을 미칠 가능성이 높다. 비디오의 압축률을 나타내는 양자화 파라미터는 네트워크 또는 저장 조건에 크게 의존한다. 일반적으로 비디오 캡션에 사용될 영상의 화질에 대한 정보는 알 수 없기 때문에, 도 4의 화질 측정 딥러닝 네트워크(400)에 따른 화질 분류가 비디오 캡션 이전에 적용될 수 있다. 따라서 입력 영상 내의 모든 프레임의 화질을 저레벨, 중레벨, 또는 고레벨로 분류한다. 또 다른 예로, 입력 영상의 프레임의 화질은 2개 이상의 레벨로 구분될 수 있다.

[0108] 일 실시 예에 따르면, 화질 측정 딥러닝 네트워크(400)는 원본 입력 프레임의 특성을 유지하기 위해 전처리 없이 64×64 RGB 패치를 직접 사용한다. 화질 측정 딥러닝 네트워크의 출력은 클래스 수와 같은 차원을 가진 벡터이다. 예를 들어, 프레임을 3개의 그룹(저레벨, 중레벨, 또는 고레벨)으로 분류하므로 출력 벡터의 차원은 3이 된다. 상기 화질 측정 딥러닝 네트워크는 VGG16을 기반으로 설계될 수 있다. 상기 VGG16에는 컨볼루션 레이어(conv), 평균 풀링 레이어(avgpool), 평면 레이어 및 완전 연결 레이어(FC)가 포함된다.

[0109] 도 4의 화질 측정 딥러닝 네트워크(400)에서, 64x64 크기 RGB 패치를 입력 받아 Convolutional Layer(conv)를 통해 압축 에러가 포함된 영상의 특징점이 추출된다. 그리고 Fully-Connected Layer(FC)를 통해 압축 열화의 정도에 따라 영상의 화질 레벨이 결정된다.

[0110] 예를 들어, 도 4와 같이 2conv3-16, avgpool2, 2conv3-32, avgpool2, 2conv3-64, avgpool2, 2conv3-128,

avgpool2, flatten, 2FC-256, FC-3이 순서대로 사용될 수 있다. 2conv3-16은 3×3 커널의 필터가 16개 있는 2개의 convolutional 레이어를 나타낸다. 그리고 소프트맥스 활성화 함수를 사용하는 마지막 FC 레이어를 제외한 모든 레이어에서 ReLU (Rectified Linear Unit) 활성화 함수가 사용될 수 있다. 컨볼루션 레이어를 통과하는 특징(feature)의 공간 차원을 유지하기 위해, 경계면의 특징에 대하여 제로 패딩이 사용될 수 있다. 또한 과적합 문제를 해결하기 위해 2FC-256에 드롭아웃 정규화(dropout regularization)가 적용될 수 있다. 출력 벡터의 값은 입력 프레임의 카테고리 확률 분포(categorical distribution)를 의미한다. 그 결과, 각 프레임은 저품질, 중품질, 고품질 그룹 중 확률이 가장 높은 그룹에 속하게 된다. 분류된 그룹에 따라, 세 가지 다른 SR 네트워크 중 하나가 다음 단계에서 적용된다.

[0111] 일 실시 예에 따르면, 영상의 화질에 관한 정보는, 복호화단에서 화질 측정 딥러닝 네트워크에 의하여 결정되는 대신, 비트스트림에 포함된 정보로부터 도출될 수 있다. 따라서, 영상의 화질에 관한 정보는 부호화단에서 결정되어 복호화단에서 영상의 화질 분류에 사용될 수 있다.

[0112] 또는, 영상의 화질에 관한 정보는, 비트스트림에 포함된 양자화 파라미터 정보, 잔차 신호, 움직임 정보 등에 기초하여 복호화단에서 도출될 수 있다. 또는 영상의 화질에 관한 정보는, 영상의 화질을 직접적으로 나타내는 화질 정보일 수 있다.

[0113] 일 실시 예에 따르면, 화질 측정 딥러닝 네트워크(400)에서 입력되는 영상은 RGB 형식의 텍스처 영상 뿐만 아니라 YUV, YCbCr 포맷의 텍스처 영상일 수 있다. 또는 화질 측정 딥러닝 네트워크(400)에서 입력되는 영상은 텍스처 영상, 깊이 영상 및 특징 맵 중 적어도 하나를 포함할 수 있다.

[0114] 일 실시 예에 따르면, 화질 측정 딥러닝 네트워크(400)에서 결정된 화질 정보에 따라, 영상의 업샘플링 뿐만 아니라, 영상의 후처리 여부 및 머신 비전 수행 여부 등이 결정될 수 있다.

[0116] 도 5는 초해상도 딥러닝 네트워크(500)의 일 실시 예를 도시한다.

[0117] 초해상도 딥러닝 네트워크(500)는 얇은 특징 추출(shallow feature extraction), 확장 잔차 블록(Extended Residual Blocks, ERB)에 의한 깊은 특징 추출 (deep feature extraction), 공간 채널 주의(spatial-channel attention), 및 블록 재구성의 네 부분으로 구성될 수 있다.

[0118] 도 5에서 I_{LR} 와 I_{SR} 는 제안된 REBN(Residual Extended Block Network)의 입출력 이미지를 나타낸다. 첫 번째 얇은 특징(Shallow Feature) F_{-1} 은 I_{SR} 에서 수학적 식 1에 따라 다음과 같이 추출된다.

[0119] [수학적 식 1]

$$\mathbf{F}_{-1} = H_{SF1}(\mathbf{I}_{LR})$$

[0120]

[0121] 여기서 $H_{SF1}(\cdot)$ 는 첫 번째 컨볼루션 연산을 나타낸다. 비슷한 방법으로 두 번째 컨볼루션 레이어에서 두 번째 얇은 특징 F_0 을 수학적 식 2에 따라 다음과 같이 추출한다.

[0122] [수학적 식 2]

$$\mathbf{F}_0 = H_{SF2}(\mathbf{F}_{-1})$$

[0123]

[0124] 여기서 $H_{SF2}(\cdot)$ 는 두 번째 컨볼루션 연산을 나타낸다. REBN 네트워크는 도 5(점선 상자)와 같이 ERB를 기반으로 하는 하위 네트워크를 통해 F_0 에서 특징을 추출한다. 상기 네트워크를 심층 훈련하기 위해, ERB는 빠르고 향상된 수렴을 위한 ReLU 활성화 기능(activation function)과 로컬 스킵 연결(local skip connection)이 있는 두 개의 컨볼루션 계층을 사용할 수 있다. 중간 특징 F_d 는 번째 ERB의 출력이며, 수학적 식 3에 따라 다음과 같이 계산된다.

