(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0079697
(43) 공개일자 2020년07월06일

(51) 국제특허분류(Int. Cl.)
H04N 21/4402 (2011.01) *G06N 3/08* (2006.01)
H04N 21/4728 (2011.01)
(52) CPC특허분류
H04N 21/440272 (2013.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2018-0169105
(22) 출원일자 2018년12월26일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
임형준
경기도 수원시 영통구 삼성로 129(매탄동)
강석주
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(뒷면에 계속)
(74) 대리인
정홍식, 김태헌

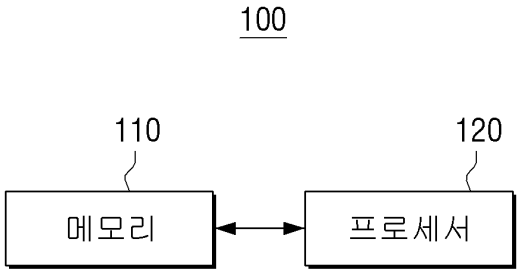
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **영상 처리 장치 및 그 영상 처리 방법**

(57) 요 약

영상 처리 장치가 개시된다. 영상 처리 장치는, 적어도 하나의 명령어를 저장하는 메모리 및, 메모리와 전기적으로 연결된 프로세서를 포함하고, 프로세서는, 명령어를 실행함으로써, 입력 영상 프레임을 학습 네트워크 모델에 적용하여 획득된 관심 영역에 대한 정보에 기초하여 출력 영상 프레임을 획득하며, 학습 네트워크 모델은, 입력 영상 프레임에서 관심 영역에 대한 정보를 획득하도록 학습된 모델일 수 있다.

대 표 도 - 도2



(52) CPC특허분류
H04N 21/4728 (2013.01)
(72) 발명자
이승준
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
문영수
경기도 수원시 영통구 삼성로 129(매탄동)
이시영
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
조성인
경상북도 경산시 진량읍 대구대로 201 (대구대학교)

명 세 서

청구범위

청구항 1

적어도 하나의 명령어를 저장하는 메모리; 및

상기 메모리와 전기적으로 연결된 프로세서;를 포함하고,

상기 프로세서는,

상기 명령어를 실행함으로써, 입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하고, 상기 획득된 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 출력 영상 프레임을 획득하며,

상기 학습 네트워크 모델은,

상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델인, 영상 처리 장치.

청구항 2

제1항에 있어서,

상기 관심 영역에 대한 정보는,

상기 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함하는, 영상 처리 장치.

청구항 3

제1항에 있어서,

상기 관심 영역에 대한 정보는,

상기 관심 영역에 대응되는 상기 입력 영상 프레임의 수평 방향 위치 정보 및 크기 정보를 포함하는 1차원 정보인, 영상 처리 장치.

청구항 4

제3항에 있어서,

상기 프로세서는,

상기 1차원 정보의 크기 또는 상기 1차원 정보에 대응되는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하는, 영상 처리 장치.

청구항 5

제1항에 있어서,

상기 프로세서는,

상기 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 상기 입력 영상 프레임을 리타겟팅하며,

상기 제2 변환 가중치는,

상기 입력 영상 프레임에 대한 해상도 정보 및 상기 출력 영상 프레임에 대한 해상도 정보에 기초하여 획득되는, 영상 처리 장치.

청구항 6

제5항에 있어서,

상기 프로세서는,

상기 관심 영역에 대응되는 픽셀에 대해서는 상기 제1 변환 가중치를 적용하여 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 상기 제2 변환 가중치를 적용하여 확대 스케일링하는, 영상 처리 장치.

청구항 7

제1항에 있어서,

상기 프로세서는,

제1 입력 영상 프레임 및 상기 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 상기 관심 영역에 대한 정보를 획득하는, 영상 처리 장치.

청구항 8

제7항에 있어서,

상기 프로세서는,

상기 제1 입력 영상 프레임 및 상기 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득하고,

상기 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득하는, 영상 처리 장치.

청구항 9

제1항에 있어서,

상기 학습 네트워크 모델은,

상기 입력 영상 프레임에 포함된 객체 관련 정보를 검출하는 특징 검출부 및 상기 관심 영역의 크기 정보, 위치 정보 및 타입 정보를 획득하는 특징 맵 추출부를 포함하는, 영상 처리 장치.

청구항 10

제1항에 있어서,

상기 학습 네트워크 모델은,

관심 영역에 대한 정보 및 상기 관심 영역에 대한 정보에 대응되는 복수의 이미지를 이용하여 상기 학습 네트워크 모델에 포함된 뉴럴 네트워크의 가중치를 학습하는, 영상 처리 장치.

청구항 11

제1항에 있어서,

디스플레이;를 더 포함하며,

상기 프로세서는,

상기 획득된 출력 영상 프레임을 디스플레이하도록 상기 디스플레이를 제어하는, 영상 처리 장치.

청구항 12

영상 처리 장치의 영상 처리 방법에 있어서,

입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하는 단계; 및

상기 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 출력 영상 프레임을 획득하는 단계;를 포함하며,

상기 학습 네트워크 모델은,

상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델인, 영상 처리 방법.

청구항 13

제12항에 있어서,

상기 관심 영역에 대한 정보는,

상기 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함하는, 영상 처리 방법.

청구항 14

제12항에 있어서,

상기 관심 영역에 대한 정보는,

상기 관심 영역에 대응되는 상기 입력 영상 프레임의 수평 방향 위치 정보 및 크기 정보를 포함하는 1차원 정보인, 영상 처리 방법.

청구항 15

제14항에 있어서,

상기 출력 영상 프레임을 획득하는 단계는,

상기 1차원 정보의 크기 또는 상기 1차원 정보에 대응되는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하는, 영상 처리 방법.

청구항 16

제12항에 있어서,

상기 출력 영상 프레임을 획득하는 단계는,

상기 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 상기 입력 영상 프레임을 리타겟팅하며,

상기 제2 변환 가중치는,

상기 입력 영상 프레임에 대한 해상도 정보 및 상기 출력 영상 프레임에 대한 해상도 정보에 기초하여 획득되는, 영상 처리 방법.

청구항 17

제16항에 있어서,

상기 출력 영상 프레임을 획득하는 단계는,

상기 관심 영역에 대응되는 픽셀에 대해서는 상기 제1 변환 가중치를 적용하여 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 상기 제2 변환 가중치를 적용하여 확대 스케일링하는, 영상 처리 방법.

청구항 18

제12항에 있어서,

상기 관심 영역에 대한 정보를 획득하는 단계는,

제1 입력 영상 프레임 및 상기 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 상기 관심 영역에 대한 정보를 획득하는, 영상 처리 방법.

청구항 19

제18항에 있어서,

상기 관심 영역에 대한 정보를 획득하는 단계는,

상기 제1 입력 영상 프레임 및 상기 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득하며,

상기 출력 영상 프레임을 획득하는 단계는,

상기 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득하는, 영상 처리 방법.

청구항 20

영상 처리 장치의 프로세서에 의해 실행되는 경우 상기 영상 처리 장치가 동작을 수행하도록 하는 컴퓨터 명령을 저장하는 비일시적 컴퓨터 판독 가능 매체에 있어서, 상기 동작은,

입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하는 단계; 및

상기 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 상기 출력 영상 프레임을 획득하는 단계;를 포함하며,

상기 학습 네트워크 모델은,

상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델인, 비일시적 컴퓨터 판독 가능 매체.

발명의 설명

기 술 분 야

[0001] 본 개시는 영상 처리 장치 및 그 영상 처리 방법에 관한 것으로, 더욱 상세하게는 입력 영상을 리타겟팅 처리하여 출력 영상을 획득하는 영상 처리 장치 및 그 영상 처리 방법에 관한 것이다.

배 경 기 술

[0002] 전자 기술의 발달에 힘입어 다양한 유형의 전자기기가 개발 및 보급되고 있다. 특히, 가정, 사무실, 공공 장소 등 다양한 장소에서 이용되는 디스플레이 장치는 최근 수년 간 지속적으로 발전하고 있다.

[0003] 최근에는 고해상도 영상 서비스, 실시간 스트리밍 서비스에 대한 요구가 크게 증가하고 있다.

[0004] 경우에 따라 입력 영상의 해상도와 출력 해상도가 상이한 경우 입력 영상을 출력 해상도에 맞추기 위한 영상 처리를 적용한다. 다만 입력 영상의 해상도와 출력 해상도의 종횡비(가로세로비율)가 동일한 경우에는 영상 왜곡이 없으나, 종횡비가 상이한 경우 종횡비 조정으로 인해 영상 왜곡이 발생하게 되는 문제점이 있다.

발명의 내용

해결하려는 과제

[0005] 본 개시은 상술한 필요성에 따른 것으로, 본 개시의 목적은, 관심 영역 검출을 통해 영상의 왜곡을 최소화하면서 입력 영상의 종횡비를 조정하여 출력 영상을 획득할 수 있는 영상 처리 장치 및 그 영상 처리 방법을 제공함에 있다.

과제의 해결 수단

[0006] 이상과 같은 목적을 달성하기 위한 본 개시의 일 실시 예에 따른 영상 처리 장치는, 적어도 하나의 명령어를 저장하는 메모리 및, 상기 메모리와 전기적으로 연결된 프로세서를 포함하고, 상기 프로세서는, 상기 명령어를 실행함으로써, 입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하고, 상기 획득된 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 출력 영상 프레임을 획득하며, 상

기 학습 네트워크 모델은, 상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델일 수 있다.

[0007] 여기서, 상기 관심 영역에 대한 정보는, 상기 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함할 수 있다.

