(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2020 년 5 월 22 일 (22.05.2020)

WIPO | PCT

(10) 국제공개번호
WO 2020/101434 A1

(51) 국제특허분류:
G06T 3/40 (2006.01) *G06N 3/02* (2006.01)
G06T 5/20 (2006.01)

(21) 국제출원번호: PCT/KR2019/015668

(22) 국제출원일: 2019 년 11 월 15 일 (15.11.2019)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2018-0141141 2018 년 11 월 15 일 (15.11.2018) KR

(71) 출원인: 삼성전자 주식회사 (**SAMSUNG ELECTRONICS CO., LTD.**) [KR/KR]; 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 한국과학기술원 (**KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY**) [KR/KR]; 34141 대전시 유성구 대학로 291, Daejeon (KR).

(72) 발명자: 박두복 (**PARK, Dubok**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 조동현 (**CHO, Donghyeon**); 34141 대전시 유성구 대학로 291, Daejeon (KR). 권인소 (**KWEON, Inso**); 34141 대전시 유성구 대학로 291, Daejeon (KR). 남현우 (**NAM, Hyunwoo**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 전강원 (**JEON, Kangwon**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).

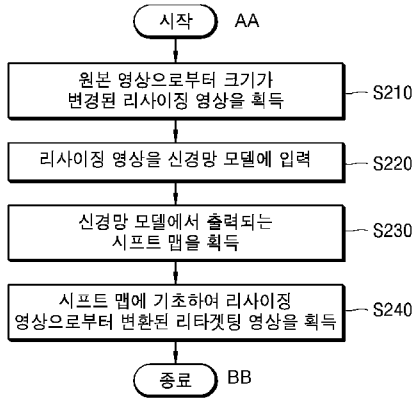
(74) 대리인: 리앤목 특허법인 (**Y.P.LEE, MOCK & PARTNERS**); 06292 서울시 강남구 언주로30길 13 대림아크로텔 12층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

(54) **Title:** IMAGE PROCESSING DEVICE AND METHOD FOR RETARGETING OF IMAGE

(54) **발명의 명칭:** 영상의 리타겟팅을 위한 영상 처리 장치 및 방법



S210 ... Acquire, from original image, resizing image having been changed in size
S220 ... Input resizing image in neural network model
S230 ... Acquire shift map outputted from neural network model
S240 ... Acquire transformed retargeting image from resizing image on basis of shift map
AA ... Start
BB ... End

(57) **Abstract:** Disclosed is an image processing method, according to one embodiment, comprising the steps of: acquiring, from an original image, a resizing image having been changed in size, according to a predetermined aspect ratio; inputting the resizing image in a neural network model trained in advance; and acquiring a transformed retargeting image from the resizing image on the basis of a shift map outputted from the neural network model.

(57) **요약서:** 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하는 단계; 리사이징 영상을 미리 훈련된 신경망 모델에 입력하는 단계; 및 신경망 모델에서 출력되는 시프트 맵에 기초하여 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함하는 것을 특징으로 하는 일 실시예에 따른 영상 처리 방법이 개시된다.

WO 2020/101434 A1

[다음 쪽 계속]

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 영상의 리타겟팅을 위한 영상 처리 장치 및 방법

기술분야

[1] 본 개시는 영상 처리 분야에 관한 것이다. 보다 구체적으로, 본 개시는 영상 내 중요 부분의 왜곡이 최소화되도록 영상을 리타겟팅하는 영상 처리 장치 및 방법에 관한 것이다.

배경기술

[2] 카메라, 디스플레이 장치들의 발전에 따라 다양한 해상도 및 종횡비를 갖는 기기들이 출시되고 있다. 이러한 환경하에서, 카메라에 의해 촬영된 영상의 종횡비가 디스플레이 장치의 종횡비에 맞지 않는 상황이 발생할 수 있다. 예를 들어, 카메라에 의해 촬영된 영상이 4:3의 종횡비를 가지고 있고, 디스플레이 장치가 16:9의 종횡비를 가지고 있는 경우, 해당 영상을 디스플레이 장치에서 풀(full) 스크린으로 재생하기 위해서는 영상을 스트레칭(stretching)시켜야 하는데, 이러한 스트레칭은 영상의 왜곡을 야기한다.

[3] 영상의 종횡비 조절을 위한 영상 리타겟팅 기술이 이용되고 있다. 영상 리타겟팅 기술이란, 원본 영상 내 픽셀들의 위치를 변경하거나, 새로운 픽셀을 추가하거나, 기존 픽셀을 제거하는 방식 등으로 원본 영상으로부터 원하는 종횡비를 갖는 영상을 생성하는 기술을 의미한다. 그러나, 아직까지는 소정 종횡비를 갖는 영상을 생성하기 위한 효율적인 처리 방법이 제안되어 있지 않은 실정이다.

발명의 상세한 설명

기술적 과제

[4] 일 실시예에 따른 영상의 처리 장치 및 방법은 영상 내 중요 부분의 왜곡이 최소화되도록 영상을 리타겟팅 하는 것을 기술적 과제로 한다.

[5] 또한, 일 실시예에 따른 영상의 처리 장치 및 방법은 시간적으로 관련된 복수의 영상들 사이의 연속성이 보존되도록 하는 것을 기술적 과제로 한다.

과제 해결 수단

[6] 일 실시예에 따른 영상 처리 방법은, 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하는 단계; 상기 리사이징 영상을 미리 훈련된 신경망 모델에 입력하는 단계; 및 상기 신경망 모델에서 출력되는 시프트 맵에 기초하여 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함할 수 있다.

발명의 효과

[7] 일 실시예에 따른 영상의 처리 장치 및 방법은 영상 내 중요 부분의 왜곡이 최소화되도록 영상을 리타겟팅할 수 있다.

[8] 또한, 일 실시예에 따른 영상의 처리 장치 및 방법은 시간적으로 관련된 복수의

영상들 사이의 연속성이 보존되도록 할 수 있다.

[9] 다만, 일 실시예에 따른 영상의 처리 장치 및 방법이 달성할 수 있는 효과는 이상에서 언급한 것들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[10] 본 명세서에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 간단한 설명이 제공된다.

[11] 도 1은 일 실시예에 따른 영상 처리 장치가 적용되는 환경을 설명하기 위한 도면이다.

[12] 도 2는 일 실시예에 따른 영상 처리 방법을 설명하기 위한 순서도이다.

[13] 도 3은 일 실시예에 따른 리타겟팅 과정을 설명하기 위한 도면이다.

[14] 도 4는 일 실시예에 따른 신경망 모델의 구조를 설명하기 위한 도면이다.

[15] 도 5는 일 실시예에 따른 신경망 모델의 구조를 상세히 설명하기 위한 도면이다.

[16] 도 6 및 도 7은 제 2 서브 모델에 입력되는 위치 맵을 도시하는 예시적인 도면이다.

[17] 도 8은 훈련용 원본 영상에 대응하는 GT 영상을 생성하는데 이용되는 마스크를 도시하는 예시적인 도면이다.

[18] 도 9는 훈련용 원본 영상 및 훈련용 리사이징 영상에 기초하여 GT 영상을 생성하는 방법을 설명하기 위한 도면이다.

[19] 도 10 및 도 11은 신경망 모델의 훈련에 이용되는 전경 마스크와 배경 마스크를 생성하는 방법을 설명하기 위한 도면이다.

[20] 도 12는 신경망 모델의 훈련에 이용되는 손실 정보들을 설명하기 위한 도면이다.

[21] 도 13은 일 실시예에 따른 영상 처리 장치의 구성을 도시하는 블록도이다.