[0125] [수학적 식 3]

$$\mathbf{F}_d = H_{ERB}^d(\mathbf{F}_{d-1}), \quad d = 1, 2, \dots, D$$

[0126]

[0127] 여기서 $H_{ERB}^d(.)$ 는 번째 ERB의 동작을 나타내고, D는 네트워크에 있는 ERB의 수이다. ERB의 집합으로 이러한 특징을 추출한 후 합성된 특징 F_{CF} 은 수학적 식 4에 따라 다음과 같이 앞의 모든 특징 F_d 을 합성하여 나타낼 수 있다.

[0128] [수학적 식 4]

$$F_{CF} = H_{CF}(Concatenate(F_1, \dots, F_D))$$

[0129]

[0130] 여기서 $H_{CF}(.)$ 는 복합 함수를 나타낸다. 또한 Attention 모듈은 네트워크 성능을 더욱 향상시키는 데 사용된다. 상기 모듈에는 spatial-channel attention 모듈을 가질 수 있다. spatial attention는 입력 특징의 공간 간 관계를 학습한다. 반면에 channel attention는 입력 특징의 채널 간 관계를 학습한다. attention 모듈의 출력은 F_{SC} 로 표현되며, F_{-1} 과 F_{SC} 에서 수학적 식 5에 따라 다음과 같이 전역 스킵 연결을 통해 딥 피쳐 F_{DF} 가 생성될 수 있다.

[0131] [수학적 식 5]

$$F_{DF} = F_{-1} + F_{SC}$$

[0132]

[0133] 마지막으로 업샘플링된 이미지 I_{SR} 는 다음과 같이 업샘플링 및 재구성 레이어에 F_{DF} 를 전달하여 생성된다.

[0134] [수학적 식 6]

$$I_{SR} = H_{RE}(H_{UP}(F_{DF}))$$

[0135]

[0136] 여기서 $H_{UP}(.)$ 및 $H_{RE}(.)$ 는 각각 업샘플링 프로세스와 재구성을 위한 컨볼루션 연산을 나타낸다. 결과적으로 I_{SR} 은 I_{LR} 대신 비디오 캡션 네트워크에서 사용된다.

[0138] 도 6은 비디오 프레임의 시퀀스를 입력으로 사용하고 단어 시퀀스를 출력으로 생성하는 인코더-디코더 구조(600)의 일 실시 예를 도시한다.

[0139] 인코더-디코더 구조(600)는 조건부 확률 $p(Y|X)$ 을 학습한다. 여기서 $Y = (y_1, \dots, y_m)$ 및 $X = (x_1, \dots, x_n)$ 은 각각 단어와 프레임 특징(feature)을 나타낸다. 첫 번째 인코딩 계층은 CNN(602)에서 추출한 특징에서 숨겨진 표현 h_t 를 학습한다. 두 번째 인코딩 계층(604)은 null padded 태그 <pad>와 연결된 h_t 를 디코딩 단계로 전달한다. h_t 를 단어 시퀀스로 변환하는 것은 두 번째 디코딩 LSTM 계층(606)이 beginning-of-sentence <bos> 태그를 수신할 때 시작된다. 출력 단어 Y는 end-of-sentence <eos> 태그를 통해 완성된다. LSTM 파라미터는 예측 단어의 로그 가능성을 최대화하도록 훈련된다.

[0140] 일반적으로 LSTM은 정보를 장기간 기억하여 장기 종속성을 효율적으로 해결하기 위해 사용된다. 시간 인스턴스에서 LSTM 장치는 입력 게이트 i_t , 망각 게이트 f_t 및 출력 게이트 o_t 를 통해 은닉 상태 h_t 및 메모리 셀 상태 c_t 를 생성한다. i_t , f_t , o_t 는 각각 셀 상태에서 어떤 정보를 저장, 삭제, 및 추출해야 하는지 나타낸다. 각 게이트는 다음과 같이 수학적 식 7에 기초하여 수행된다.

[0141] [수학식 7]

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \odot \phi(c_t) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ g_t &= \sigma(W_g x_t + U_g h_{t-1} + b_g) \end{aligned}$$

[0142]

[0143] 여기서 σ , ϕ 및 $\odot$ 는 각각 시그모이드 함수, 쌍곡선 탄젠트 함수 및 요소별 곱을 나타낸다. g_t 셀 입력 활성화 특징을 의미한다. W , U , b 는 네트워크에 의해 학습된다.

[0144] 도 6과 같이 인코딩 LSTM 계층은 입력된 특징 X 을 읽어서 은닉 상태 h_t 를 계산한다. 마지막으로, 출력 단어 Y 에 대한 분포는 다음과 같이 수학식 8에 기초하여 계산된다.

[0145] [수학식 8]

$$p(Y|X) = \prod_{t=1}^m p(y_t | h_{n+t-1}, y_{t-1})$$

[0146]

[0147] 여기서 $p(y_t | h_{n+t-1})$ 의 확률 분포는 어휘 패키지의 모든 단어에 대한 softmax 함수에 의해 계산된다.

[0149] 이하, 비디오 캡션에 사용될 수 있는 특징 맵의 생성 및 부호화에 관련된 VCM(Video Coding for Machine)이 설명된다. VCM은 이미지로부터 기계가 인식 가능한 특징을 추출하여, 그 특징으로 구성된 특징 맵 또는 특징 이미지를 부호화 및 복호화하는 기술을 의미한다. 산업 감시 시스템, 지능형 운송 등에 사용될 수 있다. VCM을 구현함에 있어서, 복호화단에 VCM에 관련된 기능이 집중될 경우, 복호화단은 데이터를 계산하고 분석하는 데 많은 부담을 가질 수 있다. 따라서, 에너지 소비 및 계산 지연 측면에서 VCM의 알고리즘의 최적화가 필요하다.