[0008] 또한, 상기 관심 영역에 대한 정보는, 상기 관심 영역에 대응되는 상기 입력 영상 프레임의 수평 방향 위치 정보 및 크기 정보를 포함하는 1차원 정보일 수 있다.

[0009] 또한, 상기 프로세서는, 상기 1차원 정보의 크기 또는 상기 1차원 정보에 대응되는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅할 수 있다.

[0010] 또한, 상기 프로세서는, 상기 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 상기 입력 영상 프레임을 리타겟팅하며, 상기 제2 변환 가중치는, 상기 입력 영상 프레임에 대한 해상도 정보 및 상기 출력 영상 프레임에 대한 해상도 정보에 기초하여 획득될 수 있다.

[0011] 또한, 상기 프로세서는, 상기 관심 영역에 대응되는 픽셀에 대해서는 상기 제1 변환 가중치를 적용하여 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 상기 제2 변환 가중치를 적용하여 확대 스케일링할 수 있다.

[0012] 또한, 상기 프로세서는, 제1 입력 영상 프레임 및 상기 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 상기 관심 영역에 대한 정보를 획득할 수 있다.

[0013] 또한, 상기 프로세서는, 상기 제1 입력 영상 프레임 및 상기 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득하고, 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득할 수 있다.

[0014] 또한, 상기 학습 네트워크 모델은, 상기 입력 영상 프레임에 포함된 객체 관련 정보를 검출하는 특징 검출부 및 상기 관심 영역의 크기 정보, 위치 정보 및 타입 정보를 획득하는 특징 맵 추출부를 포함할 수 있다.

[0015] 또한, 상기 학습 네트워크 모델은, 관심 영역에 대한 정보 및 상기 관심 영역에 대한 정보에 대응되는 복수의 이미지를 이용하여 상기 학습 네트워크 모델에 포함된 뉴럴 네트워크의 가중치를 학습할 수 있다.

[0016] 또한, 디스플레이를 더 포함하며, 상기 프로세서는, 상기 획득된 출력 영상 프레임을 디스플레이하도록 상기 디스플레이를 제어할 수 있다.

[0017] 한편, 본 개시의 일 실시 예에 따른 영상 처리 장치의 영상 처리 방법은, 입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하는 단계 및, 상기 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 출력 영상 프레임을 획득하는 단계를 포함하며, 상기 학습 네트워크 모델은, 상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델일 수 있다.

[0018] 여기서, 상기 관심 영역에 대한 정보는, 상기 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함할 수 있다.

[0019] 또한, 상기 관심 영역에 대한 정보는, 상기 관심 영역에 대응되는 상기 입력 영상 프레임의 수평 방향 위치 정보 및 크기 정보를 포함하는 1차원 정보일 수 있다.

[0020] 또한, 상기 출력 영상 프레임을 획득하는 단계는, 상기 1차원 정보의 크기 또는 상기 1차원 정보에 대응되는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅할 수 있다.

[0021] 또한, 상기 출력 영상 프레임을 획득하는 단계는, 상기 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 상기 입력 영상 프레임을 리타겟팅하며, 상기 제2 변환 가중치는, 상기 입력 영상 프레임에 대한 해상도 정보 및 상기 출력 영상 프레임에 대한 해상도 정보에 기초하여 획득될 수 있다.

[0022] 또한, 상기 출력 영상 프레임을 획득하는 단계는, 상기 관심 영역에 대응되는 픽셀에 대해서는 상기 제1 변환 가중치를 적용하여 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 상기 제2 변환 가중치를 적용하여 확대 스케일링할 수 있다.

[0023] 또한, 상기 관심 영역에 대한 정보를 획득하는 단계는, 제1 입력 영상 프레임 및 상기 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 상기 관심 영역에 대한 정보를 획득할 수 있다.

[0024] 또한, 상기 관심 영역에 대한 정보를 획득하는 단계는, 상기 제1 입력 영상 프레임 및 상기 제2 입력 영상 프레임 각각에서 획득된 상기 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득하며, 상기 출력 영상 프레임을 획득하는 단계는, 상기 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 상기 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득할 수 있다.

[0025] 또한, 본 개시의 일 실시 예에 따른 영상 처리 장치의 프로세서에 의해 실행되는 경우 상기 영상 처리 장치가 동작을 수행하도록 하는 컴퓨터 명령을 저장하는 비일시적 컴퓨터 판독 가능 매체에 있어서, 상기 동작은, 입력 영상 프레임을 상기 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득하는 단계 및, 상기 관심 영역에 대한 정보에 기초하여 상기 입력 영상 프레임을 리타겟팅하여 출력 영상 프레임을 획득하는 단계를 포함하며, 상기 학습 네트워크 모델은, 상기 입력 영상 프레임에서 상기 관심 영역에 대한 정보를 획득하도록 학습된 모델일 수 있다.

발명의 효과

[0026] 본 개시의 다양한 실시 예에 따르면, 영상의 주요 영역 왜곡 없이 입력 영상의 종횡비를 조정하여 출력 영상을 획득할 수 있게 된다.

도면의 간단한 설명

[0027] 도 1은 본 개시의 일 실시 예에 따른 영상 처리 장치의 구현 예를 설명하기 위한 도면이다.

도 2는 본 개시의 일 실시 예에 따른 영상 처리 장치의 구성을 나타내는 블럭도이다.

도 3은 본 개시의 일 실시 예에 따른 학습 네트워크 모델을 설명하기 위한 도면이다.

도 4는 본 개시의 일 실시 예에 따른 학습 네트워크 모델을 설명하기 위한 도면이다.

도 5는 본 개시의 일 실시 예에 따른 학습 네트워크 모델을 설명하기 위한 도면이다.

도 6은 본 개시의 일 실시 예에 따른 관심 영역 정보의 예시를 나타내는 도면이다.

도 7은 본 개시의 다른 실시 예에 따른 관심 영역 정보의 예시를 나타내는 도면이다.

도 8a는 본 개시의 일 실시 예에 따른 영상 크기 조정 방법을 설명하기 위한 도면이다.

도 8b는 본 개시의 다른 실시 예에 따른 영상 크기 조정 방법을 설명하기 위한 도면이다.

도 9는 본 개시의 다른 실시 예에 따른 영상 크기 조정 방법을 설명하기 위한 도면이다.

도 10은 본 개시의 다른 실시 예에 따른 관심 영역 정보의 예시를 나타내는 도면이다.

도 11은 본 개시의 다른 실시 예에 따른 영상 처리 장치의 일 구현 예를 나타내는 도면이다.

도 12는 본 개시의 일 실시 예에 따른 영상 처리 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0028] 이하에서는 첨부 도면을 참조하여 본 개시를 상세히 설명한다.

[0029] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 개시에 대해 구체적으로 설명하기로 한다.

[0030] 본 개시의 실시 예에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 개시의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그

용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0031] 본 개시의 실시 예들은 다양한 변환을 가할 수 있고 여러 가지 실시 예를 가질 수 있는바, 특정 실시 예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나 이는 특정한 실시 형태에 대해 범위를 한정하려는 것이 아니며, 개시된 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 실시 예들을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0032] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[0033] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "구성되다" 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.

[0034] A 및 B 중 적어도 하나라는 표현은 "A" 또는 "B" 또는 "A 및 B" 중 어느 하나를 나타내는 것으로 이해되어야 한다.

[0035] 본 개시에서 "모듈" 혹은 "부"는 적어도 하나의 기능이나 동작을 수행하며, 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 "모듈" 혹은 복수의 "부"는 특정한 하드웨어로 구현될 필요가 있는 "모듈" 혹은 "부"를 제외하고는 적어도 하나의 모듈로 일체화되어 적어도 하나의 프로세서(미도시)로 구현될 수 있다.

[0036] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 개시를 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0037] 도 1은 본 개시의 일 실시 예에 따른 영상 처리 장치의 구현 예를 설명하기 위한 도면이다.

[0038] 영상 처리 장치(100)는 도 1에 도시된 바와 같이 TV 또는 set-top box 로 구현될 수 있으나, 이에 한정되는 것은 아니며 스마트 폰, 태블릿 PC, 노트북 PC, HMD(Head mounted Display), NED(Near Eye Display), LFD(large format display), Digital Signage(디지털 간판), DID(Digital Information Display), 비디오 월(video wall), 프로젝터 디스플레이, 카메라 등과 같이 영상 처리 및/또는 디스플레이 기능을 갖춘 장치라면 한정되지 않고 적용 가능하다.

[0039] 영상 처리 장치(100)는 다양한 압축 영상 또는 다양한 해상도의 영상을 수신할 수 있다. 예를 들어, 영상 처리 장치(100)는 MPEG(Moving Picture Experts Group)(예를 들어, MP2, MP4, MP7 등), JPEG(joint photographic coding experts group), AVC(Advanced Video Coding), H.264, H.265, HEVC(High Efficiency Video Codec) 등으로 압축된 형태로 영상을 수신할 수 있다. 또는 영상 처리 장치(100)는 SD(Standard Definition), HD(High Definition), Full HD, Ultra HD 영상 중 어느 하나의 영상을 수신할 수 있다.

[0040] 일 실시 예에 따라 입력되는 영상의 해상도와 출력 해상도가 종횡비가 상이한 경우, 입력 영상의 해상도를 출력 해상도에 맞추기 위한 리타겟팅(retargeting) 처리가 요구된다. 예를 들어, 입력 영상의 해상도는 HD 또는 Full HD 영상이고 출력 해상도는 Ultra HD인 경우 출력 해상도에 맞추어 영상을 리타겟팅(retargeting) 처리하게 된다. 이 경우, 입력 영상의 종횡비(가로세로비율)를 출력 영상의 종횡비에 맞게 조정함에 따라 영상 왜곡이 발생하게 되는 문제점이 있다.