발명의 실시를 위한 최선의 형태

[22] 일 실시예에 따른 영상 처리 방법은, 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하는 단계; 상기 리사이징 영상을 미리 훈련된 신경망 모델에 입력하는 단계; 및 상기 신경망 모델에서 출력되는 시프트 맵에 기초하여 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함할 수 있다.

[23] 예시적인 실시예에서, 상기 신경망 모델은, 상기 리사이징 영상을 입력받는 제 1 서브 모델, 상기 제 1 서브 모델로부터 출력되는 특징 맵을 히든 스테이트에 따라 처리하는 컨볼루션 LSTM(long short-term memory) 레이어 및 상기 컨볼루션 LSTM 레이어에서 출력되는 데이터를 처리하여 상기 시프트 맵을 출력하는 제 2 서브 모델을 포함할 수 있다.

[24] 예시적인 실시예에서, 상기 컨볼루션 LSTM 레이어는, 이전 시간대의 원본 영상에 대응하는 리사이징 영상을 처리한 후 업데이트된 상기 히든 스테이트를 이용하여 상기 원본 영상에 대응하는 리사이징 영상을 처리할 수 있다.

[25] 예시적인 실시예에서, 상기 제 1 서브 모델 및 제 2 서브 모델 각각은, 소정의 필터 커널을 이용하여 입력 데이터를 컨볼루션 처리하는 적어도 하나의 컨볼루션 레이어를 포함할 수 있다.

[26] 예시적인 실시예에서, 상기 리사이징 영상의 픽셀들의 제 1 위치 맵 및 제 2 위치 맵이 상기 제 2 서브 모델로 입력되며, 상기 제 1 위치 맵의 첫 번째 행과 마지막 행에는 상기 원본 영상과 상기 리사이징 영상 사이의 세로 크기들 사이의 비율에 대응하는 값이 할당되고, 상기 제 2 위치 맵의 첫 번째 열과 마지막 열에는, 상기 원본 영상과 상기 리사이징 영상 사이의 가로 크기들 사이의 비율에 대응하는 값이 할당될 수 있다.

[27] 예시적인 실시예에서, 상기 제 1 위치 맵의 상기 첫 번째 행으로부터 상기 마지막 행까지 소정 차이를 갖는 값들이 순차적으로 각 행에 할당되고, 상기 제 2 위치 맵의 상기 첫 번째 열로부터 상기 마지막 열까지 소정 차이를 갖는 값들이 순차적으로 각 열에 할당될 수 있다.

[28] 예시적인 실시예에서, 상기 리타겟팅 영상을 획득하는 단계는, 상기 리사이징 영상으로부터 제 1 방향으로의 리타겟팅을 위한 시프트 맵에 기초하여 변환된 영상을, 제 2 방향으로의 리타겟팅을 위한 시프트 맵을 기초로 변환하여 상기 리타겟팅 영상을 획득하는 단계를 포함할 수 있다.

[29] 예시적인 실시예에서, 상기 리타겟팅 영상을 획득하는 단계는, 이전 시간대의 시프트 맵과 현재 시간대의 상기 시프트 맵을 조합하여, 상기 현재 시간대의 시프트 맵을 갱신하는 단계; 및 상기 갱신된 상기 현재 시간대의 시프트 맵에 기초하여, 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함할 수 있다.

[30] 예시적인 실시예에서, GT(ground truth) 영상과 훈련용 리타겟팅 영상의 차이에 마스크를 적용한 결과에 대응하는 제 1 손실 정보가 감소되도록 상기 신경망 모델이 훈련될 수 있다.

[31] 예시적인 실시예에서, 훈련용 리사이징 영상을 제 1 방향으로의 리타겟팅을 위한 시프트 맵에 따라 변환시켜 획득된 영상과 GT 영상의 차이에 제 1 배경 마스크를 적용한 결과에 대응하는 제 2 손실 정보와, 상기 훈련용 리사이징 영상을 상기 제 1 방향으로의 리타겟팅을 위한 시프트 맵 및 제 2 방향으로의 리타겟팅을 위한 시프트 맵에 따라 변환시켜 획득된 훈련용 리타겟팅 영상과 상기 GT 영상의 차이에 제 1 전경 마스크를 적용한 결과에 대응하는 제 3 손실 정보가 감소되도록 상기 신경망 모델이 훈련될 수 있다.

[32] 예시적인 실시예에서, 상기 제 1 전경 마스크는, 훈련용 원본 영상에 대응하는 제 2 전경 마스크를 패딩하여 획득되며, 상기 제 1 배경 마스크는, 상기 제 2 전경 마스크에 대응하는 제 2 배경 마스크와, 상기 제 2 전경 마스크의 크기를

변경시켜 획득한 제 3 전경 마스크에 대응하는 제 3 배경 마스크를 서로 결합하여 획득될 수 있다.

[33] 예시적인 실시예에서, 상기 GT 영상은, 상기 훈련용 원본 영상에 상기 제 2 전경 마스크를 적용하여 획득한 전경 오브젝트와, 상기 훈련용 리사이징 영상에 상기 제 3 배경 마스크를 적용하여 획득한 배경 오브젝트를 결합하여 획득될 수 있다.

[34] 예시적인 실시예에서, 이전 시간대의 훈련용 원본 영상에 대응하는 시프트 맵과 현재 시간대의 훈련용 원본 영상에 대응하는 시프트 맵 사이의 관계에 기초하여 도출되는 제 4 손실 정보가 감소되도록 상기 신경망 모델이 훈련될 수 있다.

[35] 일 실시예에 따른 영상 처리 장치는, 프로세서; 및 미리 훈련된 신경망 모델 및 적어도 하나의 프로그램을 저장하는 메모리를 포함하되, 상기 프로세서는, 상기 메모리에 저장된 상기 적어도 하나의 프로그램이 실행됨에 따라, 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하고, 상기 리사이징 영상을 상기 신경망 모델에 입력하고, 상기 신경망 모델에서 출력되는 시프트 맵에 기초하여 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득할 수 있다.

[36] 예시적인 실시예에서, 상기 영상 처리 장치는, 상기 리타겟팅 영상을 출력하는 디스플레이를 포함하되, 상기 프로세서는, 상기 디스플레이의 종횡비에 기초하여 상기 리사이징 영상의 종횡비를 결정할 수 있다.

[37] 예시적인 실시예에서, 상기 프로세서는, 상기 리사이징 영상으로부터 제 1 방향으로의 리타겟팅을 위한 시프트 맵에 기초하여 변환된 영상을, 제 2 방향으로의 리타겟팅을 위한 시프트 맵을 기초로 변환하여 상기 리타겟팅 영상을 획득할 수 있다.

[38] 예시적인 실시예에서, 상기 프로세서는, 이전 시간대의 시프트 맵과 현재 시간대의 상기 시프트 맵을 조합하여, 상기 현재 시간대의 시프트 맵을 갱신하고, 상기 갱신된 상기 현재 시간대의 시프트 맵에 기초하여, 상기 리사이징 영상을 변환시킬 수 있다.

[39] 예시적인 실시예에서, 상기 프로세서는, 훈련용 원본 영상에 대응하는 훈련용 리사이징 영상을 상기 신경망 모델로 입력하되, 상기 신경망 모델은, 훈련용 리타겟팅 영상과 GT 영상 사이의 차이에 마스크를 적용한 결과에 대응하는 손실 정보가 감소되도록 훈련될 수 있다.

[40] 예시적인 실시예에서, 상기 영상 처리 장치는, 상기 미리 훈련된 신경망 모델의 데이터를 외부 장치로부터 수신하는 통신 모듈을 더 포함할 수 있다.

발명의 실시를 위한 형태

[41] 본 개시는 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고, 이를 상세한 설명을 통해 설명하고자 한다.