[0150] 일 실시 예에 따르면, VCM을 위한 심층 모델이 분할되어, 부호화단과 복호화단 간에 계산이 분산될 수 있다. 부호화단과 복호화단 간에 전송되는 데이터는 시각적 신호일 수도 있고 특징 신호일 수도 있다. 머신 비전 애플리케이션에서는 벤더 간에 서로 다른 작업을 수행하기 위해 비디오 텍스처보다 중요한 정보나 특징 맵을 공유하는 것이 더 효율적일 수 있다. 이러한 분산 컴퓨팅 시나리오 외에도 대부분의 감시 또는 운송 응용 프로그램의 동의 없는 비디오 캡처 및 전송에 대한 개인 정보 보호 문제가 증가함에 따라 사람이 인지할 수 있는 시각적 신호를 직접 코딩하는 대신 특징 신호를 압축하는 기술이 필요하다.

[0152] 도 7 및 8은 VCM을 구현하기 위한 2개의 파이프 라인이 도시한다.

[0153] 도 7은 VCM을 구현하기 위한 파이프 라인의 제1 실시 예(700)를 나타낸다.

[0154] 도 7에 따르면, 기계 분석(Machine Analysis)은 비디오 디코더에 의하여 복호화된 이미지에 대하여 수행된다. 도 7의 제1 실시 예(700)에 따르면 송신자(710)는 시각적 이미지를 부호화하여 전송하고, 수신자(720)가 부호화된 시각적 이미지를 복호화한 후, 복호화된 이미지를 분석하여 특징 맵을 출력한다. 따라서 송신자(720)가 기계 분석을 위한 전체 딥 러닝 네트워크를 운용하여야 한다. 따라서, 제1 실시 예(700)에 따르면 수신자가 딥 러닝 네트워크의 모든 파라미터를 계산하고 저장하기 위하여 필요한 고성능 컴퓨팅 환경을 구축할 부담이 있다.

[0155] 도 8은 VCM을 구현하기 위한 파이프 라인의 제2 실시 예(800)를 나타낸다.

[0156] 도 8에 따르면, 송신자(810)가 기계 분석을 위한 일부 기능을 보유하고 있다. 따라서 송신자(810)는 기계 분석, 특징 변환, 및 비디오 부호화에 따른 부호화된 특징 맵을 생성하고, 상기 부호화된 특징 맵을 포함한 비트스트림

림을 출력한다.

- [0157] 수신자(820)는 부호화된 특징 맵을 수신하고, 상기 부호화된 특징 맵을 복호화 및 특징 역변환하여 특징 맵을 획득한다. 그리고 특징 맵을 머신 비전 작업에 활용하므로, 수신자(820)가 기계 분석을 위한 전체 딥 러닝 네트워크를 보유할 필요가 없다. 즉, 수신자(820)가 전체 딥러닝 네트워크를 운영하지 않기 때문에 딥러닝 네트워크의 모든 파라미터를 계산하고 저장하는 데 필요한 고성능 컴퓨팅 환경을 구축할 필요가 없다. 또한 머신 비전 태스크의 결과를 서버가 전송하는 구조에 비해 다양한 태스크의 활용을 기대할 수 있다.
- [0158] 제2 실시 예(800)에 따르면, 수신자(820)의 딥 러닝 네트워크의 동작을 위한 부하를 줄이는 장점이 있다. 하지만, 하나의 이미지 대신 전송해야 하는 특징 맵의 채널 수가 매우 많고 하나의 특징 맵의 크기가 작지 않으므로, 원본 영상 데이터에 비해 전송할 데이터의 크기가 너무 커지는 문제가 있을 수 있다. 이러한 전송 데이터의 크기 문제를 극복하기 위해 비디오 코덱에서 높은 QP 값을 적용하여 압축을 수행하면, 데이터의 왜곡이 지나치게 증가할 수 있다.
- [0159] 이러한 한계를 해결하기 위해, 비디오 압축 특성을 고려한 특징 맵 변환 방법으로 유사도가 높은 대표적인 특징 맵을 재정렬하는 방법, 특징 맵의 크기를 조정된 다음 변환된 특징 맵을 압축된 단일 이미지로 압축하는 방법, 및 개별 특징 맵 기반 비디오 프레임으로 재배열하는 방법이 적용될 수 있다. 그러나 특징 맵의 크기를 조정하고 특징 맵 채널을 재정렬하는 것과 같은 단일 이미지 기반 접근 방식은 압축률 향상에 한계가 있다. 반면, 비디오 기반 접근 방식은 상당한 압축률 향상이 있지만, 양자화 파라미터가 증가할수록 화질 저하가 증가하는 문제점이 있다. 따라서, 화질 저하를 방지하기 위하여, 부호화율을 높이면서 특징 맵의 압축 손실을 최소화할 수 있는 방법이 필요하다.
- [0160] 따라서, 본 개시는 공간적, 시간적 중복성 증가를 고려하여 다양한 수의 특징 맵을 갖는 특징 프레임을 생성하는 특징 프레임 기반 특징 맵 압축 방법을 제공한다.
- [0162] 도 9는 복수의 특징 맵을 포함하는 특징 프레임을 구성하는 방법의 일 실시 예를 설명한다.
- [0163] 본 개시의 특징 프레임은 다양한 특징으로 생성된 것을 의미한다. 도 9에 따르면 16x16개의 특징 맵은 하나의 이미지(900)로 표현될 수 있다. 하나의 이미지(900)는 분할되어 복수의 프레임으로 표현될 수 있다. 예를 들어, 하나의 프레임에 8x8 개의 특징 맵이 포함된 경우, 상기 이미지(900)는 4개의 프레임(902)으로 분할될 수 있다. 마찬가지로, 하나의 프레임에 4x4, 2x2, 1x1개의 특징 맵이 포함된 경우, 각각 상기 이미지(900)는 각각 16개의 프레임(904), 64개의 프레임(906), 256개의 프레임(908)로 분할될 수 있다.
- [0165] 도 10은 특징 프레임에 포함된 특징 맵들을 재정렬하기 위한, 공간 우선 순위 할당 방법 및 시간 우선 순위 할당 방법을 설명한다.
- [0166] 특징 맵의 채널 간의 그래디언트를 기반으로 특징 맵이 재정렬될 수 있다. 특징 맵을 재정렬하기 위하여, 공간 우선 순위 할당 및 시간 우선 순위 할당 중 적어도 하나가 적용될 수 있다.
- [0167] 그림 10의 (a)와 같이 공간 우선 할당은 하나의 특징 프레임을 유사도에 따라 특징 맵을 재정렬(정렬)하여 프레임 내 중복성을 높이는 방법이다. 도 10의 (a)에 따르면, 특징 프레임의 복수의 특징 맵이 유사도에 따라 우선 순위가 재결정될 수 있다. 그리고 재결정된 우선순위에 따라, 특징 프레임 내에서 특징 맵이 배열된다. 예를 들어, 제1 특징 프레임(1000)에 1~16번의 우선순위를 가진 특징 맵이 래스터 스캔 순서대로 배열된다. 그리고, 제2 특징 프레임(1002)에 17~32번의 우선순위를 가진 특징 맵이 래스터 스캔 순서대로 배열된다. 제1 특징 프레임(1000) 및 제2 특징 프레임(1002)의 특징 맵 배열 순서와 마찬가지로 33~256번의 우선순위를 가진 특징 맵은 순서대로 나머지 특징 프레임들에 배열된다. 따라서, 인접한 특징 맵들 간의 유사도가 높기 때문에, 화면 내 예측의 효율성이 증가될 수 있다.
- [0168] 그림 10의 (b)와 같이 시간 우선 할당은 유사도에 따라 재정렬(정렬)된 특징 맵을 프레임의 동일한 위치에 순차적으로 할당하여 프레임 간 중복성을 높이는 방법이다. 도 10의 (b)에 따르면, 특징 프레임의 복수의 특징 맵이 유사도에 따라 우선 순위가 재결정될 수 있다. 그리고 재결정된 우선순위에 따라, 연속된 특징 프레임의 동일 위치에 특징 맵이 순차적으로 배열된다. 예를 들어, 1~3번의 우선순위를 가진 특징 맵이 순차적으로 제1 특징 프레임(1052), 제2 특징 프레임(1054), 제3 특징 프레임(1056)의 좌상측 모서리 위치에 배열된다. 그리고 4~6번의 우선순위를 가진 특징 맵이 순차적으로 제1 특징 프레임(1052), 제2 특징 프레임(1054), 제3 특징 프레임(1056)의