[0041] 이에 따라 본 개시에서는 입력 영상의 해상도와 출력 해상도의 종횡비가 상이한 경우 영상 왜곡을 최소화할 수 있는 다양한 실시 예에 대해 설명하도록 한다.

[0042] 도 2는 본 개시의 일 실시 예에 따른 영상 처리 장치의 구성을 나타내는 블럭도이다.

[0043] 도 2에 따르면, 영상 처리 장치(100)는 메모리(110) 및 프로세서(120)를 포함한다.

[0044] 메모리(110)는 메모리(110)는 프로세서(120)와 전기적으로 연결되며, 본 개시의 다양한 실시 예를 위해 필요한 데이터를 저장할 수 있다. 예를 들어, 메모리(110)는 프로세서(120)에 포함된 롬(ROM)(예를 들어,

EEPROM(electrically erasable programmable read-only memory)), 램(RAM) 등의 내부 메모리로 구현되거나, 프로세서(120)와 별도의 메모리로 구현될 수도 있다. 이 경우, 메모리(110)는 데이터 저장 용도에 따라 영상 처리 장치(100)에 임베디드된 메모리 형태로 구현되거나, 영상 처리 장치(100)에 탈부착이 가능한 메모리 형태로 구현될 수도 있다. 예를 들어, 영상 처리 장치(100)의 구동을 위한 데이터의 경우 영상 처리 장치(100)에 임베디드된 메모리에 저장되고, 영상 처리 장치(100)의 확장 기능을 위한 데이터의 경우 영상 처리 장치(100)에 탈부착이 가능한 메모리에 저장될 수 있다. 한편, 영상 처리 장치(100)에 임베디드된 메모리의 경우 휘발성 메모리(예: DRAM(dynamic RAM), SRAM(static RAM), 또는 SDRAM(synchronous dynamic RAM) 등), 비휘발성 메모리(non-volatile Memory)(예: OTPROM(one time programmable ROM), PROM(programmable ROM), EPROM(erasable and programmable ROM), EEPROM(electrically erasable and programmable ROM), mask ROM, flash ROM, 플래시 메모리(예: NAND flash 또는 NOR flash 등), 하드 드라이브, 또는 솔리드 스테이트 드라이브(solid state drive(SSD)) 중 적어도 하나로 구현되고, 영상 처리 장치(100)에 탈부착이 가능한 메모리의 경우 메모리 카드(예를 들어, CF(compact flash), SD(secure digital), Micro-SD(micro secure digital), Mini-SD(mini secure digital), xD(extreme digital), MMC(multi-media card) 등), USB 포트에 연결가능한 외부 메모리(예를 들어, USB 메모리) 등과 같은 형태로 구현될 수 있다.

[0045] 메모리(110)는 프로세서(120)가 입력 영상 프레임을 학습 네트워크 모델에 적용하여 획득된 관심 영역에 대한 정보에 기초하여 출력 영상 프레임을 획득하도록 제어하는 명령어를 저장한다. 여기서, 학습 네트워크 모델은, 입력 영상 프레임에서 관심 영역에 대한 정보를 획득하도록 학습된 모델이 될 수 있다.

[0046] 일 실시 예에 따라 메모리(110)는 본 개시의 일 실시 예에 따른 학습 네트워크 모델을 저장할 수 있다. 다만, 다른 실시 예에 따르면, 학습 네트워크 모델은 외부 서버 또는 외부 장치 중 적어도 하나에 저장될 수도 있다.

[0047] 프로세서(120)는 메모리(110)와 전기적으로 연결되어 영상 처리 장치(100)의 전반적인 동작을 제어한다.

[0048] 일 실시 예에 따라 프로세서(120)는 디지털 영상 신호를 처리하는 디지털 시그널 프로세서(digital signal processor(DSP), 마이크로 프로세서(microprocessor), TCON(Time controller)으로 구현될 수 있다. 다만, 이에 한정되는 것은 아니며, 중앙처리장치(central processing unit(CPU)), MCU(Micro Controller Unit), MPU(micro processing unit), 컨트롤러(controller), 어플리케이션 프로세서(application processor(AP)), 또는 커뮤니케이션 프로세서(communication processor(CP)), ARM 프로세서 중 하나 또는 그 이상을 포함하거나, 해당 용어로 정의될 수 있다. 또한, 프로세서(140)는 프로세싱 알고리즘이 내장된 SoC(System on Chip), LSI(large scale integration)로 구현될 수도 있고, FPGA(Field Programmable gate array) 형태로 구현될 수도 있다.

[0049] 프로세서(120)는 메모리(110)에 저장된 명령어를 실행함으로써, 학습 네트워크 모델에 입력 영상 프레임을 입력하고, 학습 네트워크 모델로부터 출력되는 관심 영역에 대한 정보에 기초하여 출력 영상 프레임을 획득할 수 있다.

[0050] 일 실시 예에 따라, 프로세서(120)는 학습 네트워크 모델로부터 출력되는 관심 영역에 대한 정보에 기초하여 입력 영상 프레임을 리타겟팅 즉, 종횡비를 조정하여 출력 영상 프레임을 획득한다. 여기서, 관심 영역에 대한 정보는, 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함할 수 있다. 예를 들어, 관심 영역에 대한 정보는, 입력 영상 프레임의 수평 방향(또는 수직 방향)으로 식별되는 위치 정보 및 크기 정보를 포함하는 1차원 정보를 포함할 수 있다. 예를 들어 입력 영상 프레임을 수평 방향으로 확대 스케일링하는 경우 관심 영역에 대한 정보는, 수평 방향으로 식별되는 위치 정보 및 크기 정보를 포함하는 1차원 정보일 수 있다. 다만, 입력 영상을 수직 방향으로 확대 스케일링하는 경우 관심 영역에 대한 정보는, 수직 방향으로 식별되는 위치 정보 및 크기 정보를 포함하는 1차원 정보일 수 있다. 다만, 이하에서는 관심 영역에 대한 정보가 입력 영상 프레임의 수평 방향으로 식별되는 위치 정보 및 크기 정보를 포함하는 1차원 정보인 경우로 상정하여 설명하도록 한다. 다만, 필요에 따라서는 수직 및 수평 방향으로 식별되는 위치 정보 및 크기 정보를 모두 포함하는 것도 가능하다.

[0051] 한편, 학습 네트워크 모델은, 관심 영역에 대한 정보 및 관심 영역에 대한 정보에 대응되는 복수의 이미지를 이용하여 학습 네트워크 모델에 포함된 뉴럴 네트워크의 가중치를 학습할 수 있다. 여기서, 복수의 이미지는 별개의 정지 영상 이미지, 동영상을 구성하는 연속된 복수의 이미지 등 다양한 타입의 이미지를 포함할 수 있다.

[0052] 본 개시의 일 실시 예에 따르면, 학습 네트워크 모델은, 입력 영상 프레임에 포함된 객체 관련 정보를 검출하는 특징 검출부 및, 관심 영역의 크기 정보, 위치 정보 및 타입 정보를 획득하는 특징 맵 추출부를 포함할 수 있다. 예를 들어, 특징 검출부는, 에지 정보, 코너 정보, 컬러 정보 등을 포함하는 특징 정보의 조합에 기초하

여 일부 객체 정보 또는 전체 객체 정보 중 적어도 하나를 포함하는 객체 관련 정보를 검출할 수 있다. 특징 맵 추출부는, 본 개시의 일 실시 예에 따른 관심 영역에 대한 크기 정보 및 위치 정보를 1차원 정보로 추출하고 관심 영역에 포함된 객체의 타입 정보를 획득할 수 있다. 여기서, 1차원 정보는, 영상 프레임의 수평 방향으로 식별되는 위치 정보 및 크기 정보를 포함할 수 있다. 또한, 타입 정보는 기 정의된 개수(예를 들어 20개)의 클래스(또는 타입) 중 하나에 대한 정보가 될 수 있다.

[0053] 도 3 내지 도 5는 본 개시의 일 실시 예에 따른 학습 네트워크 모델을 설명하기 위한 도면들이다.

[0054] 본 개시의 일 실시 예에 따른 학습 네트워크 모델은, 입력 영상에 대해 연속적인 합성 곱 연산을 통하여 원하는 출력 데이터를 얻을 수 있도록 설계하고, 이를 학습하여 획득된 모델일 수 있다. 특히, 학습 네트워크 모델은, 입력 영상에 대해서 의미있는 특정 객체들의 위치, 크기, 명칭을 예측하는 모델일 수 있다.

[0055] 본 개시의 일 실시 예에 따른 학습 네트워크 모델(또는 Classification network model)은 기존의 학습 모델(예를 들어, Darknet-19(C 언어 기반 VGG-19 network)을 객체 검출이 가능하도록 fine-tuning하여 획득될 수 있다. 예를 들어, 학습 네트워크 모델은 기존 심층 인공 신경망 기반 객체 검출 시스템을 본 개시의 실시 예에 따라 변형시켜 재학습시킨 모델이 될 수 있다.

[0056] 예를 들어, 도 3에 따르면, 본 개시의 일 실시 예에 따른 학습 네트워크 모델은 Darknet-19(300)에 포함된 복수의 레이어(310, 320, 330) 중 15 layer(330) 이후 부분을 제거한 후, 7 layers(340, 350)를 추가하여 획득된 총 22 layer를 포함하는 학습 네트워크 모델일 수 있다. 여기서, 기본 레이어(310, 320)는 상술한 특징 검출부에 해당하며, 추가 레이어(340, 350)는 특징 맵 검출부에 해당할 수 있으나, 이에 한정되는 것은 아니다. 예를 들어, 기본 레이어(310, 320)를 선 학습시킨 후, 추가 레이어(340, 350)를 결합하여 추가 학습시킬 수 있다.