그러나, 이는 본 개시를 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[42] 실시예를 설명함에 있어서, 관련된 공지 기술에 대한 구체적인 설명이 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 실시예의 설명 과정에서 이용되는 숫자(예를 들어, 제 1, 제 2 등)는 하나의 구성요소를 다른 구성요소와 구분하기 위한 식별기호에 불과하다.

[43] 또한, 본 명세서에서 일 구성요소가 다른 구성요소와 "연결된다" 거나 "접속된다" 등으로 언급된 때에는, 상기 일 구성요소가 상기 다른 구성요소와 직접 연결되거나 또는 직접 접속될 수도 있지만, 특별히 반대되는 기재가 존재하지 않는 이상, 중간에 또 다른 구성요소를 매개하여 연결되거나 또는 접속될 수도 있다고 이해되어야 할 것이다.

[44] 또한, 본 명세서에서 '~부(유닛)', '모듈' 등으로 표현되는 구성요소는 2개 이상의 구성요소가 하나의 구성요소로 합쳐지거나 또는 하나의 구성요소가 보다 세분화된 기능별로 2개 이상으로 분화될 수도 있다. 또한, 이하에서 설명할 구성요소 각각은 자신이 담당하는 주기능 이외에도 다른 구성요소가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성요소 각각이 담당하는 주기능 중 일부 기능이 다른 구성요소에 의해 전담되어 수행될 수도 있음은 물론이다.

[45] 이하, 본 개시의 기술적 사상에 의한 실시예들을 차례로 상세히 설명한다.

[46] 도 1은 일 실시예에 따른 영상 처리 장치(100)가 적용되는 환경을 설명하기 위한 도면이다.

[47] 일 실시예에 따른 영상 처리 장치(100)는 원본 영상을 획득하고, 원본 영상에 기초하여 리타겟팅 영상을 획득한다. 영상 처리 장치(100)는 소정 종횡비와 상이한 종횡비를 갖는 원본 영상으로부터 소정 종횡비를 갖는 리타겟팅 영상을 생성할 수 있다. 영상 처리 장치(100)는 리타겟팅 영상을 디스플레이 장치(200)를 통해 디스플레이할 수 있다.

[48] 일 실시예에서, 영상 처리 장치(100)는 디스플레이 장치(200) 내에 포함될 수 있다. 또는, 일 실시예에서, 영상 처리 장치(100)는 서버로 구현될 수 있고, 이 경우, 영상 처리 장치(100)는 리타겟팅 영상을 네트워크를 통해 디스플레이 장치(200)로 전송할 수도 있다. 일 실시예에서 네트워크는 유선 네트워크와 무선 네트워크를 포함할 수 있으며, 구체적으로, 근거리 네트워크(LAN: Local Area Network), 도시권 네트워크(MAN: Metropolitan Area Network), 광역 네트워크(WAN: Wide Area Network) 등의 다양한 네트워크를 포함할 수 있다. 또한, 네트워크는 공지의 월드 와이드 웹(WWW: World Wide Web)을 포함할 수도 있다. 그러나, 네트워크는 상기 열거된 네트워크에 국한되지 않고, 공지의 무선 데이터 네트워크나 공지의 전화 네트워크, 공지의 유무선 텔레비전 네트워크를 적어도 일부로 포함할 수도 있다.

[49] 일 실시예에 따른 영상 처리 장치(100)는 리타겟팅 영상을 생성할 때, 원본 영상 내 중요한 부분이 리타겟팅 영상에서도 원본 영상 내에서의 종횡비를 최대한 유지하도록 할 수 있다.

[50] 도 2는 일 실시예에 따른 영상 처리 방법을 설명하기 위한 순서도이고, 도 3은 일 실시예에 따른 리타겟팅 과정을 설명하기 위한 도면이다.

[51] S210 단계에서, 영상 처리 장치(100)는 소정 종횡비에 따라 원본 영상(10)으로부터 크기가 변경된 리사이징 영상(20)을 획득한다.

[52] 리사이징 영상(20)의 크기는 원본 영상(10)의 크기보다 클 수 있다. 예를 들어, 리사이징 영상(20)의 세로 방향 크기는 원본 영상(10)의 세로 방향 크기와 동일하되, 리사이징 영상(20)의 가로 방향 크기는 원본 영상(10)의 가로 방향 크기보다 클 수 있다.

[53] 상기 소정 종횡비는 후술하는 리타겟팅 영상을 출력하는 디스플레이의 종횡비에 기초하여 결정될 수 있다. 예를 들어, 소정 종횡비는 리타겟팅 영상을 출력하는 디스플레이가 가지는 종횡비와 동일하게 결정될 수 있다.

[54] 영상 처리 장치(100)는 원본 영상(10)을 스케일링하여 크기가 변경된 리사이징 영상(20)을 획득할 수 있다. 일 예에서, 영상 처리 장치(100)는 원본 영상(10) 내 픽셀들을 보간(interpolation)하여 크기가 변경된 리사이징 영상(20)을 획득할 수 있다.

[55] S220 단계에서, 영상 처리 장치(100)는 리사이징 영상(20)을 미리 훈련된 신경망 모델(300)에 입력한다. 신경망 모델(300)은 예를 들어, 복수의 레이어로 이루어진 딥러닝(deep learning) 모델일 수 있다. 신경망 모델(300)의 구조 및 훈련 방법에 대해서는 도 8 이하를 참조하여 상세히 설명한다.

[56] S230 단계에서, 영상 처리 장치(100)는 신경망 모델(300)에서 출력되는 시프트 맵(90)을 획득한다. 시프트 맵(90)은 리사이징 영상(20) 내 픽셀들의 위치를 변경하기 위한 정보를 포함한다. 예를 들어, 시프트 맵(90)은 리사이징 영상(20) 내 각 픽셀들이 리타겟팅 영상 내 어느 위치로 이동되어야 하는지의 정보를 포함할 수 있다. 시프트 맵(90)은 리사이징 영상(20)의 크기와 동일한 크기를 가질 수 있다.

[57] S240 단계에서, 영상 처리 장치(100)는 시프트 맵(90)에 기초하여 리사이징 영상(20)으로부터 변환된 리타겟팅 영상을 획득한다.

[58] 일 실시예에 따른 신경망 모델(300)은, 신경망 모델(300)은 출력되는 시프트 맵(90)에 따라 리사이징 영상(20)을 리타겟팅하더라도 원본 영상(10) 내 중요 부분의 비율이 리타겟팅 영상 내에서도 유지되도록 훈련될 수 있다.

[59] 도 4는 일 실시예에 따른 신경망 모델(300)의 구조를 설명하기 위한 도면이다.

[60] 도 4를 참조하면, 신경망 모델(300)은 제 1 서브 모델(310), 컨볼루션 LSTM 레이어(320) 및 제 2 서브 모델(330)을 포함할 수 있다.

[61] 제 1 서브 모델(310)은 리사이징 영상(20)을 입력받아 처리한다. 제 1 서브 모델(310)의 출력 데이터, 예를 들어, 특징 맵(feature map)은 컨볼루션 LSTM(long

short-term memory) 레이어로 입력된다. 그리고, 컨볼루션 LSTM 레이어(320)에서 출력되는 데이터는 제 2 서브 모델(330)로 입력된다. 제 2 서브 모델(330)에서는 리사이징 영상(20)의 리타겟팅을 위한 시프트 맵(90)이 출력된다.

[62] 일 예에서, 제 2 서브 모델(330)에서는 리사이징 영상(20) 내 전경 오브젝트를 추출하기 위한 전경 마스크(92), 리사이징 영상(20)을 제 1 방향을 따라 리타겟팅하기 위한 제 1 시프트 맵(94) 및 리사이징 영상(20)을 제 2 방향을 따라 리타겟팅하기 위한 제 2 시프트 맵(96)이 출력될 수 있다.