6)의 1~3번의 우선순위를 가진 특징 맵이 위치의 우측에 배열된다. 나머지 특징 맵들도 우선순위에 따라, 동일 위치에서 제1 특징 프레임(1052), 제2 특징 프레임(1054), 제3 특징 프레임(1056)의 순서대로 배열될 수 있다. 이 때 배열되는 위치의 순서는 래스터 스캔 순서일 수 있다. 따라서, 동일 위치의 특징 맵들 간의 유사도가 높기 때문에, 화면 간 예측의 효율성이 증가될 수 있다.

[0169] 도 10의 (a) 및 (b)의 배열 순서가 래스터 스캔 순서에 기초하고 있으나, 실시 예에 따라, 수직 스캔, 수평 스캔, 대각선 스캔의 순서에 기초하여 배열될 수 있다. 또한 상기 언급된 스캔 순서의 역방향으로 배열될 수도 있다.

[0171] 도 11은 초해상도(SR) 기술에 따른 특징 맵 압축 방법의 일 실시 예(1100)를 도시한다.

[0172] 특징 추출기(1104)는 원본 이미지 (1102)로부터 고해상도 특징 맵(1106)을 생성할 수 있다. 그리고 고해상도 특징 맵(1106)에 대하여 양자화 및 특징 맵 재배열과 같은 전처리가 수행될 수 있다. 그리고 전처리된 특징 맵에 대하여 노이즈 제거 SR을 사용하여 특징 맵의 아티팩트가 제거될 수 있다.

[0173] 다운샘플링에 의하여 고해상도 특징 맵(1106)은 저해상도 특징 맵(1108)으로 변환될 수 있다. 이 때 고해상도 특징 맵(1106)의 다운샘플링은 bicubic 보간법에 의하여 수행될 수 있다. 일 실시 예에 따르면, 저해상도 특징 맵(1108)에 대하여 양자화 및 특징 맵 재배열과 같은 전처리가 수행될 수 있다.

[0174] 저해상도 특징 맵(1108)은 영상 압축을 위한 인코더(1110)에 의하여 압축될 수 있다. 그리고 압축된 특징 맵은 디코더(1112)에 의하여 복호화됨으로써, 복원된 저해상도 특징 맵(1114)이 생성될 수 있다. 이 때, 복원된 저해상도 특징 맵(1114)은 고해상도 특징 맵(1106)의 다운샘플링 및 영상 압축에 의한 아티팩트를 가질 수 있다.

[0175] 따라서, SR 기술을 이용하여 복원된 저해상도 특징 맵(1114)을 업샘플링함으로써, 상기 언급된 아티팩트가 제거된 복원된 고해상도 특징 맵(1118)이 생성될 수 있다. 그리고 복원된 고해상도 특징 맵(1118)에 기초하여 컴퓨터 비전이 수행될 수 있다. 이 때 SR 기술을 구현하기 위한 SR 네트워크(1116)는, SRCNN 등의 딥러닝 네트워크일 수 있다. 상기 SR 네트워크(1116)는 특징 맵의 압축 및 공간 업샘플링으로 인한 아티팩트를 제거하기 위하여 학습될 수 있다. SR 네트워크(1116)의 학습은, 복원된 저해상도 특징 맵(1114)로부터 고해상도 특징 맵(1108)과 매우 유사한 복원된 고해상도 특징 맵(1118)을 획득하는 것을 목표로 한다.

[0177] 도 12는 영상 처리 방법의 일 실시예를 도시한다.

[0178] 단계 1202에서, 영상의 화질을 나타내는 화질 정보가 결정된다.

[0179] 일 실시 예에 따르면, 상기 화질 정보는 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 결정될 수 있다.

[0180] 일 실시 예에 따르면, 상기 화질 정보는 화질 측정 딥러닝 네트워크에 결정될 수 있다. 상기 화질 정보는, 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나를 화질 측정 딥러닝 네트워크에 입력함으로써, 화질 측정 딥러닝 네트워크로부터 출력될 수 있다.

[0181] 일 실시 예에 따르면, 화질 정보는 화질 측정 딥러닝 네트워크로부터 출력된 출력 벡터에 따라 결정되며, 출력 벡터는 복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타낼 수 있다.

[0182] 일 실시 예에 따르면, 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정될 수 있다.

[0183] 단계 1204에서, 단계 1202에서 결정된 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크가 결정된다.

[0184] 단계 1206에서, 단계 1204에서 결정된 초해상도 딥러닝 네트워크에 기초하여, 영상이 업샘플링된다.