[0057] 본 개시의 일 실시 예에 따르면, 입력 영상 프레임의 크기를 수평 방향으로 재조정하기 위하여, 학습 네트워크 모델을 통해 영상 프레임의 각 행 단위 관심 영역 정보를 획득할 수 있다. 이에 따라 학습 네트워크 모델의 출력 데이터는 X(수평 방향 위치 정보), W(크기 정보), 명칭이 될 수 있다. 이 경우, Y(수직 방향 위치 정보)는 필요로 하지 않기 때문에 인공 신경망의 합성 곱 층이 Darknet-19(REDMON, Joseph; FARHADI, Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.)의 81 개보다 작더라도 정확한 예측 데이터를 획득할 수 있게 된다. 즉, 본 개시에서는 도 3에서 설명한 바와 같이 22개의 합성 곱 층을 이용하여 객체 검출 예측에 필요한 파라미터 수를 대폭 감소시키면서 종래 기술과 비슷한 성능을 유지할 수 있게 된다.

[0058] 구체적으로, 도 5에 도시된 바와 같이 분류 네트워크, 예를 들어, Darknet-19(510)에 포함된 종단 15 레이어(330)를 제거한 후 7 layers(340, 350)를 추가하여 총 22 layer의 네트워크로 재구성(520)한 후, 샘플 데이터 셋을 기반으로 객체 검출을 학습(1차 학습)(530)시킨 후, 연속된 프레임의 예측값을 획득(540)하고, 획득된 예측값에 대하여 추가 학습(2차 학습)(550)시킬 수 있다. 다만, 상술한 바와 같이 7 layers(340, 350)를 결합하기 전에 기본 레이어(310, 320)는 객체 검출에 대해 선 학습될 수 있으나, 이에 한정되는 것은 아니다.

[0059] 여기서, 샘플 데이터 셋은 다양한 타입의 영상 및 영상 내의 객체에 대한 1차원 정보(후술하는 박스 정보) 및 기설정된 개수(예를 들어 20개)의 명칭 정보를 포함할 수 있다. 즉, 샘플 데이터 셋을 기반으로 X(수평 방향 위치 정보), W(크기 정보), 명칭에 대해 인공 신경망을 학습시킬 수 있다.

[0060] 추가 학습시에는 이전 영상 프레임에 대한 예측값을 다음 영상 프레임의 진리값으로 취급하여 인공 신경망을 학습시킬 수 있다. 이 경우, 인공 신경망이 연속적으로 입력되는 영상 프레임에 대해 객체의 균일한 크기 및 위치 예측을 할 수 있도록 시간 손실 함수를 정의하여 시간 손실(Temporal Loss)을 최소화 하는 방향으로 학습을 시킬 수 있다. 하기 수학식 1은 일 실시 예에 따라 정의된 시간 손실 함수를 나타낸다.

수학식 1

[0061] $$L_{width} = \beta(\sqrt{w_t} - \sqrt{w_{t-1}})^2$$

수학식 2

$$L_{center} = \alpha(x_t - x_{t-1})^2$$

[0062]

[0063] 수학식 1 및 2와 같은 시간 손실 함수를 사용함으로써 동영상의 연속된 프레임에 대하여 예측된 객체의 W, X의 값이 유사하게 지속되는 방향으로 인공 신경망을 추가 학습할 수 있다. 해당 추가 학습에는 연속된 영상 프레임이 필요하기 때문에 여러 동영상의 영상 프레임으로 시간 손실 함수에 대하여 추가 학습시킬 수 있다.

[0064] 구체적으로, 수학식 1 및 2에서 이전 영상 프레임의 예측 결과인 x_{t-1}, w_{t-1}은 현재 영상 프레임의 예측 결과 x_t, w_t와 비교되는 진리 값이 되며, 인공 신경망은 이 값들의 차이가 최소화되도록 학습될 수 있다. 즉, 연속된 영상 프레임에서 예측 결과들이 유사해지는 것을 기대할 수 있도록 학습될 수 있다. 이 경우, 네트워크 종단 부분을 제거하지 않고 22 layers 모두 학습에 참여시킬 수 있다. 예를 들어, Learning rate=0.01, Weight decay=0.0005, momentum=0.9로 학습시킬 수 있으나, 이에 한정되는 것은 아니다.

[0065] 도 2로 돌아와서, 다른 실시 예에 따르면, 프로세서(120)는 1차원 정보의 크기 또는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 입력 영상 프레임에서 관심 영역을 식별할 수 있다. 여기서, 1차원 정보는 상술한 바와 같이 입력 영상 프레임의 수평 방향으로 식별되는 위치 정보(X) 및 크기 정보(Y)를 포함할 수 있다.

[0066] 일 예로 프로세서(120)는 기설정된 크기 이상의 1차원 정보 만을 관심 영역으로 식별할 수 있다. 다른 예로 프로세서(120)는 기설정된 타입(예를 들어, 사람, 동물)의 1차원 정보 만을 관심 영역으로 식별할 수 있다. 또 다른 예로, 프로세서(120)는 기설정된 크기 이상의 1차원 정보이면서 해당 정보가 기설정된 타입(예를 들어, 사람, 동물)에 해당하는 경우에만 해당 1차원 정보를 관심 영역으로 식별할 수 있다.

[0067] 도 6은 본 개시의 일 실시 예에 따른 관심 영역을 획득하는 방법을 설명하기 위한 도면이다.

[0068] 본 개시의 일 실시 예에 따르면, 학습 네트워크 모델은 객체의 중심 좌표 X, 가로 크기 W, 명칭 정보를 예측하여 1차원 정보로서 관심 영역 정보를 출력할 수 있다. 예를 들어, 도 6에 도시된 바와 같이 관심 영역에 대응되는 영역에 1차원 박스 형태의 정보(또는 바운딩 박스 정보)를 출력할 수 있는데, 이 경우, 1차원 박스 정보는 1의 값으로, 다른 영역은 0의 값으로 설정될 수 있다. 즉, 1차원 박스 정보는 가로 영역 즉, 위치 및 크기 정보를 특정할 수 있으며, 이 경우 1차원 박스 정보에 기초하여 대응되는 세로 영역을 투사하여 1차원 관심 영역을 획득할 수 있다. 예를 들어, 1차원 관심 영역은 영상의 각 행에 대해서 관심 영역인지 아닌지를 1 또는 0의 값으로 나타내는 1차원 행렬로 구현될 수 있다.

[0069] 다만 학습 네트워크 모델을 통해 예측된 영상의 모든 1차원 박스 정보를 관심 영역으로 투사하게 되면, 불필요한 관심 영역이 획득될 수도 있다. 예를 들어, 사람이 너무 많은 영상에서 모든 사람을 관심 영역으로 취급하는 것은 불필요한 작업이 될 수 있다. 또한, 너무 많은 개수의 명칭(또는 타입)에 대해 재학습된 학습 네트워크 모델을 이용하는 경우, 자동차와 사람 등의 예측 가능한 모든 객체를 전부 1차원 박스로 출력하기 때문에 불필요한 작업이 될 수 있다. 이에 따라, 영상의 특성에 기초하여 기설정된 특정 명칭(또는 타입)의 객체를 포함하는 1차원 박스 만을 1차원 관심 영역으로 투사하도록 구현할 수 있다. 예를 들어, 도 6에 도시된 바와 같이, 입력된 영상 프레임(610)에서 복수의 1차원 박스(621, 622)가 획득되면, 1차원 박스의 크기 또는 1차원 박스에 대응되는 객체의 명칭(또는 타입)에 기초하여 특정 1차원 박스(621) 만을 선택할 수 있다. 이 경우, 해당 1차원 박스(621) 만이 1차원 관심 영역으로 투사될 수 있다. 즉, 해당 차원 박스(621) 만에 대응되는 영역 만이 관심 영역으로 식별될 수 있다.

[0070] 다만, 상술한 실시 예에서는 학습 네트워크 모델이 식별된 모든 1차원 박스를 출력하고, 프로세서(120)가 1차원 박스의 크기 또는 1차원 박스에 대응되는 객체의 명칭(또는 타입)에 기초하여 특정 1차원 박스를 식별하는 것으로 설명하였지만, 특정 1차원 박스의 식별 역시 구현 예에 따라 학습 네트워크 모델에서 수행될 수 있음은 물론이다. 즉, 학습 네트워크 모델은, 임계 개수 이상의 1차원 박스가 식별되면, 1차원 박스의 크기 또는 1차원 박스에 대응되는 객체의 명칭(또는 타입)에 기초하여 특정 1차원 박스를 식별하여 식별된 1차원 박스 정보 만을 출력하도록 학습될 수 있다.

[0071] 한편, 학습 네트워크 모델이 특정 프레임에서 객체를 검출하지 못하거나, 동일한 장면에 포함된 동일한 객체에 대해 예측한 1차원 정보의 크기가 매 프레임마다 상이할 수 있다. 이는 연속된 프레임의 관심 영역이 균일하지 못할 수 있음을 의미하고, 이 경우 리타겟팅된 영상이 흔들릴 우려가 있게 된다. 이에 따라 관심 영역에 대한 정보에 있어 시간적 일관성을 고려해야 할 수 있다.

[0072] 이에 따라 본 개시의 다른 실시 예에 따르면, 프로세서(120)는 제1 입력 영상 프레임 및 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 제1 입력 영상 프레임에 대응되는 관심 영역에 대한 정보를 획득할 수 있다. 여기서, 제1 입력 영상 프레임 및 적어도 하나의 제2 입력 영상 프레임은 동일한 씬에 속하는 프레임들일 수 있으나, 이에 한정되는 것은 아니다. 다만, 해당 동작을 학습 네트워크를 통해 학습시키는 경우, 학습 네트워크를 통해 해당 정보를 획득할 수 있음은 물론이다.