[63] 예를 들어, 상기 제 1 시프트 맵(94)은 리사이징 영상(20) 내 픽셀들을 세로 방향으로 리타겟팅하는데 이용되며, 제 2 시프트 맵(96)은 리사이징 영상(20) 내 픽셀들을 가로 방향으로 리타겟팅하는데 이용될 수 있다. 일 실시예에서, 원본 영상(10)의 종횡비를 변환하기 위해 원본 영상(10)의 세로 크기만 변경시키고자 하는 경우, 제 2 서브 모델(330)은 제 1 시프트 맵(94)만을 출력하도록 구성될 수 있다. 반대로, 원본 영상(10)의 종횡비를 변환하기 위해 원본 영상(10)의 가로 크기만 변경시키고자 하는 경우, 제 2 서브 모델(330)은 제 2 시프트 맵(96)만을 출력하도록 구성될 수 있다.

[64] 제 2 서브 모델(330)에서 출력되는 전경 마스크(92)는 신경망 모델(300)이 제대로 훈련되었는지를 평가하기 위한 것으로서, 구현예에 따라, 제 2 서브 모델(330)은 전경 마스크(92)가 출력되지 않도록 구성될 수도 있다.

[65] 컨볼루션 LSTM 레이어(320)는 제 1 서브 모델(310)에서 출력되는 데이터를 입력받아 히든 스테이트(hidden state)에 따라 처리하고, 처리된 결과를 제 2 서브 모델(330)로 전달한다. 컨볼루션 LSTM 레이어(320) 내 히든 스테이트는 제 1 서브 모델(310)에서 출력되는 데이터를 처리한 후, 업데이트되고, 업데이트된 히든 스테이트는 다음 시간대의 영상 처리를 위해 이용될 수 있다.

[66] 예를 들어, t 시간 대의 원본 영상(10)에 대응하는 리사이징 영상(20)이 제 1 서브 모델(310)에서 처리되고, 제 1 서브 모델(310)에서 출력되는 데이터는 컨볼루션 LSTM 레이어(320)에서 히든 스테이트에 따라 처리된다. 여기서, t 시간대의 리사이징 영상(20)을 처리하기 위해 이용되는 히든 스테이트는, 컨볼루션 LSTM 레이어(320)에서 t-1 시간대의 리사이징 영상(20)을 처리한 후 업데이트된 히든 스테이트일 수 있다. t 시간 대의 리사이징 영상(20)에 대해 제 1 서브 모델(310)에서 출력되는 데이터가 컨볼루션 LSTM 레이어(320)에서 히든 스테이트에 따라 처리되면, 히든 스테이트는 업데이트되고, 업데이트된 히든 스테이트는 t+1 시간대의 리사이징 영상(20)을 처리하는데 이용된다.

[67] 즉, 일 실시예에 따른 신경망 모델(300)에서, 현재 시간대의 리사이징 영상(20)을 처리하는데 이용된 컨볼루션 LSTM 레이어(320)의 히든 스테이트가, 다음 시간대의 리사이징 영상(20)을 처리하는데 이용되기 때문에 여러 프레임들로 구성된 비디오를 리타겟팅하는데 있어, 프레임 간 연속성이 유지될 수 있다.

[68] 도 5는 일 실시예에 따른 신경망 모델(300)의 구조를 상세히 설명하기 위한

도면이다.

[69] 도 5를 참조하면, 제 1 서브 모델(310)은 복수의 컨볼루션 레이어(convolution layer)를 포함할 수 있다. 각각의 컨볼루션 레이어는 미리 결정된 크기 및 개수의 필터 커널을 이용하여 입력 데이터를 컨볼루션 처리한 후, 컨볼루션 처리된 출력 데이터를 다음 레이어로 전달한다. 도 5는 제 1 서브 모델(310)이 5개의 컨볼루션 레이어를 포함하는 것으로 도시하고 있으나, 컨볼루션 레이어의 개수는 당업자에게 자명한 범위 내에서 다양하게 변경될 수 있다. 또한, 도 5에는 도시되어 있지 않지만, 제 1 서브 모델(310) 내 컨볼루션 레이어들 중 적어도 일부의 앞 또는 뒤에는 활성화 레이어 및/또는 풀링(pooling) 레이어가 위치할 수 있다.

[70] 어느 하나의 컨볼루션 레이어의 앞에 활성화 레이어 및/또는 풀링(pooling) 레이어가 위치하는 경우, 활성화 레이어 및/또는 풀링(pooling) 레이어의 출력 데이터는 상기 어느 하나의 컨볼루션 레이어로 입력될 수 있다. 또한, 어느 하나의 컨볼루션 레이어의 뒤에 활성화 레이어 및/또는 풀링(pooling) 레이어가 위치하는 경우, 상기 어느 하나의 컨볼루션 레이어의 출력 데이터는 활성화 레이어 및/또는 풀링(pooling) 레이어로 입력될 수 있다.

[71] 활성화 레이어는 이전 레이어의 출력 결과에 대해 비선형(Non-linear) 특성을 부여할 수 있다. 활성화 레이어는 활성화 함수를 이용할 수 있다. 활성화 함수는 시그모이드 함수(sigmoid function), Tanh 함수, ReLU(Rectified Linear Unit) 함수 등을 포함할 수 있으나, 이에 한정되는 것은 아니다.

[72] 제 2 서브 모델(330)은 복수의 디컨볼루션 레이어(deconvolution layer)를 포함할 수 있다. 각각의 디컨볼루션 레이어는 미리 결정된 크기 및 개수의 필터 커널을 이용하여 입력 데이터를 컨볼루션 처리한 후, 컨볼루션 처리된 출력 데이터를 다음 레이어로 전달한다. 도 5는 제 2 서브 모델(330)이 최종적으로 출력되는 데이터 각각을 위하여 5개의 디컨볼루션 레이어를 포함하는 것으로 도시하고 있으나, 디컨볼루션 레이어의 개수는 당업자에게 자명한 범위 내에서 다양하게 변경될 수 있다.

[73] 앞서 설명한 바와 같이, 제 2 서브 모델(330)에서 전경 마스크(92), 제 1 시프트 맵(94) 및 제 2 시프트 맵(96)이 출력되는 경우, 전경 마스크(92), 제 1 시프트 맵(94) 및 제 2 시프트 맵(96) 각각의 출력을 위한 디컨볼루션 레이어들이 서로 병렬로 배치될 수 있다. 제 2 서브 모델(330)에서 제 1 시프트 맵(94)과 제 2 시프트 맵(96)만이 출력되는 경우, 전경 마스크(92)의 출력을 위한 디컨볼루션 레이어(dconv4-1, dconv5-1)는 제 2 서브 모델(330)에 포함되지 않을 수 있다. 마찬가지로, 제 2 서브 모델(330)에서 제 2 시프트 맵(96)만이 출력되는 경우, 전경 마스크(92)의 출력을 위한 디컨볼루션 레이어(dconv4-1, dconv5-1) 그리고, 제 1 시프트 맵(94)의 출력을 위한 디컨볼루션 레이어(dconv4-2, dconv5-2)가 제 2 서브 모델(330)에 포함되지 않을 수 있다.

[74] 또한, 도 5에는 도시되어 있지 않지만, 제 2 서브 모델(330) 내 디컨볼루션

레이어들 중 적어도 일부의 앞 또는 뒤에는 활성화 레이어 및/또는 풀링 레이어가 위치할 수 있다.