[0185] 일 실시 예에 따르면, 영상 처리 방법은, 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 단계를 더 포함할 수 있다. 상기 플래그 정보가 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 단계 1202 내지 1206이 수행될 수 있다. 상기 플래그 정보는 시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득될 수 있다.

- [0186] 일 실시 예에 따르면, 상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 포함하고, 상기 화질 정보는 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 결정될 수 있다.
- [0187] 일 실시 예에 따르면, 영상 처리 방법은, 업샘플링된 영상에 대한 비디오 캡션을 수행하는 단계를 더 포함할 수 있다.
- [0188] 상기 영상 처리 방법은 컴퓨터 프로그램으로 구현될 수 있다. 또한, 상기 영상 처리 방법이 구현된 컴퓨터 프로그램은 기록 매체에 저장되어 유통될 수 있다. 또는 상기 영상 처리 방법이 구현된 컴퓨터 프로그램은 온라인으로 유통될 수 있다.
- [0190] 도 13은 영상 처리 장치(1300)의 일 실시예를 도시한다.
- [0191] 영상 처리 장치(1300)는 복수의 초해상도 딥러닝 네트워크(1302), 및 영상의 화질을 나타내는 화질 정보를 결정하고, 상기 화질 정보에 따라, 영상을 업샘플링하기 위한 초해상도 딥러닝 네트워크를 결정하는 화질 결정 유닛(1304)을 포함한다. 그리고 영상 처리 장치(1300)에서 복수의 초해상도 딥러닝 네트워크 중 화질 정보에 따라 선택된 초해상도 딥러닝 네트워크에 의하여 상기 영상이 업샘플링된다.
- [0192] 일 실시 예에 따르면, 화질 결정 유닛(1304)은 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 결정할 수 있다.
- [0193] 일 실시 예에 따르면, 화질 결정 유닛(1304)은 화질 측정 딥러닝 네트워크를 포함할 수 있다. 그리고 화질 측정 딥러닝 네트워크에 의하여 상기 화질 정보가 결정될 수 있다. 화질 측정 딥러닝 네트워크는 영상의 양자화 파라미터 정보, 잔차 신호 정보, 움직임 벡터 정보 중 적어도 하나에 기초하여 상기 화질 정보를 출력할 수 있다.
- [0194] 일 실시 예에 따르면, 화질 측정 딥러닝 네트워크는, 복수의 초해상도 딥러닝 네트워크에 대한 각각의 확률 값을 나타내는 출력 벡터를 출력하고, 출력 벡터에 따라 상기 화질 정보를 결정할 수 있다.
- [0195] 일 실시 예에 따르면, 화질 정보는 비트스트림으로부터 획득된 영상의 화질을 나타내는 화질 인덱스 정보로부터 결정될 수 있다.
- [0196] 일 실시 예에 따르면, 영상 처리 장치는, 영상을 초해상도 딥러닝 네트워크를 이용하여 업샘플링하는지 여부를 나타내는 플래그 정보를 비트스트림으로부터 획득하는 영상 복호화 유닛(1308)을 더 포함할 수 있다. 그리고 플래그 정보가 영상이 초해상도 딥러닝 네트워크를 이용하여 업샘플링됨을 나타낼 때, 화질 결정 유닛, 및 복수의 초해상도 딥러닝 네트워크의 기능이 수행될 수 있다.
- [0197] 상기 플래그 정보는 시퀀스, 프레임, 슬라이스, 및 블록 중 적어도 하나의 단위에 대하여 획득될 수 있다.
- [0198] 일 실시 예에 따르면, 상기 영상은 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 포함하고, 상기 화질 결정 유닛은 상기 텍스처 영상, 깊이 영상, 및 특징 맵 중 적어도 하나를 평가함으로써 상기 화질 정보를 결정할 수 있다.
- [0199] 일 실시 예에 따르면, 상기 영상 처리 장치는, 상기 업샘플링된 영상에 대한 비디오 캡션을 수행하는 비디오 캡션 유닛(1306)을 더 포함할 수 있다.
- [0201] 상술한 실시예들에서, 방법들은 일련의 단계 또는 유닛으로서 순서도를 기초로 설명되고 있으나, 본 발명은 단계들의 순서에 한정되는 것은 아니며, 어떤 단계는 상술한 바와 다른 단계와 다른 순서로 또는 동시에 발생할 수 있다. 또한, 당해 기술 분야에서 통상의 지식을 가진 자라면 순서도에 나타난 단계들이 배타적이지 않고, 다른 단계가 포함되거나, 순서도의 하나 또는 그 이상의 단계가 본 발명의 범위에 영향을 미치지 않고 삭제될 수 있음을 이해할 수 있을 것이다.
- [0202] 상술한 실시예는 다양한 양태의 예시들을 포함한다. 다양한 양태들을 나타내기 위한 모든 가능한 조합을 기술할 수는 없지만, 해당 기술 분야의 통상의 지식을 가진 자는 다른 조합이 가능함을 인식할 수 있을 것이다. 따라서, 본 발명은 이하의 특허청구범위 내에 속하는 모든 다른 교체, 수정 및 변경을 포함한다고 할 것이다.
- [0203] 이상 설명된 본 발명에 따른 실시예들은 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로

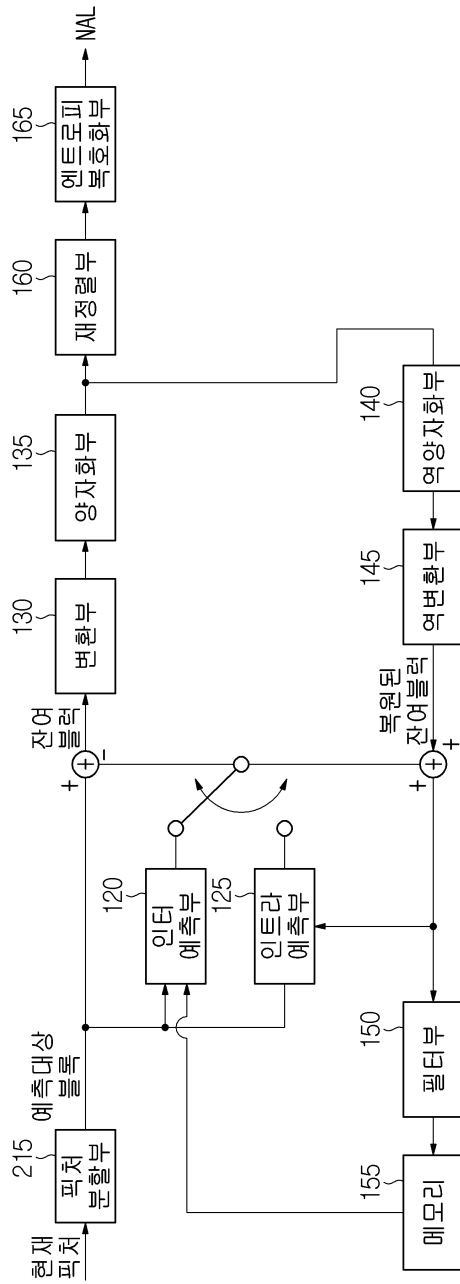
그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0204] 이상에서 본 발명이 구체적인 구성요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나, 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명이 상기 실시예들에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형을 꾀할 수 있다.