[0073] 구체적으로, 프로세서(120)는 제1 입력 영상 프레임 및 제2 입력 영상 프레임 각각에서 획득된 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 제1 입력 영상 프레임에 대응되는 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득할 수 있다. 이 경우, 프로세서(120)는, 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득할 수 있다. 다만, 해당 동작을 학습 네트워크를 통해 학습시키는 경우, 학습 네트워크를 통해 해당 정보를 획득할 수 있음은 물론이다.

[0074] 예를 들어, 프로세서(120)는, 동일한 장면에 포함된 각 영상 프레임의 1차원 정보에 대해 이동 평균 필터를 적용할 수 있다.

[0075] 예를 들어, 도 7에 도시된 바와 같이 프로세서(120)는 이전 영상 프레임(예를 들어, n-2, n-1 번째 프레임) 및 현재 영상 프레임(예를 들어, n번째 프레임)의 관심 영역 정보를 누적하여, 현재 영상 프레임에 대응되는 관심 영역 정보를 획득할 수 있다. 여기서, 이전 영상 프레임은 현재 영상 프레임과 동일한 씬을 구성하는 영상 프레임일 수 있다. 구체적으로, 프로세서(120)는 도시된 바와 같이 이전 영상 프레임 및 현재 영상 프레임 각각에서 획득된 관심 영역에 대한 크기 정보 및 위치 정보에 대한 평균 크기 정보 및 평균 위치 정보를 획득하고, 평균 크기 정보 및 평균 위치 정보를 현재 영상 프레임에 대응되는 관심 영역의 크기 정보 및 위치 정보로 결정할 수 있다.

[0076] 상술한 바와 같이 이동 평균이 적용된 관심 영역을 이용하면 관심 영역의 유실 또는 움직임에 따른 시간적 왜곡을 감소시킬 수 있게 된다.

[0077] 도 2로 돌아와서, 프로세서(120)는 관심 영역 및 나머지 영역(또는 비관심 영역)이 식별되면, 각 영역에 포함된 입력 영상 프레임의 각 행(또는 각 열)에 대한 변환 가중치, 즉 크기 변환 비율을 결정할 수 있다. 프로세서(120)는 관심 영역에 포함된 행에 대해서는 수평 크기 변환 비율을 1로 할당하고, 나머지 영역에 포함된 행에 대해서는 타겟 종횡비를 만족시키는 변환 비율을 산출할 수 있다. 예를 들어 프로세서(120)는 나머지 영역에 포함된 행에 대해서는 타겟 종횡비를 만족시킬 때까지 변환 비율을 증가시켜 변환 비율을 산출할 수 있다.

[0078] 설명의 편의를 위하여 입력 영상 프레임의 수평 방향의 폭을 W (픽셀 단위)라 하고, 타겟 수평 방향의 폭을 W' (픽셀 단위)라 가정한다. 1차원 관심 영역의 해당하는 행의 개수를 X 라 하면, 관심 영역을 제외한 다른 영역의 크기 변환 비율 γ 는 하기 수학식 3과 같이 산출될 수 있다.

수학식 3

[0079] $$\gamma = \frac{W' - X}{W - X}$$

[0080] 이에 따라, 프로세서(120)는 제1 가중치를 1로 결정하고, 제2 가중치를 γ 로 결정하여 영상 크기를 조정할 수 있다.

[0081] 일 실시 예에 따라, 프로세서(120)는 학습 네트워크 모델로부터 관심 영역에 대한 정보가 획득되면, 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역(또는 비관심 영역)에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 입력 영상 프레임을 리타겟팅할 수 있다. 여기서, 제2 변환 가중치는, 리타겟팅 정보에 기초하여 획득될 수 있다. 여기서, 리타겟팅 정보는, 입력 영상 프레임의 해상도 정보 및 출력

영상 프레임의 해상도 정보를 포함할 수 있다. 또는, 리타겟팅 정보는 입력 영상 프레임의 종횡비 및 출력 영상 프레임의 종횡비를 포함할 수 있다. 또는, 리타겟팅 정보는 입력 영상 프레임의 해상도 정보 및 출력 영상 프레임의 해상도 정보에 기초하여 산출된 입력 영상 프레임의 종횡비 조정 정보를 포함할 수 있다. 예를 들어, 프로세서(120)는 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하여 대응되는 영역 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 대응되는 영역 크기를 확대 스케일링할 수 있다.

[0082] 구체적으로, 프로세서(120)는 관심 영역 정보(또는 이동 평균이 적용된 관심 영역 정보)로부터 비관심 영역에 대한 수평 크기 변환 비율이 계산되면 입력 영상 프레임의 각 행 데이터가 리타겟팅 영상 프레임에서 새로 배치되는 행 위치를 식별할 수 있다. 예를 들어, 입력 영상 프레임과 수평 크기 변환 비율 행렬을 곱 연산하여 행 영상 데이터를 재배치할 수 있다. 이 과정에서 빈 공간의 행 데이터가 발생할 수 있으며, 영상 보간을 수행하여 빈 공간의 행 데이터를 인접 행의 영상 데이터와 유사하게 채워 넣어 리타겟팅된 영상 프레임을 획득할 수 있다. 이 경우, 각 영상 프레임에 대응되는 리타겟팅 영상 프레임을 재생함으로써 리타겟팅 동영상을 획득할 수 있다.

[0083] 도 8a 및 도 8b은 본 개시의 일 실시 예에 따른 리타겟팅 방법을 설명하기 위한 도면들이다.

[0084] 도 6에서 설명한 바와 같이 입력 영상 프레임에서 1차원 박스 정보(621)에 기초하여 관심 영역(630)이 식별되면, 관심 영역 및 나머지 영역을 구분하여, 각 영역의 크기 변환 비율을 결정할 수 있다. 예를 들어 프로세서(120)는 관심 영역에 포함된 행에 대해서는 수평 크기 변환 비율을 1로 할당하고, 나머지 영역(비관심 영역)에 포함된 행에 대해서는 타겟 종횡비를 만족시키는 변환 비율을 산출할 수 있다. 예를 들어 프로세서(120)는 나머지 영역(비관심 영역)에 포함된 행에 대해서는 출력 영상 프레임의 해상도를 구현하기 위한 타겟 종횡비를 만족시키는 변환 비율을 산출할 수 있다.

[0085] 이 경우, 도 8a에 도시된 바와 같이 관심 영역(630)의 크기는 유지되고, 나머지 영역의 크기는 확대 스케일링된 형태의 출력 영상 프레임(810)이 획득될 수 있다.

[0086] 여기서, 확대 스케일링이란, 픽셀 값을 추가하는 형태가 될 수 있다. 예를 들어, 나머지 영역(비관심 영역)에 포함된 행의 수평 변환 비율이 2인 경우, 각 행에 포함된 픽셀 값과 동일한 픽셀 값을 가지는 행을 각 행의 인접 행으로 추가하여 수평 방향 크기를 2배로 확대할 수 있다. 다만, 이는 일 예를 든 것이며 종래의 영역의 크기를 확대할 수 있는 다양한 영상 처리 방법에 본 개시에 적용될 수 있음은 물론이다.

[0087] 한편, 도 8a에서는 입력 영상 프레임의 세로 크기(또는 가로 크기)가 조정되지 않아도 되는 경우를 설명한 것이며, 출력 영상 프레임의 해상도를 맞추기 위해 입력 영상 프레임의 가로 크기(또는 세로 크기) 뿐 아니라, 세로 크기(또는 가로 크기)도 조정되어야 할 수 있다.

[0088] 이 경우, 프로세서(120)는 도 8b에 도시된 바와 같이 관심 영역의 수직/수평 변환 비율을 입력 영상 프레임 및 출력 영상 프레임의 세로 크기에 기초하여 결정할 수 있다. 즉, 프로세서(120)는 입력 영상 프레임의 세로 크기가 출력 영상의 세로 크기와 동일해지도록 하는 수직 변환 비율을 결정하고, 해당 수직 변환 비율과 동일한 수평 변환 비율에 기초하여 관심 영역(630)을 수직 및 수평 방향으로 스케일링할 수 있다. 예를 들어, 입력 영상 프레임(610)의 세로 길이가 480 픽셀이고, 출력 영상 프레임(810)의 세로 길이가 960인 경우, 관심 영역의 수직/수평 변환 비율을 4.5(2160/480)로 결정하고, 나머지 영역의 수직/수평 변환 비율을 입력 영상 프레임 및 출력 영상 프레임의 가로 크기에 기초하여 결정할 수 있다.

[0089] 도 9는 본 개시의 다른 실시 예에 따른 리타겟팅 방법을 설명하기 위한 도면이다.

[0090] 본 개시의 다른 실시 예에 따라 관심 영역 정보는 가로 방향 정보(921) 뿐 아니라, 세로 방향 정보(922)를 포함하는 2차원 정보일 수 있다. 이 경우, 프로세서(120)는 2차원 정보에 기초하여 입력 영상 프레임(610)에서 관심 영역(910)을 특정하고, 관심 영역에 포함된 픽셀에 대해서는 수평/수직 변환 비율을 1로 할당하고, 나머지 영역(비관심 영역)에 포함된 픽셀에 대해서는 타겟 종횡비를 만족시키는 변환 비율을 산출할 수 있다. 예를 들어 프로세서(120)는 나머지 영역(비관심 영역)에 포함된 픽셀에 대해서는 출력 영상 프레임의 해상도를 구현하기 위한 타겟 종횡비를 만족시키는 변환 비율을 산출할 수 있다. 이 경우, 프로세서(120)는 입력 영상 프레임 및 출력 영상 프레임의 해상도에 기초하여 수평 변환 비율 및 수직 변환 비율을 각각 산출할 수 있다. 프로세서(120)는 나머지 영역에 포함된 픽셀에 대해 산출된 수평 변환 비율 및 수직 변환 비율을 적용하여 출력 영상 프레임(920)을 획득할 수 있다.