[75] 일 실시예에서, 제 2 서브 모델(330)에 포함된 디컨볼루션 레이어 중 적어도 하나의 디컨볼루션 레이어로 리사이징 영상(20) 내 픽셀들의 위치를 나타내는 위치 맵이 입력될 수도 있다.

[76] 도 6 및 도 7은 제 2 서브 모델(330)에 입력되는 위치 맵을 도시하는 예시적인 도면이다.

[77] 제 2 서브 모델(330)에 포함된 디컨볼루션 레이어 중 적어도 하나의 디컨볼루션 레이어로 제 1 위치 맵(610, 710)과 제 2 위치 맵(630, 730)이 입력될 수 있는데, 제 1 위치 맵(610, 710)은 리사이징 영상(20) 내 픽셀들의 세로 방향 위치를 나타내고, 제 2 위치 맵(630, 730)은 리사이징 영상(20) 내 픽셀들의 가로 방향 위치를 나타낼 수 있다.

[78] 일 실시예에서, 영상 처리 장치(100)는 제 1 위치 맵(610, 710) 및 제 2 위치 맵(630, 730을 생성하는 경우, 제 1 위치 맵(610, 710)과 제 2 위치 맵(630, 730)이 원본 영상(10)과 리사이징 영상(20) 사이의 크기 비율 정보를 포함하도록 할 수 있다.

[79] 일 예에서, 영상 처리 장치(100)는 제 1 위치 맵(610, 710)의 첫 번째 행(611, 711)과 마지막 행(619, 719)에 포함된 픽셀들에, 원본 영상(10)과 리사이징 영상(20)의 세로 크기들 사이의 비율에 대응하는 값을 할당할 수 있다. 그리고, 영상 처리 장치(100)는 제 2 위치 맵(630, 730)의 첫 번째 열(631, 731)과 마지막 열(639, 739)에 포함된 픽셀들에, 원본 영상(10)과 리사이징 영상(20)의 가로 크기들 사이의 비율에 대응하는 값을 할당할 수 있다.

[80] 도 6은 원본 영상(10)과 리사이징 영상(20)의 크기 및 종횡비가 동일할 때의 제 1 위치 맵(610) 및 제 2 위치 맵(630)을 도시하고 있다.

[81] 도 6을 참조하면, 제 1 위치 맵(610)의 첫 번째 행(611)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 세로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 -로 결정될 수 있다. 그리고, 제 1 위치 맵(610)의 마지막 행(619)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 세로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 +로 결정될 수 있다. 제 1 위치 맵(610)의 첫 번째 행(611)으로부터 마지막 행(619)까지 소정 차이를 갖는 값들이 순차적으로 각 행의 픽셀들에 할당될 수 있다. 예를 들어, 제 1 위치 맵(610)의 두 번째 행의 픽셀들에는, 첫 번째 행(611)의 픽셀들에 할당된 -1보다 a만큼 큰 값들이 할당될 수 있고, 제 1 위치 맵(610)의 세 번째 행의 픽셀들에는, 두 번째 행의 픽셀들에 할당된 값보다 a만큼 큰 값들이 할당될 수 있다.

[82] 제 2 위치 맵(630)의 첫 번째 열(631)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 가로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 -로 결정될 수 있다. 그리고, 제 2 위치 맵(630)의 마지막 열(639)에

포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 가로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 +로 결정될 수 있다. 제 2 위치 맵(630)의 첫 번째 열(631)으로부터 마지막 열(639)까지 소정 차이를 갖는 값들이 순차적으로 각 열의 픽셀들에 할당될 수 있다. 예를 들어, 제 2 위치 맵(630)의 두 번째 열의 픽셀들에는, 첫 번째 열(631)의 픽셀들에 할당된 -1보다 a만큼 큰 값들이 할당될 수 있고, 제 2 위치 맵(630)의 세 번째 열의 픽셀들에는, 두 번째 열의 픽셀들에 할당된 값보다 a만큼 큰 값들이 할당될 수 있다.

[83] 도 7은 리사이징 영상(20)의 세로 크기와 원본 영상(10)의 세로 크기가 동일하고, 리사이징 영상(20)의 가로 크기가 원본 영상(10)의 가로 크기보다 1.5배 클 때의 제 1 위치 맵(710) 및 제 2 위치 맵(730)을 도시하고 있다.

[84] 도 7을 참조하면, 제 1 위치 맵(710)의 첫 번째 행(711)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 세로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 -로 결정될 수 있다. 그리고, 제 1 위치 맵(710)의 마지막 행(719)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 세로 크기들 사이의 비율에 대응하는 값, 즉 1이 할당되되, 그 부호가 +로 결정될 수 있다. 제 1 위치 맵(710)의 첫 번째 행(711)으로부터 마지막 행(719)까지 소정 차이를 갖는 값들이 순차적으로 각 행의 픽셀들에 할당될 수 있다. 예를 들어, 제 1 위치 맵(710)의 두 번째 행의 픽셀들에는, 첫 번째 행(711)의 픽셀들에 할당된 -1보다 a만큼 큰 값들이 할당될 수 있고, 제 1 위치 맵(710)의 세 번째 행의 픽셀들에는, 두 번째 행의 픽셀들에 할당된 값보다 a만큼 큰 값들이 할당될 수 있다.

[85] 제 2 위치 맵(730)의 첫 번째 열(731)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 가로 크기들 사이의 비율에 대응하는 값, 즉 1.5가 할당되되, 그 부호가 -로 결정될 수 있다. 그리고, 제 2 위치 맵(730)의 마지막 열(739)에 포함된 픽셀들에는 원본 영상(10)과 리사이징 영상(20)의 가로 크기들 사이의 비율에 대응하는 값, 즉 1.5가 할당되되, 그 부호가 +로 결정될 수 있다. 제 2 위치 맵(730)의 첫 번째 열(731)으로부터 마지막 열(739)까지 소정 차이를 갖는 값들이 순차적으로 각 행의 픽셀들에 할당될 수 있다. 예를 들어, 제 2 위치 맵(730)의 두 번째 열의 픽셀들에는, 첫 번째 열(731)의 픽셀들에 할당된 -1.5보다 b만큼 큰 값들이 할당될 수 있고, 제 2 위치 맵(730)의 세 번째 열의 픽셀들에는, 두 번째 열의 픽셀들에 할당된 값보다 b만큼 큰 값들이 할당될 수 있다.

[86] 도 6의 제 2 위치 맵(630)의 각각의 열을 따라 할당된 픽셀 값들의 차이인 a는 도 7의 제 2 위치 맵(730)의 각각의 열을 따라 할당된 픽셀 값들의 차이인 b보다 작을 수 있다. 왜냐하면, 도 6의 제 2 위치 맵(630)과 도 7의 제 2 위치 맵(730)의 가로 크기가 서로 동일할 때, 각각의 제 2 위치 맵(630, 730)의 첫 번째 열(631, 731)과 마지막 열(639, 739)에 할당되는 픽셀 값이 고정되어 있는 상태에서 각 열에 픽셀 값들이 동일 차이의 간격으로 할당되기 때문이다.

[87] 도 6 및 도 7과 관련하여, 제 1 위치 맵(610, 710)의 첫 번째 행(611, 711)의 픽셀

값들이 -의 부호를 갖고, 제 1 위치 맵(610, 710)의 마지막 행(619, 719)의 픽셀 값들이 +의 부호를 갖는 것으로 설명하였으나, 일 예에서, 제 1 위치 맵(610, 710)의 첫 번째 행(611, 711)의 픽셀 값들이 +의 부호를 갖고, 제 1 위치 맵(610, 710)의 마지막 행(619, 719)의 픽셀 값들이 -의 부호를 가질 수 있다. 마찬가지로, 제 2 위치 맵(630, 730)의 첫 번째 열(631, 731)의 픽셀 값들이 -의 부호를 갖고, 제 2 위치 맵(630, 730)의 마지막 열(639, 739)의 픽셀 값들이 +의 부호를 갖는 것으로 설명하였으나, 일 예에서, 제 2 위치 맵(630, 730)의 첫 번째 열(631, 731)의 픽셀 값들이 +의 부호를 갖고, 제 2 위치 맵(630, 730)의 마지막 열(639, 739)의 픽셀 값들이 -의 부호를 가질 수 있다.