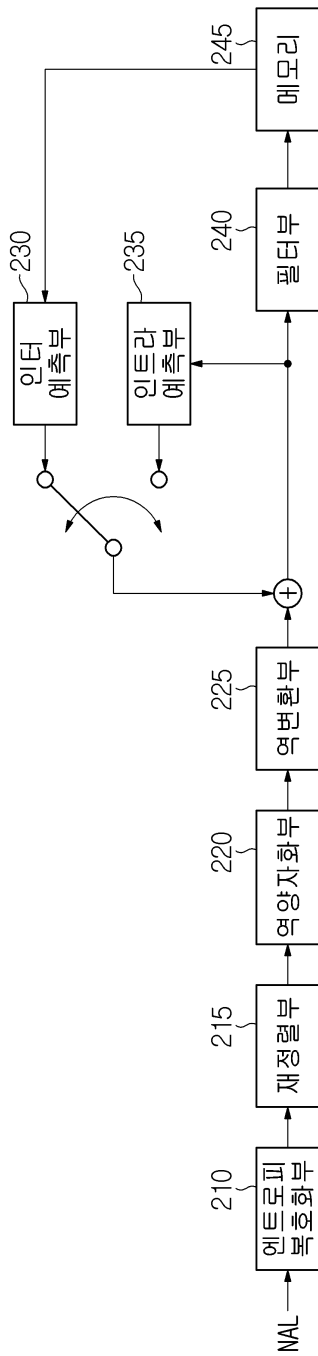
[0205] 따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등하게 또는 등가적으로 변형된 모든 것들은 본 발명의 사상의 범주에 속한다고 할 것이다.