[0091] 도 10은 본 개시의 일 실시 예에 따른 영상 처리 방법을 설명하기 위한 도면이다.

[0092] 도 10에 도시된 영상 처리 방법에 따르면, 프로세서(120)는 n번째 입력 영상 프레임(1010)을 학습 네트워크 모델(1020)에 입력하여 1차원 관심 영역 정보(1030)를 획득할 수 있다. 이 경우, 프로세서(120)는 획득된 1차원 관심 영역 정보 및 이전 프레임에서 획득된 1차원 관심 영역 정보의 누적 이동 평균 정보를 산출(1040)할 수 있다. 이어서, 프로세서(120)는 산출된 누적 이동 평균 정보에 기초하여 입력 영상 프레임의 수평/수직 방향 크기를 조정(1050)하여 n번째 입력 영상 프레임(1010)에 대응되는 출력 영상 프레임(1060)을 획득할 수 있다.

[0093] 도 11은 본 개시의 일 실시 예에 따른 학습 네트워크 모델의 세부 구성을 설명하기 위한 도면이다.

[0094] 도 11에 따르면, 본 개시의 일 실시 예에 따른 학습 네트워크 모델(1100)은, 학습부(1110) 및 인식부(1120)를 포함할 수 있다.

[0095] 학습부(1110)는 소정의 상황 판단을 위한 기준을 갖는 인식 모델을 생성 또는 학습시킬 수 있다. 학습부(1110)는 수집된 학습 데이터를 이용하여 판단 기준을 갖는 인식 모델을 생성할 수 있다. 일 예로, 학습부(1110)는 객체가 포함된 이미지를 학습 데이터로서 이용하여 이미지에 포함된 객체가 어떤 것인지 판단하는 기준을 갖는 객체 인식 모델을 생성, 학습 또는 갱신시킬 수 있다. 또 다른 예로, 학습부(1110)는 객체가 포함된 이미지에 포함된 주변 정보를 학습 데이터로서 이용하여 이미지에 포함된 객체 주변에 다양한 추가 정보를 판단하는 기준을 갖는 주변 정보 인식 모델을 생성, 학습 또는 갱신시킬 수 있다.

[0096] 인식부(1120)는 소정의 데이터를 학습된 인식 모델의 입력 데이터로 사용하여, 소정의 데이터에 포함된 인식 대상을 추정할 수 있다. 일 예로, 인식부(1120)는 객체가 포함된 객체 영역(또는, 이미지)를 학습된 인식 모델의 입력 데이터로 사용하여 객체 영역에 포함된 객체에 대한 객체 정보를 획득(또는, 추정, 추론)할 수 있다. 다른 예로, 인식부(1120)는 객체 정보 및 컨텍스트 정보 중 적어도 하나를 학습된 인식 모델에 적용하여 검색 결과를 제공할 검색 카테고리를 추정(또는, 결정, 추론)할 수 있다. 이 때, 검색 결과는 우선 순위에 따라 복수 개가 획득될 수도 있다.

[0097] 학습부(1110)의 적어도 일부 및 인식부(1120)의 적어도 일부는, 소프트웨어 모듈로 구현되거나 적어도 하나의 하드웨어 칩 형태로 제작되어 영상 처리 장치(100)에 탑재될 수 있다. 예를 들어, 학습부(1110) 및 인식부(1120) 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 전자 장치 또는 객체 인식 장치에 탑재될 수도 있다. 이 때, 인공 지능을 위한 전용 하드웨어 칩은 확률 연산에 특화된 전용 프로세서로서, 기존의 범용 프로세서보다 병렬처리 성능이 높아 기계 학습과 같은 인공 지능 분야의 연산 작업을 빠르게 처리할 수 있다. 학습부(1110) 및 인식부(1120)가 소프트웨어 모듈(또는, 인스트럭션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 이 경우, 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[0098] 도 12은 본 개시의 다른 실시 예에 따른 영상 처리 장치의 일 구현 예를 나타내는 도면이다.

[0099] 도 12에 따르면, 영상 처리 장치(100')는 메모리(110), 프로세서(120), 통신부(130), 디스플레이(140) 및 사용자 인터페이스(150)를 포함한다. 도 11에 도시된 구성 중 도 2에 도시된 구성과 중복되는 구성에 대해서는 자세한 설명을 생략하도록 한다.

[0100] 통신부(130)는 다양한 타입의 컨텐츠를 수신한다. 예를 들어 통신부(130)는 AP 기반의 Wi-Fi(와이파이, Wireless LAN 네트워크), 블루투스(Bluetooth), 지그비(Zigbee), 유/무선 LAN(Local Area Network), WAN, 이더넷, IEEE 1394, HDMI(High Definition Multimedia Interface), MHL (Mobile High-Definition Link), USB (Universal Serial Bus), DP(Display Port), 썬더볼트(Thunderbolt), VGA(Video Graphics Array)포트, RGB 포트, D-SUB(D-subminiature), DVI(Digital Visual Interface) 등과 같은 통신 방식을 통해 외부 장치(예를 들어, 소스 장치), 외부 저장 매체(예를 들어, USB), 외부 서버(예를 들어 웹 하드) 등으로부터 스트리밍 또는 다운로드 방식으로 영상 신호를 입력받을 수 있다. 여기서, 영상 신호는 디지털 신호가 될 수 있으나 이에 한정되는 것은 아니다.

[0101] 디스플레이(140)는 LCD(Liquid Crystal Display), OLED(Organic Light Emitting Diodes) 디스플레이,

LED(Light Emitting Diodes), PDP(Plasma Display Panel) 등과 같은 다양한 형태의 디스플레이로 구현될 수 있다. 디스플레이(160) 내에는 a-si TFT, LTPS(low temperature poly silicon) TFT, OTFT(organic TFT) 등과 같은 형태로 구현될 수 있는 구동 회로, 백라이트 유닛 등도 함께 포함될 수 있다. 한편, 디스플레이(140)는 터치 센서와 결합된 터치 스크린, 플렉시블 디스플레이(flexible display), 3차원 디스플레이(3D display) 등으로 구현될 수 있다.

[0102] 또한, 본 개시의 일 실시 예에 따른, 디스플레이(140)는 영상을 출력하는 디스플레이 패널뿐만 아니라, 디스플레이 패널을 하우징하는 베젤을 포함할 수 있다. 특히, 본 개시의 일 실시예에 따른, 베젤은 사용자 인터렉션을 감지하기 위한 터치 센서(미도시)를 포함할 수 있다.

[0103] 프로세서(120)는 본 개시의 다양한 실시 예에 따라 처리된 영상을 디스플레이하도록 디스플레이(140)를 제어할 수 있다.

[0104] 일 예에 따라 프로세서(120)는 그래픽 처리 기능(비디오 처리 기능)을 수행할 수 있다. 예를 들어, 프로세서(120)는 연산부(미도시) 및 렌더링부(미도시)를 이용하여 아이콘, 이미지, 텍스트 등과 같은 다양한 객체를 포함하는 화면을 생성할 수 있다. 여기서, 연산부(미도시)는 수신된 제어 명령에 기초하여 화면의 레이아웃에 따라 각 객체들이 표시될 좌표값, 형태, 크기, 컬러 등과 같은 속성값을 연산할 수 있다. 그리고, 렌더링부(미도시)는 연산부(미도시)에서 연산한 속성값에 기초하여 객체를 포함하는 다양한 레이아웃의 화면을 생성할 수 있다. 또한, 프로세서(120)는 비디오 데이터에 대한 디코딩, 스케일링, 노이즈 필터링, 프레임 레이트 변환, 해상도 변환 등과 같은 다양한 이미지 처리를 수행할 수 있다.

[0105] 다른 예에 따라, 프로세서(120)는 오디오 데이터에 대한 처리를 수행할 수 있다. 구체적으로, 프로세서(120)는 오디오 데이터에 대한 디코딩이나 증폭, 노이즈 필터링 등과 같은 다양한 처리가 수행될 수 있다.

[0106] 사용자 인터페이스(150)는 버튼, 터치 패드, 마우스 및 키보드와 같은 장치로 구현되거나, 상술한 디스플레이 기능 및 조작 입력 기능도 함께 수행 가능한 터치 스크린으로도 구현될 수 있다. 여기서, 버튼은 영상 처리 장치(100)의 본체 외관의 전면부나 측면부, 배면부 등의 임의의 영역에 형성된 기계적 버튼, 터치 패드, 휠 등과 같은 다양한 유형의 버튼이 될 수 있다.

[0107] 한편, 영상 처리 장치(100)는 구현 예에 따라 튜너 및 복조부를 추가적으로 포함할 수 있다.

[0108] 튜너(미도시)는 안테나를 통해 수신되는 RF(Radio Frequency) 방송 신호 중 사용자에 의해 선택된 채널 또는 기 저장된 모든 채널을 튜닝하여 RF 방송 신호를 수신할 수 있다.

[0109] 복조부(미도시)는 튜너에서 변환된 디지털 IF 신호(DIF)를 수신하여 복조하고, 채널 복호화 등을 수행할 수도 있다.

[0110] 도 13은 본 개시의 일 실시 예에 따른 영상 처리 방법을 설명하기 위한 흐름도이다.

[0111] 도 13에 도시된 영상 처리 방법에 따르면, 입력 영상 프레임을 학습 네트워크 모델에 적용하여 관심 영역에 대한 정보를 획득할 수 있다(S1310). 여기서, 학습 네트워크 모델은, 입력 영상 프레임에서 관심 영역에 대한 정보를 획득하도록 학습된 모델일 수 있다.