[88] 일 실시예에서, 영상 처리 장치(100)는 여러 시간대의 원본 영상(10)들에 대응하는 리타겟팅 영상들 사이의 연속성을 유지하기 위하여, 이전 시간대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력되는 시프트 맵(90)과, 현재 시간대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력되는 시프트 맵(90)을 결합하여 현재 시간대의 리사이징 영상(20)을 리타겟팅할 수도 있다.

[89] 예를 들어, 영상 처리 장치(100)는 t-1 시간 대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력되는 시프트 맵(90)과, t 시간 대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력되는 시프트 맵(90)을 결합한 후, t 시간대의 리사이징 영상(20)을 리타겟팅할 수 있다. 각 시간대별 시프트 맵(90)의 갱신 과정은 아래의 수학식 1로 표현될 수 있다.

[90] [수식1]

$$Sb'(t) = (t/(t+1))*Sb'(t-1)+(1/(t+1))*Sb(t)$$

$$Sf'(t) = (t/(t+1))*Sf'(t-1)+(1/(t+1))*Sf(t)$$

[91] 상기 수학식 1에서, Sb(t)는 t 시간대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력된 제 1 시프트 맵(94), Sb'(t)은 Sb(t)로부터 업데이트된 제 1 시프트 맵(94), Sb'(t-1)은 t-1 시간대 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력된 후, 업데이트된 제 1 시프트 맵(94), Sf(t)는 t 시간대의 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력된 제 2 시프트 맵(96), Sf'(t)은 Sf(t)로부터 업데이트된 제 2 시프트 맵(96), Sf'(t-1)은 t-1 시간대 리사이징 영상(20)에 대응하여 신경망 모델(300)에서 출력된 후, 업데이트된 제 2 시프트 맵(96)을 의미한다. 상기 제 1 시프트 맵(94)은 리사이징 영상(20)을 세로 방향으로 리타겟팅하기 위한 시프트 맵, 제 2 시프트 맵(96)은 리사이징 영상(20)을 가로 방향으로 리타겟팅하기 위한 시프트 맵을 나타낼 수 있다.

[92] 이하에서는, 도 8 내지 도 12를 참조하여, 신경망 모델(300)을 훈련시키는 방법에 대해 설명한다.

[93] 도 8은 훈련용 원본 영상(910)에 대응하는 GT 영상(930)을 생성하는데 이용되는 마스크를 도시하는 예시적인 도면이고, 도 9는 훈련용 원본 영상(910)

및 훈련용 리사이징 영상(920)에 기초하여 GT 영상(930)을 생성하는 방법을 설명하기 위한 도면이다.

[94] 일 실시예에서, 마스크는 바이너리 마스크일 수 있다. 마스크 내 포함된 픽셀들은 두 가지 값 중 어느 하나의 값을 가질 수 있다. 전경 마스크는 영상 내 전경 오브젝트를 추출하기 위한 마스크로서, 예를 들어, 영상 내 중요 영역만을 추출하기 위해 이용될 수 있다. 배경 마스크는 영상 내 배경 오브젝트를 추출하기 위한 마스크로서, 예를 들어, 영상 내 비중요 영역만을 추출하기 위해 이용될 수 있다.

[95] 영상 처리 장치(100)는 신경망 모델(300)의 훈련을 위해 훈련용 원본 영상(910)에 대응하는 GT(ground truth) 영상(930)을 생성할 수 있다. GT 영상(930)은 훈련용 리사이징 영상(920)과 동일한 종횡비 및 크기를 가질 수 있다.

[96] 영상 처리 장치(100)는 GT 영상(930)을 생성하기 위해 도 8에 도시된 것과 같은 제 1 전경 마스크(32) 및 제 2 배경 마스크(44)를 획득할 수 있다.

[97] 영상 처리 장치(100)는 훈련용 원본 영상(910)에 대응하는 제 1 전경 마스크(32)를 획득할 수 있다. 영상 처리 장치(100)는 훈련용 원본 영상(910) 및 그에 대응하는 제 1 전경 마스크(32)를 외부의 장치로부터 수신하거나, 관리자로부터 입력받을 수 있다. 훈련용 원본 영상(910)과 제 1 전경 마스크(32)는 동일한 크기 및 동일한 종횡비를 가질 수 있다. 일 실시예에서, 영상 처리 장치(100)는 훈련용 원본 영상(910)을 미리 저장된 알고리즘에 따라 분석하여, 훈련용 원본 영상(910) 내 중요 부분을 식별하고, 식별된 부분을 추출하기 위한 제 1 전경 마스크(32)를 생성할 수도 있다.

[98] 영상 처리 장치(100)는 소정 종횡비에 따라 제 1 전경 마스크(32)를 스트레칭(stretching)한 제 2 전경 마스크(34)를 획득할 수 있다. 영상 처리 장치(100)는 제 1 전경 마스크(32)를 보간 처리하여 제 2 전경 마스크(34)를 생성할 수 있다.

[99] 영상 처리 장치(100)는 제 1 전경 마스크(32)의 픽셀 값들을 반전시켜 제 1 전경 마스크(32)에 대응하는 제 1 배경 마스크(42)를 획득할 수 있다. 제 1 배경 마스크(42)는 후술하는 바와 같이, 제 3 배경 마스크(46)를 생성하는데 이용될 수 있다. 그리고, 영상 처리 장치(100)는 제 2 전경 마스크(34)의 픽셀 값들을 반전시켜 제 2 전경 마스크(34)에 대응하는 제 2 배경 마스크(44)를 획득할 수 있다. 여기서, 전경 마스크의 픽셀 값들을 반전시킨다는 것은, 전경 마스크 내 픽셀들에 할당된 어느 하나의 값을, 마스크에 할당 가능한 2가지 값 중 다른 하나로 변경하는 것을 의미할 수 있다. 예를 들어, 제 1 전경 마스크(32)에서 1의 값을 가지는 픽셀은, 제 1 배경 마스크(42)에서 0의 값을 가질 수 있다. 일 실시예에서, 상기 제 2 배경 마스크(44)는 제 1 배경 마스크(42)의 크기를 변경시켜 획득될 수도 있다.

[100] 도 9를 참조하면, 영상 처리 장치(100)는 훈련용 원본 영상(910)에 제 1 전경

마스크(32)를 적용하여 훈련용 원본 영상(910)으로부터 전경 오브젝트(915)를 추출하고, 훈련용 리사이징 영상(920)에 제 2 배경 마스크(44)를 적용하여 훈련용 리사이징 영상(920)으로부터 배경 오브젝트(927)를 추출한다. 그리고, 영상 처리 장치(100)는 훈련용 원본 영상(910)에서 추출되는 전경 오브젝트(915)와 훈련용 리사이징 영상(920)에서 추출되는 배경 오브젝트(927)를 결합하여 GT 영상(930)을 생성한다. 여기서, 결합한다는 것은, 훈련용 원본 영상(910)에서 추출되는 오브젝트(915)와 훈련용 리사이징 영상(920)에 추출되는 오브젝트(927)를 합하여 하나의 GT 영상(930)을 생성한다는 것을 의미할 수 있다.