도면

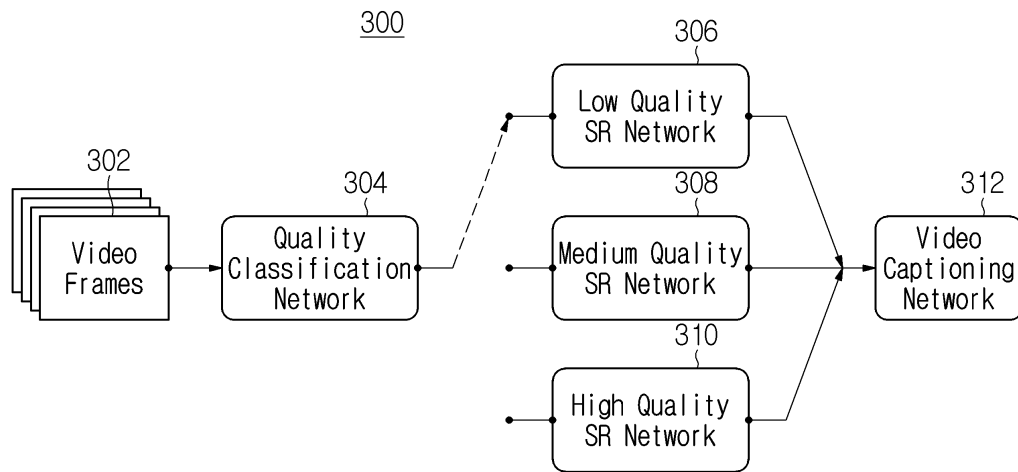
도면1



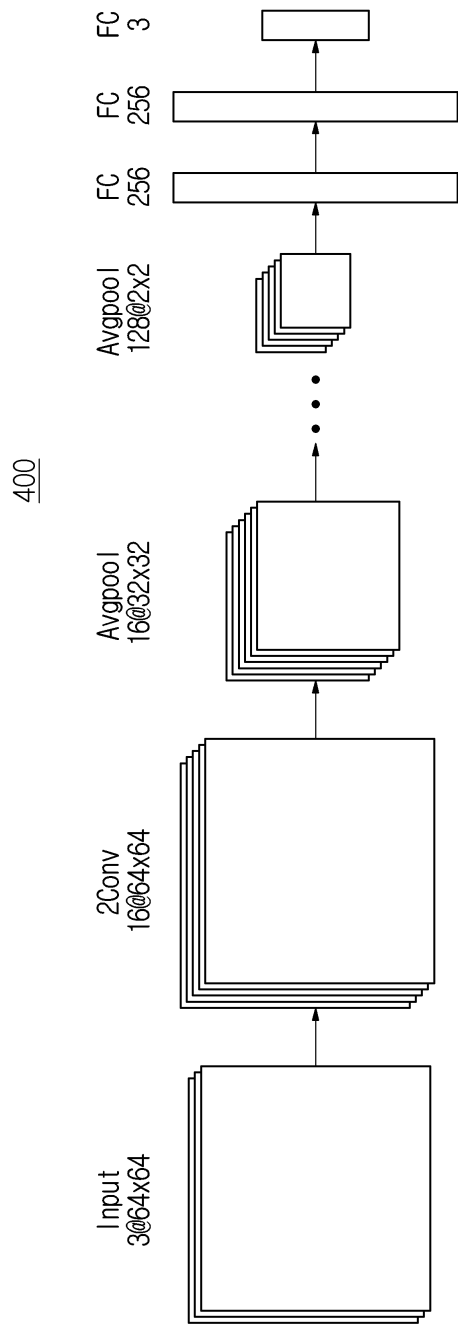
도면2



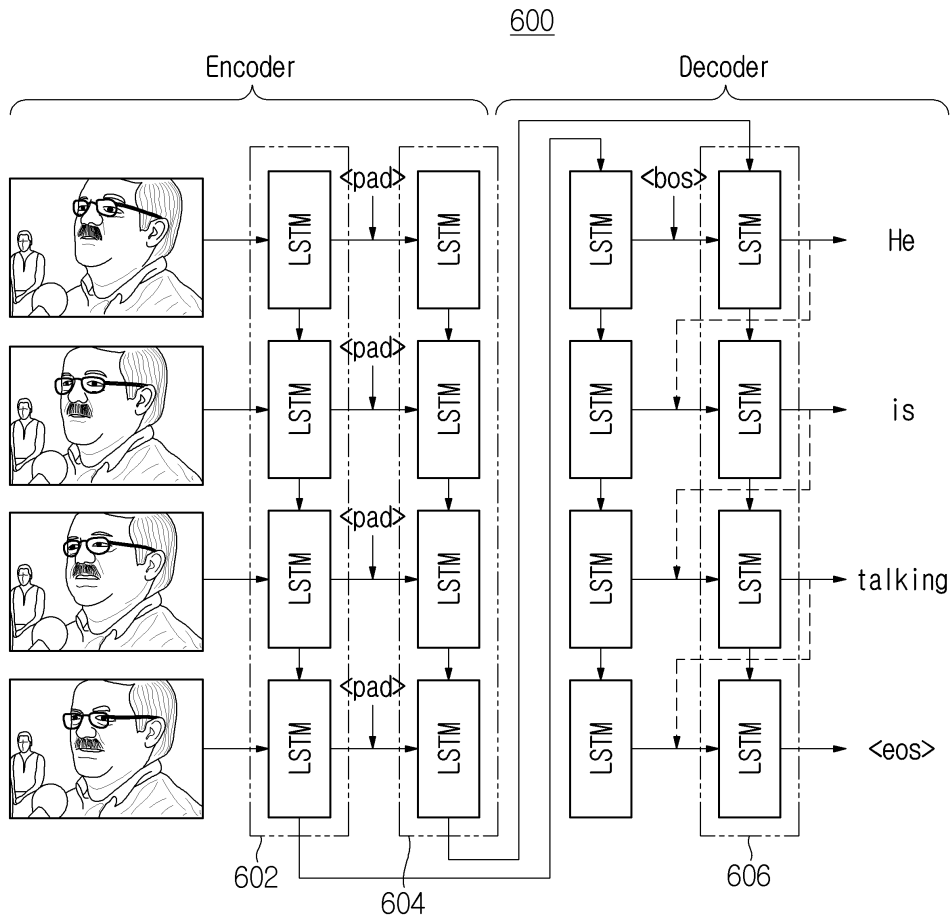
도면3



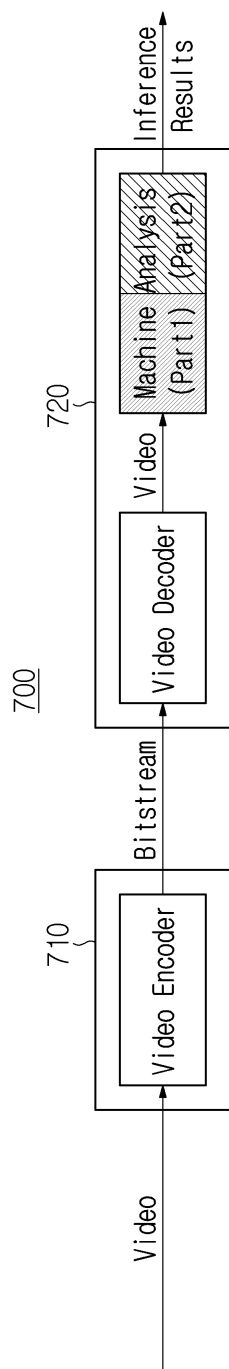
도면4



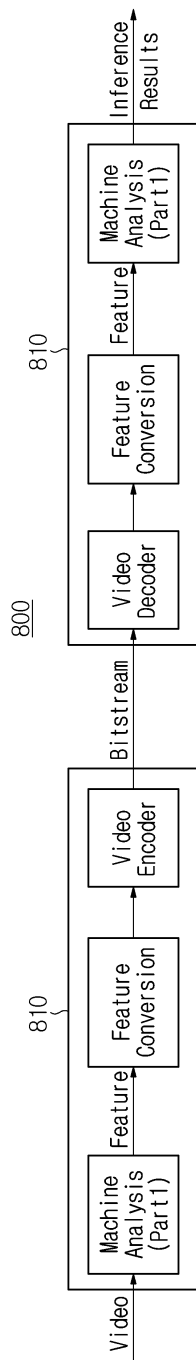