[0112] 이어서, 관심 영역에 대한 정보에 기초하여 입력 영상 프레임을 리타겟칭하여 출력 영상 프레임을 획득할 수 있다(S1320).

[0113] 여기서, 관심 영역에 대한 정보는, 관심 영역의 크기 정보, 위치 정보 또는 타입 정보 중 적어도 하나를 포함할 수 있다.

[0114] 또한, 관심 영역에 대한 정보는, 입력 영상 프레임의 수평 방향에 대응되는 위치 정보 및 크기 정보를 포함하는 1차원 정보일 수 있다. 이 경우, 출력 영상 프레임을 획득하는 S1320 단계에서는, 1차원 정보의 크기 또는 1차원 정보에 대응되는 객체의 타입 중 적어도 하나에 기초하여 복수의 1차원 정보 중 적어도 하나의 1차원 정보를 식별하고, 식별된 1차원 정보에 기초하여 입력 영상 프레임을 리타겟팅할 수 있다.

[0115] 또한, 출력 영상 프레임을 획득하는 S1320 단계에서는, 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 입력 영상 프레임을 리타겟팅할 수 있다. 여기서, 제2 변환 가중치는, 입력 영상 프레임에 대한 해상도 정보 및 출력 영상 프레임에 대한 해상도 정보에 기초하여 획득될 수 있다.

[0116] 또한, 출력 영상 프레임을 획득하는 S1320 단계에서는, 관심 영역에 대응되는 픽셀에 대해서는 제1 변환 가중치를 적용하여 크기를 유지하고, 나머지 영역에 대응되는 픽셀에 대해서는 제2 변환 가중치를 적용하여 확대 스케일링할 수 있다.

[0117] 또한, 관심 영역에 대한 정보를 획득하는 S1310 단계에서는, 제1 입력 영상 프레임 및 제1 입력 영상 프레임 이전에 입력된 적어도 하나의 제2 입력 영상 프레임 각각에서 획득된 관심 영역에 대한 정보를 누적하고, 누적된 정보에 기초하여 제1 입력 영상 프레임에 대응되는 관심 영역에 대한 정보를 획득할 수 있다.

[0118] 또한, 관심 영역에 대한 정보를 획득하는 S1310 단계에서는, 제1 입력 영상 프레임 및 제2 입력 영상 프레임 각각에서 획득된 관심 영역에 대한 크기 정보 및 위치 정보에 기초하여 관심 영역의 평균 크기 정보 및 평균 위치 정보를 획득할 수 있다. 이 경우, 출력 영상 프레임을 획득하는 S1220 단계에서는, 관심 영역의 평균 크기 정보 및 평균 위치 정보에 기초하여 제1 입력 영상 프레임에 대응되는 출력 영상 프레임을 획득할 수 있다.

[0119] 한편, 학습 네트워크 모델은, 입력 영상 프레임에 포함된 객체 관련 정보를 검출하는 특징 검출부 및 관심 영역의 크기 정보, 위치 정보 및 타입 정보를 획득하는 특징 맵 추출부를 포함할 수 있다.

[0120] 또한, 학습 네트워크 모델은, 관심 영역에 대한 정보 및 관심 영역에 대한 정보에 대응되는 복수의 이미지를 이용하여 학습 네트워크 모델에 포함된 뉴럴 네트워크의 가중치를 학습할 수 있다.

[0121] 상술한 다양한 실시 예들에 따르면, 기존 cropping 기반 방법(영상 특성에 관계 없이 영상의 가로 혹은 세로 축을 기준으로 일부 영역만 잘라내는 기법), seam carving 기반 방법(영상 내에 중요하지 않은 연결된 선 (seam)들을 찾아 찾아진 선이 있는 영역을 늘리거나 줄여 종횡비를 조정하는 기법) 혹은 warping 기반 방법(영상 내 픽셀별로 중요도를 판단하여 중요도에 따라서 부분적으로 영상을 늘리거나 줄여 원하는 종횡비를 조정하는 기법)에서도 달성할 수 없었던 시간적 일관성(temporal coherency)을 유지하면서 영상 컨텐츠의 왜곡을 최소화할 수 있게 된다.

[0122] 다만, 본 개시의 다양한 실시 예들은 영상 처리 장치 뿐 아니라, 셋탑 박스와 같은 영상 수신 장치, TV와 같은 디스플레이 장치 등 영상 처리가 가능한 모든 영상 처리 장치에 적용될 수 있음은 물론이다.

[0123] 한편, 상술한 본 개시의 다양한 실시 예들에 따른 방법들은, 기존 영상 처리 장치에 설치 가능한 어플리케이션 형태로 구현될 수 있다.

[0124] 또한, 상술한 본 개시의 다양한 실시 예들에 따른 방법들은, 기존 영상 처리 장치에 대한 소프트웨어 업그레이드, 또는 하드웨어 업그레이드 만으로도 구현될 수 있다.

[0125] 또한, 상술한 본 개시의 다양한 실시 예들은 영상 처리 장치에 구비된 임베디드 서버, 또는 영상 처리 장치 및 디스플레이 장치 중 적어도 하나의 외부 서버를 통해 수행되는 것도 가능하다.

[0126] 한편, 본 개시의 일시 예에 따르면, 이상에서 설명된 다양한 실시 예들은 기기(machine)(예: 컴퓨터)로 읽을 수 있는 저장 매체(machine-readable storage media에 저장된 명령어를 포함하는 소프트웨어로 구현될 수 있다. 기기는, 저장 매체로부터 저장된 명령어를 호출하고, 호출된 명령어에 따라 동작이 가능한 장치로서, 개시된 실시 예들에 따른 영상 처리 장치(예: 영상 처리 장치(A))를 포함할 수 있다. 명령이 프로세서에 의해 실행될 경우, 프로세서가 직접, 또는 프로세서의 제어 하에 다른 구성요소들을 이용하여 명령에 해당하는 기능을 수행할 수 있다. 명령은 컴파일러 또는 인터프리터에 의해 생성 또는 실행되는 코드를 포함할 수 있다. 기기로 읽을 수 있는 저장 매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 신호(signal)를 포함하지 않으며 실재(tangible)한다는 것을 의미할 뿐 데이터가 저장매체에 반영구적 또는 임시적으로 저장됨을 구분하지 않는다.

[0127] 또한, 본 개시의 일 실시 예에 따르면, 이상에서 설명된 다양한 실시 예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로, 또는 어플리케이션 스토어(예: 플레이 스토어TM)를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

[0128] 또한, 상술한 다양한 실시 예들에 따른 구성 요소(예: 모듈 또는 프로그램) 각각은 단수 또는 복수의 개체로 구성될 수 있으며, 전술한 해당 서브 구성 요소들 중 일부 서브 구성 요소가 생략되거나, 또는 다른 서브 구성 요소가 다양한 실시 예에 더 포함될 수 있다. 대체적으로 또는 추가적으로, 일부 구성 요소들(예: 모듈 또는 프로

그램)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성 요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 다양한 실시 예들에 따른, 모듈, 프로그램 또는 다른 구성 요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

[0129] 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시에 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안될 것이다.