[101] 훈련용 리사이징 영상(920) 내 포함된 전경 오브젝트(925)는, 훈련용 원본 영상(910) 내 포함된 전경 오브젝트(915) 대비 크기 비율이 변경된 것이므로, 도 9에 도시된 바와 같이, GT 영상(930) 내에는 빈 영역(935)이 존재하게 된다. 다시 말하면, 제 2 배경 마스크(44)를 훈련용 리사이징 영상(920)에 적용하였을 때, 훈련용 리사이징 영상(920)에 포함된 전경 오브젝트(925)는 훈련용 리사이징 영상(920)에 포함된 그대로 추출되지 않고, 제 2 배경 마스크(44)를 거쳐 예를 들어, 0의 픽셀 값을 가지게 된다. 훈련용 원본 영상(910)에 포함된 전경 오브젝트(915)의 크기는 훈련용 리사이징 영상(920)에 포함된 전경 오브젝트(925)의 크기보다 작으므로, 훈련용 원본 영상(910)으로부터 추출된 전경 오브젝트(915)를 훈련용 리사이징 영상(920)으로부터 추출되는 배경 오브젝트(927)에 결합시키는 경우, 0의 픽셀 값을 가진 부분은 그대로 빈 공간(935)이 되는 것이다.

[102] 영상 처리 장치(100)는 신경망 모델(300)의 훈련을 위해, 훈련용 리사이징 영상(920)을 신경망 모델(300)로 입력한다. 신경망 모델(300)은 훈련용 리사이징 영상(920)에 대응하여 시프트 맵을 출력한다. 그리고, 신경망 모델(300)은 시프트 맵에 기초하여 훈련용 리사이징 영상(920)으로부터 변환된 훈련용 리타겟팅 영상과 GT 영상(930) 사이의 차이가 감소되도록 훈련할 수 있다. 일 예에서, 신경망 모델(300)은 훈련용 리타겟팅 영상과 GT 영상(930) 사이의 차이에 마스크를 적용한 결과에 대응하는 손실 정보가 감소되도록 훈련될 수 있다. 상기 손실 정보를 도출하기 위한 마스크는 앞서 도 8의 제 1 전경 마스크(32), 제 1 배경 마스크(42) 및 제 2 배경 마스크(44)에 기초하여 생성될 수 있는데, 이에 대해서는 도 10 및 도 11을 참조하여 설명한다.

[103] 도 10 및 도 11은 신경망 모델(300)의 훈련에 이용되는 전경 마스크와 배경 마스크를 생성하는 방법을 설명하기 위한 도면이다.

[104] 도 10을 참조하면, 영상 처리 장치(100)는 제 1 전경 마스크(32)를 패딩(padding)하여 제 3 전경 마스크(36)를 생성할 수 있다. 영상 처리 장치(100)는 제 3 전경 마스크(36)의 크기 및 종횡비가 GT 영상(930)의 크기 및 종횡비와 동일하도록 제 1 전경 마스크(32)를 패딩할 수 있다. 패딩에 따라 새롭게 생성되는 픽셀들에는, 제 1 전경 마스크(32)에 포함된 주변 픽셀 값들과

동일한 값들이 할당될 수 있다. 제 3 전경 마스크(36)는 GT 영상(930)으로부터 전경 오브젝트를 추출하는데 이용될 수 있다.

[105] 도 11을 참조하면, 영상 처리 장치(100)는 제 1 배경 마스크(42)와 제 2 배경 마스크(44)를 결합하여 제 3 배경 마스크(46)를 생성할 수 있다. 영상 처리 장치(100)는 제 1 배경 마스크(42)와 제 2 배경 마스크(44)를 합하여 하나의 제 3 배경 마스크(46)를 생성할 수 있다. 제 3 배경 마스크(46)는 GT 영상(930)으로부터 배경 오브젝트를 추출하는데 이용될 수 있다.

[106] GT 영상(930), 제 3 전경 마스크(36) 및 제 3 배경 마스크(46)를 이용하여 신경망 모델(300)을 훈련시키는 방법은 도 12를 참조하여 설명한다.

[107] 도 12는 신경망 모델(300)의 훈련에 이용되는 손실 정보들을 설명하기 위한 도면이다.

[108] 영상 처리 장치(100)는 훈련용 원본 영상(910)에 대응하는 훈련용 리사이징 영상(920)을 신경망 모델(300)로 입력하고, 신경망 모델(300)에서는 훈련용 리사이징 영상(920)의 제 1 방향(예를 들어, 세로 방향)으로의 리타겟팅을 위한 제 1 시프트 맵(994)과 훈련용 리사이징 영상(920)의 제 2 방향(예를 들어, 가로 방향)으로의 리타겟팅을 위한 제 2 시프트 맵(996)이 출력된다.

[109] 훈련용 리사이징 영상(920)은 제 1 시프트 맵(994)에 기초하여 리타겟팅되고, 리타겟팅된 영상과 GT 영상(930) 사이의 차이에 제 3 배경 마스크(46)가 적용된다. 제 3 배경 마스크(46)가 적용된 결과에 대응하는 제 1 손실 정보(1210)가 추출될 수 있다. 신경망 모델(300)은 제 1 손실 정보(1210)가 감소되도록 훈련될 수 있다. 일 예에서, 신경망 모델(300)은 제 1 손실 정보(1210)가 최소화되도록 훈련될 수 있다. 제 1 손실 정보(1210)는 아래의 수학식 2로 계산될 수 있다.

[110] [수식2]

$$Lb = \Sigma \left| \left(I(x, y+Sb(x, y)) - GT(x, y) \right) \cdot Mb(x, y) \right|$$

[111] 상기 수학식 2에서, Lb는 제 1 손실 정보(1210), (x, y)는 영상 내 픽셀의 위치, I는 훈련용 리사이징 영상(920), Sb는 제 1 시프트 맵(994), GT는 GT 영상(930), Mb는 제 3 배경 마스크(46)를 나타낸다.

[112] 또한, 훈련용 리사이징 영상(920)은 제 1 시프트 맵(994) 및 제 2 시프트 맵(996)에 기초하여 리타겟팅되고, 훈련용 리사이징 영상(920)이 리타겟팅되어 생성된 훈련용 리타겟팅 영상과 GT 영상(930) 사이의 차이에 제 3 전경 마스크(36)가 적용된다. 제 3 전경 마스크(36)가 적용된 결과에 대응하는 제 2 손실 정보(1230)가 추출될 수 있다. 신경망 모델(300)은 제 2 손실 정보(1230)가 감소되도록 훈련될 수 있다. 일 예에서, 신경망 모델(300)은 제 2 손실 정보(1230)가 최소화되도록 훈련될 수 있다. 제 2 손실 정보(1230)는 아래의 수학식 3으로 계산될 수 있다.

[113]

[수식3]

$$Lf = \sum \left| \left(I\left(x + Sf(x, y), y + Sb(x, y)\right) - GT(x,y)\right) \cdot Mf(x, y) \right|$$

[114] 상기 수학식 3에서, Lf는 제 2 손실 정보(1230), (x, y)는 영상 내 픽셀의 위치, I는 훈련용 리사이징 영상(920), Sb는 제 1 시프트 맵(994), Sf는 제 2 시프트 맵(996), GT는 GT 영상(930), Mf는 제 3 전경 마스크(36)를 나타낸다.

[115] 일 실시예에서, 신경망 모델(300)은 이전 시간대의 훈련용 리사이징 영상(920)에 대응하는 시프트 맵과 현재 시간대의 훈련용 리사이징 영상(920)에 대응하는 시프트 맵 사이의 관계에 기초하여 도출되는 제 3 손실 정보가 감소되도록 훈련될 수 있다. 일 예에서, 신경망 모델(300)은 제 3 손실 정보가 최소화되도록 훈련될 수도 있다.