도면6



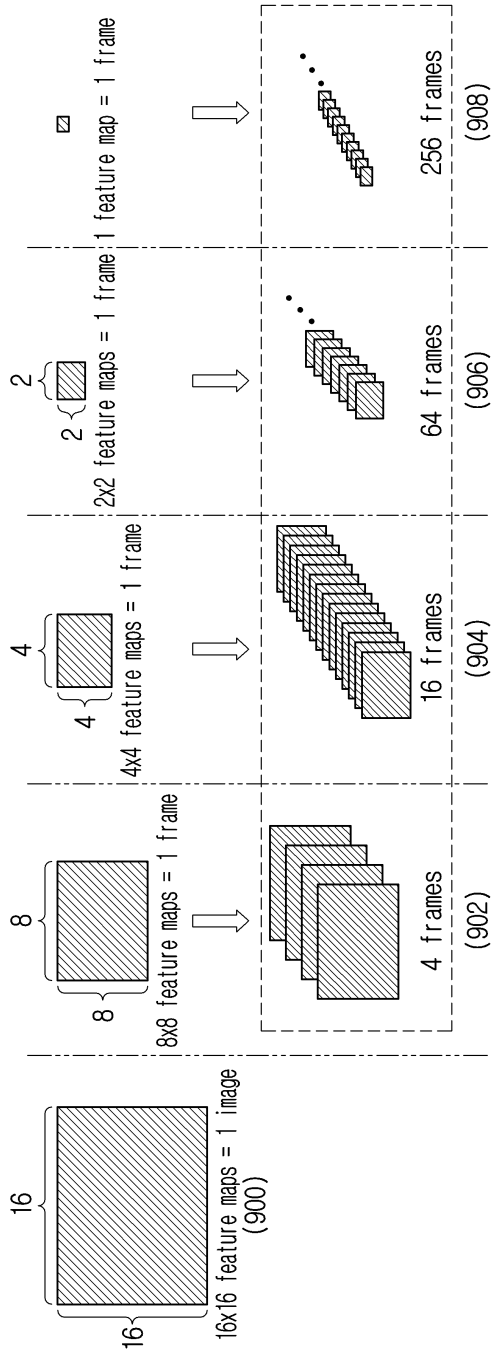
도면7



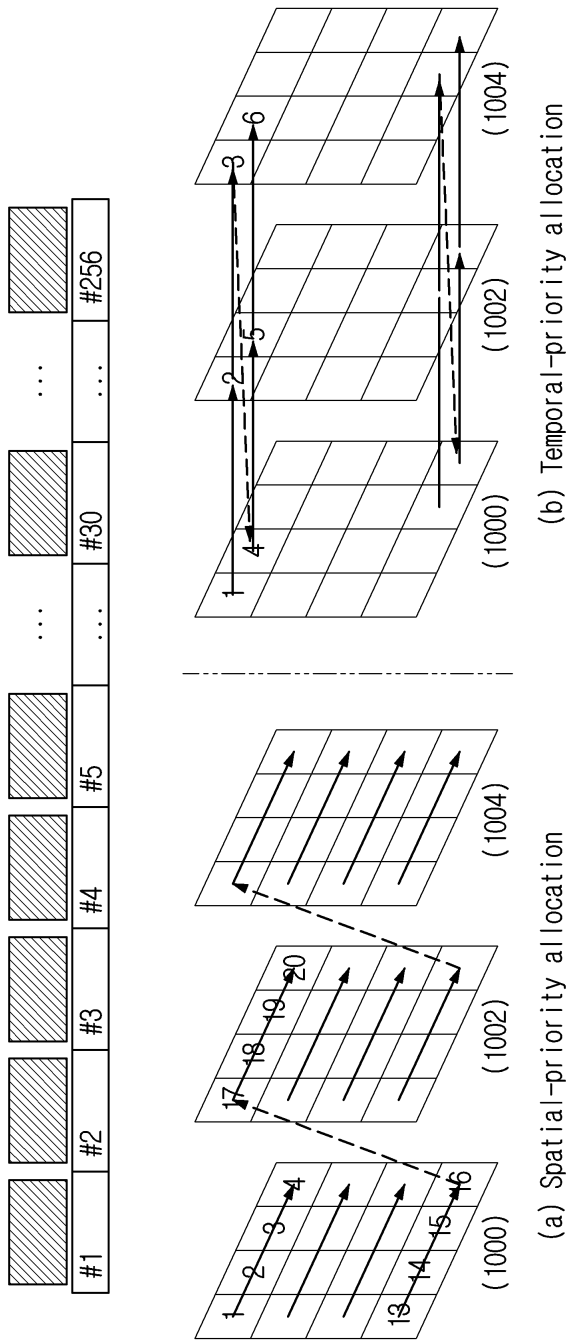
도면8



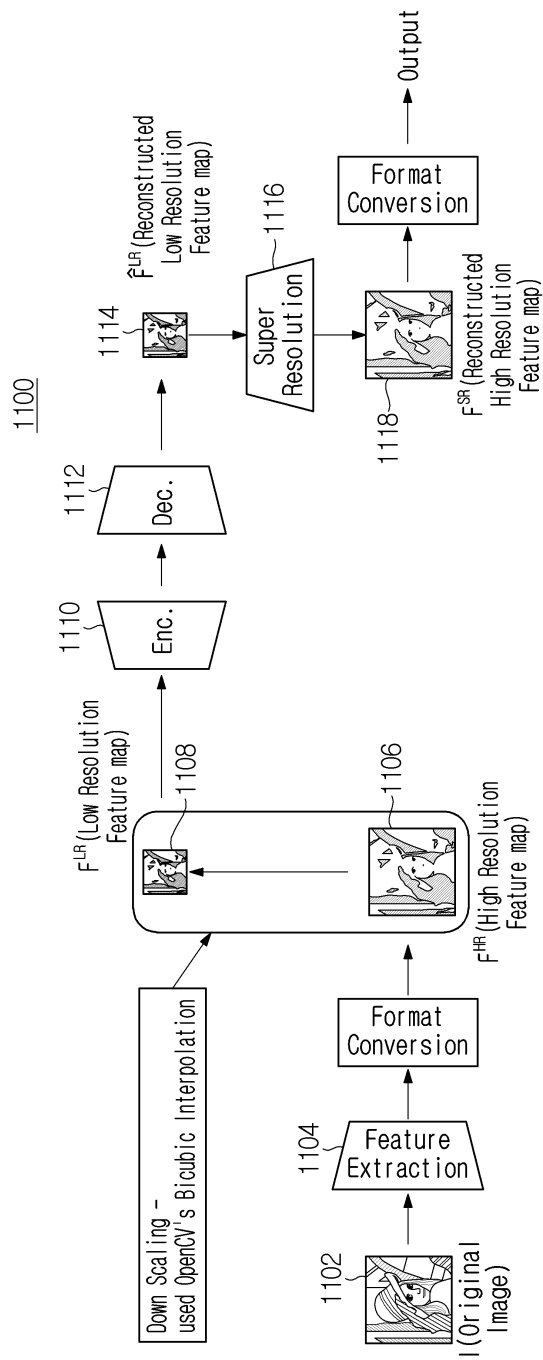
도면9



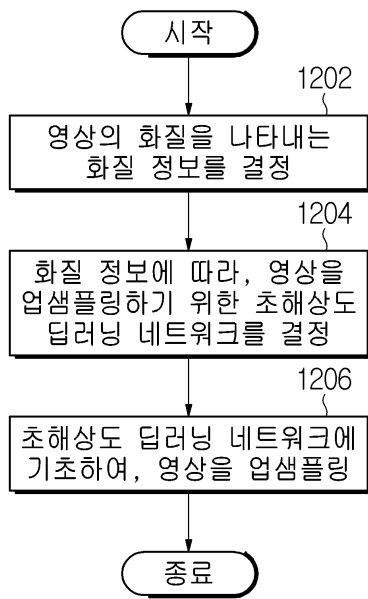
도면10



도면11



도면12



도면13