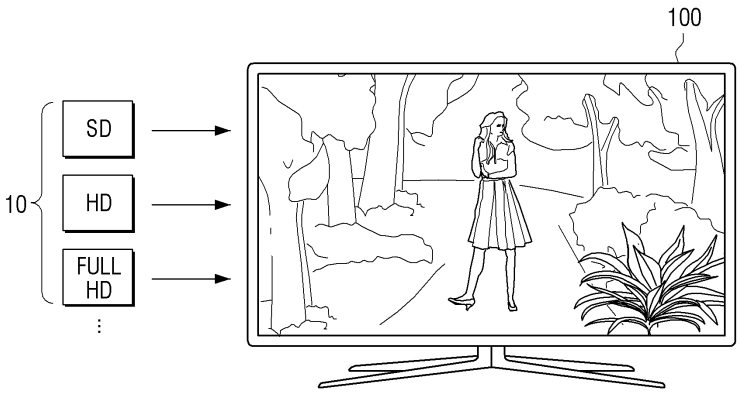
부호의 설명

[0130] 100: 영상 처리 장치 110: 메모리

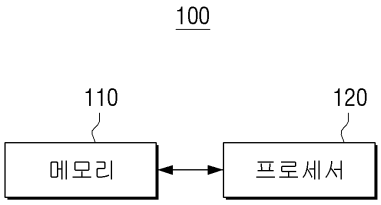
120: 프로세서

도면

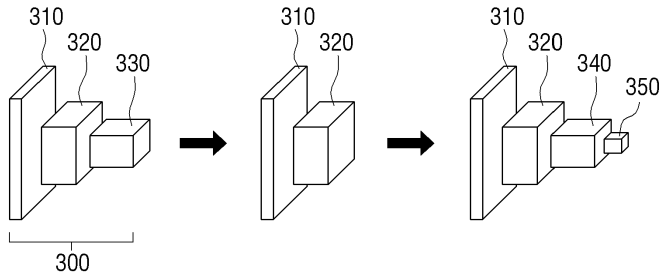
도면1



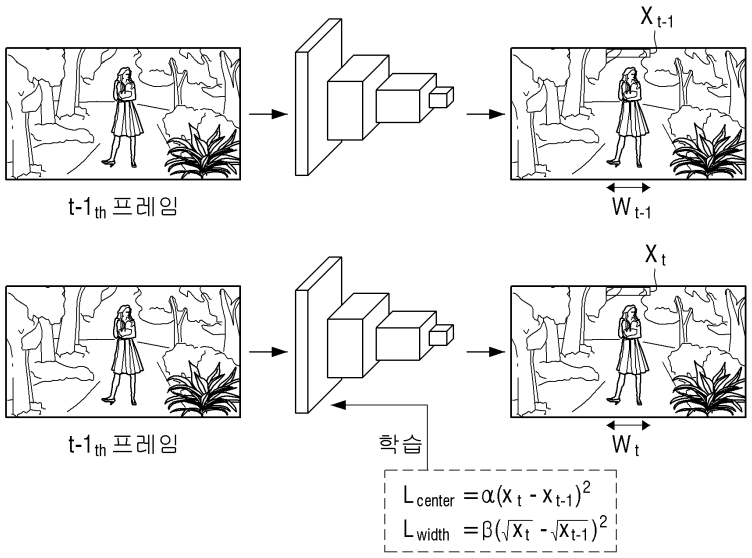
도면2



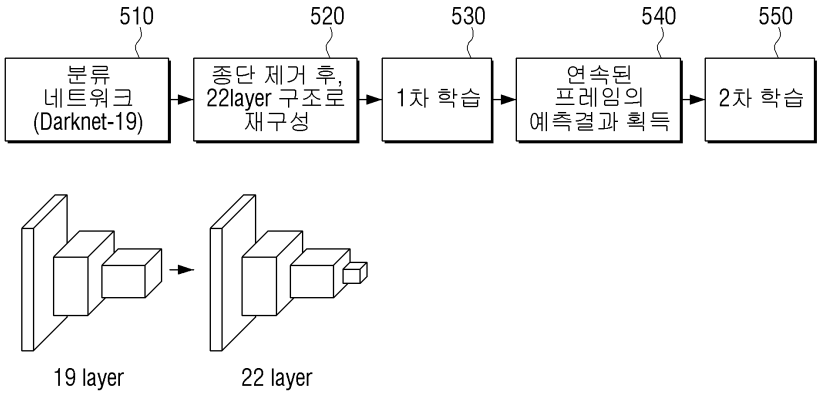
도면3


310
320
330
310
320
310
320
340
350
300

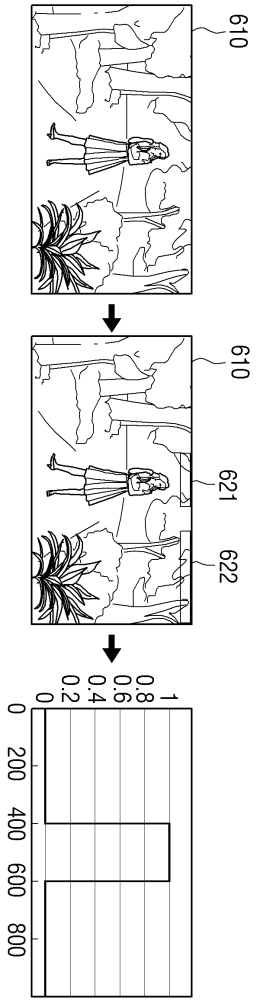
도면4


X t-1
t-1th 프레임
W t-1
X t
t-1th 프레임
학습
W t
L center =α(x t - x t-1)2
L width =β(√x t - √x t-1)2

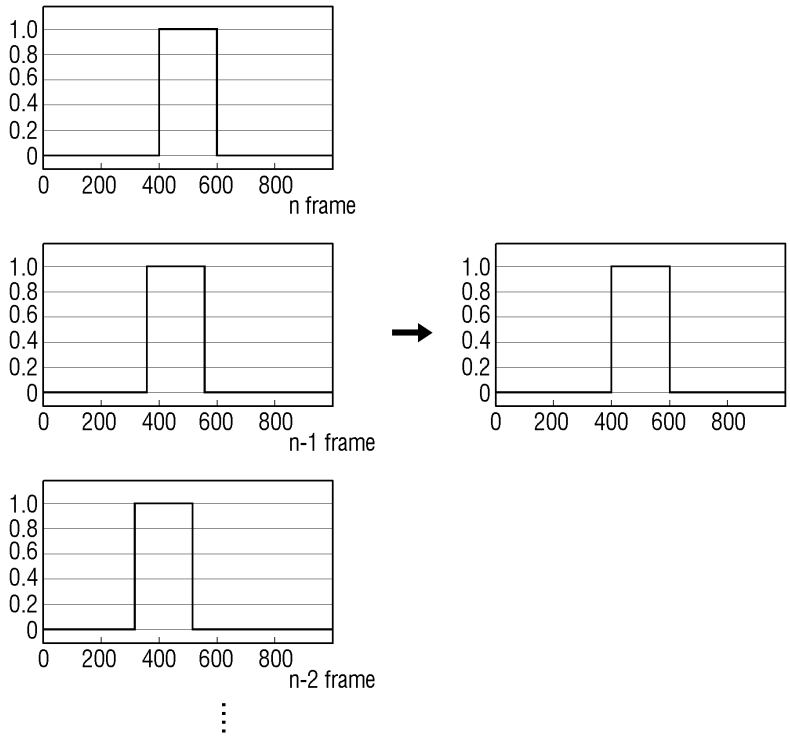
도면5


510
520
530
540
550
분류 네트워크 (Darknet-19)
종단 제거 후, 22layer 구조로 재구성
1차 학습
연속된 프레임의 예측결과 획득
2차 학습
19 layer
22 layer

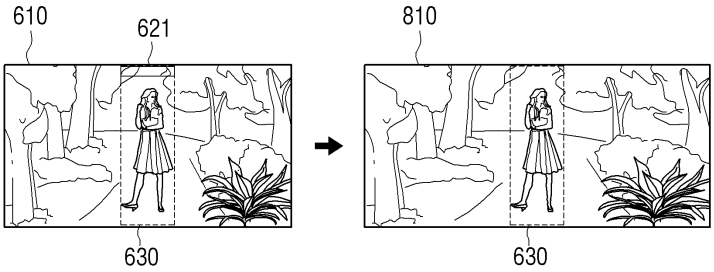
도면6



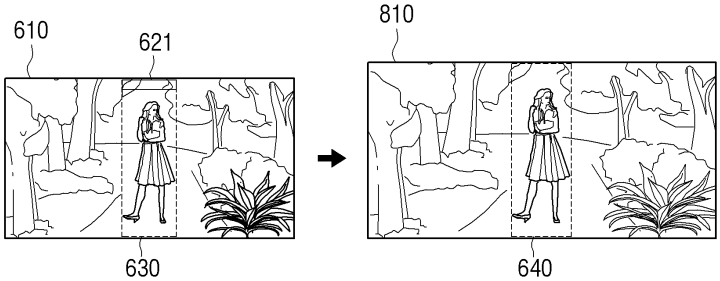
도면7



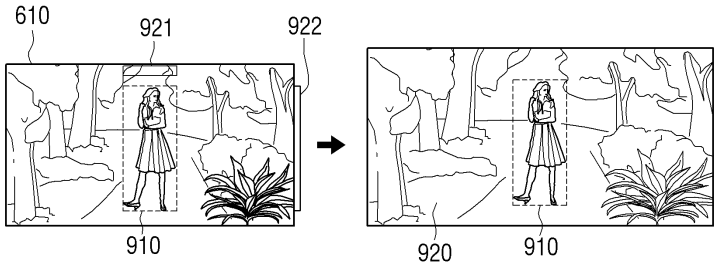
도면8a



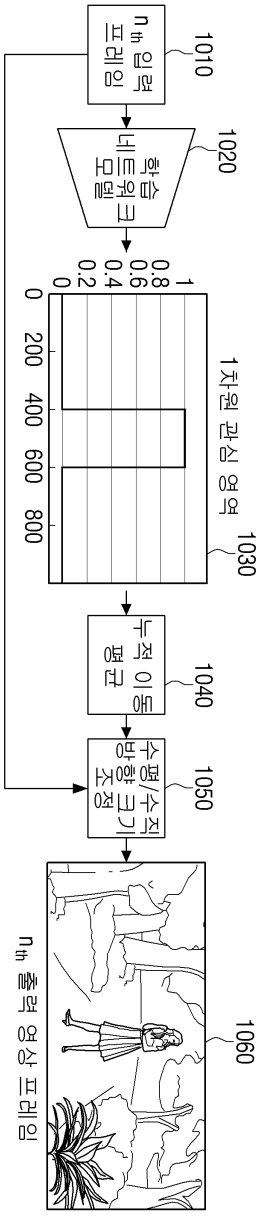
도면8b



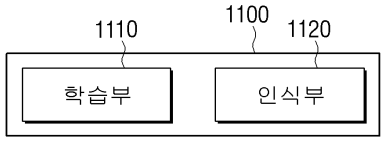
도면9


610
921
922
910
920
910

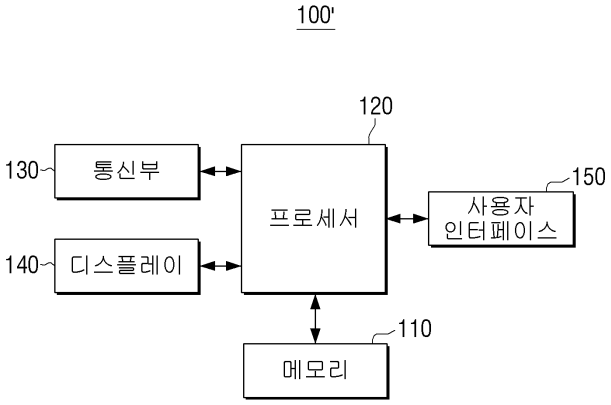
도면10


nth 입력 프레임
1010
학습 네트워크 모델
1020
1
0.8
0.6
0.4
0.2
0
0
200
400
600
800
1차원 관심 영역
1030
누적 이동 평균
1040
수평/수직 방향 크기 조정
1050
nth 출력 영상 프레임
1060

도면11



도면12



도면13