[116] 제 3 손실 정보는 여러 프레임들로 구성된 비디오에서 리타겟팅된 프레임들 사이의 연속성을 유지하기 위해 산출된다. 제 3 손실 정보는 Flow consistency loss에 대응할 수 있다. 제 3 손실 정보는 아래의 수학식 4에 따라 산출될 수 있다.

[117] [수식4]

$$Lw = \sum \left| \left(warp(Sb(t)) - Sb(t+1)\right) \cdot Mo \right| + \sum \left| \left(warp(Sf(t)) - Sf(t+1)\right) \cdot Mo \right|$$

[118] 상기 수학식 4에서 warp는 플로우넷(flownet)을 이용한 워핑 함수를 나타내고, Sb(t)는 t 시간 대의 훈련용 리사이징 영상(920)을 입력받아 신경망 모델(300)이 출력하는 제 1 시프트 맵(994), Sb(t+1)은 t+1 시간 대의 훈련용 리사이징 영상(920)을 입력받아 신경망 모델(300)이 출력하는 제 1 시프트 맵(994), Sf(t)는 t 시간 대의 훈련용 리사이징 영상(920)을 입력받아 신경망 모델(300)이 출력하는 제 2 시프트 맵(996), Sf(t+1)은 t+1 시간 대의 훈련용 리사이징 영상(920)을 입력받아 신경망 모델(300)이 출력하는 제 2 시프트 맵(996)을 나타내고, Mo는 오클루젼 마스크(occlusion mask)로서 아래의 수학식 5로 산출된다.

[119] [수식5]

$$Mo = \exp\left(-a \cdot \left| warp(I(t)) - I(t+1) \right| \right)$$

[120] 상기 수학식 5에서 a는 미리 결정된 가중치를 나타내고, I(t)는 t 시간대의 훈련용 원본 영상(910)에 대응하는 훈련용 리사이징 영상(920), I(t+1)은 t+1 시간대의 훈련용 원본 영상(910)에 대응하는 훈련용 리사이징 영상(920)을 나타낸다.

[121] 신경망 모델(300)은 제 1 손실 정보(1210), 제 2 손실 정보(1230) 및 제 3 손실 정보를 조합한 최종 손실 정보가 감소되도록 훈련될 수 있다. 최종 손실 정보는 아래의 수학식 6에 따라 결정될 수 있다.

[122] [수식6]

$$\text{최종 손실 정보} = a \cdot \text{제 1 손실 정보} + b \cdot \text{제 2 손실 정보} + c \cdot \text{제 3 손실 정보}$$

[123] 상기 수학식 6에서, a, b, c는 각 손실 정보에 적용되는 미리 결정된 가중치이다.

[124] 일 실시예에 따른 신경망 모델(300)은, 훈련용 리사이징 영상(920)을 제 1

시프트 맵(994)에 기초하여 세로 방향으로 리타겟팅시킨 영상의 배경과 GT 영상(930)의 배경 사이의 차이가 감소 또는 최소화되도록 훈련되고, 훈련용 리사이징 영상(920)을 제 1 시프트 맵(994) 및 제 2 시프트 맵(996)에 기초하여 세로 및 가로 방향으로 리타겟팅시킨 영상의 전경과 GT 영상(930)의 전경 사이의 차이가 감소 또는 최소화되도록 훈련될 수 있다. 특히, 훈련용 리타겟팅 영상과 GT 영상(930) 사이의 차이에 제 3 전경 마스크(36)가 적용되는데, 제 3 전경 마스크(36)는 훈련용 원본 영상(910)에 포함된 전경 오브젝트와 동일 크기의 오브젝트를 추출할 수 있으며, GT 영상(930)은 훈련용 원본 영상(910)에 포함된 전경 오브젝트의 크기와 동일한 크기의 오브젝트를 포함하고 있으므로, 결국, 신경망 모델(300)은 훈련용 리타겟팅 영상에 포함된 전경 오브젝트가 훈련용 원본 영상(910)에 포함된 전경 오브젝트와 동일 또는 유사해지도록 훈련될 수 있는 것이다.

[125] 도 13은 일 실시예에 따른 영상 처리 장치(100)의 구성을 도시하는 블록도이다.

[126] 도 13을 참조하면, 일 실시예에 따른 영상 처리 장치(100)는 메모리(110) 및 프로세서(130)를 포함할 수 있다.

[127] 메모리(110)는 적어도 하나의 프로그램 및 신경망 모델(300)을 저장할 수 있다. 프로세서(130)는 메모리(110)에 저장된 적어도 하나의 프로그램이 실행됨에 따라 영상의 리타겟팅을 포함한 영상 처리 동작을 수행할 수 있다.

[128] 구체적으로, 프로세서(130)는 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하고, 리사이징 영상을 신경망 모델(300)에 입력할 수 있다. 프로세서(130)는 신경망 모델(300)에서 출력되는 시프트 맵에 기초하여 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득할 수 있다. 리사이징 영상을 신경망 모델(300)에 입력하고, 신경망 모델(300)에서 출력되는 시프트 맵에 기초하여 리타겟팅 영상을 생성하는 방법에 대해서는 앞서 상세히 설명하였으므로, 여기에서는 그 설명을 생략한다.

[129] 도 13에 도시되지는 않았지만, 프로세서(130)는 리타겟팅 영상을 디스플레이로 전송하여, 리타겟팅 영상이 디스플레이에 의해 재생되도록 할 수 있다. 또한, 프로세서(130)는 원본 영상으로부터 리사이징 영상을 획득하는데 있어, 리사이징 영상의 종횡비를 디스플레이가 가지는 종횡비에 기초하여 결정할 수 있다. 예를 들어, 프로세서(130)는 리사이징 영상의 종횡비를 디스플레이가 가지는 종횡비와 동일하게 결정할 수 있다.

[130] 일 실시예에 따른 영상 처리 장치(100)는 스마트폰, 태블릿 PC, 데스크탑 PC, 텔레비전, 스마트워치, 스마트글래스, 서버 등에 포함될 수 있다.

[131] 영상 처리 장치(100)에 저장되는 신경망 모델(300)의 훈련은 다양한 장치에서 이루어질 수 있다.

[132] 일 실시예에서, 신경망 모델(300)의 훈련이 외부 장치(예를 들어, 서버 등)에서 수행되고, 훈련된 신경망 모델(300)이 영상 처리 장치(100)에 저장될 수 있다.

[133] 일 실시예에서, 신경망 모델(300)의 훈련이 외부 장치(예를 들어, 서버 등)에서

수행되고, 훈련된 신경망 모델(300)의 데이터가 네트워크를 통해 영상 처리 장치(100)의 통신 모듈(미도시)로 전송될 수도 있다.

[134] 또는, 일 예에서, 신경망 모델(300)이 영상 처리 장치(100)에 저장된 상태에서, 신경망 모델(300)의 훈련이 영상 처리 장치(100)에서 수행될 수 있다.

[135] 또는, 일 예에서, 신경망 모델(300)의 훈련이 외부 장치(예를 들어, 서버 등)에서 수행된 후, 신경망 모델(300)의 내부 가중치를 나타내는 정보가 외부 장치로부터 영상 처리 장치(100)로 네트워크를 통해 전송될 수도 있다. 영상 처리 장치(100)는 수신된 내부 가중치 정보에 따라 미리 저장된 신경망 모델(300)을 훈련시킬 수 있다.

[136] 한편, 상술한 본 개시의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 작성된 프로그램은 매체 또는 컴퓨터 프로그램 제품에 저장될 수 있다.

[137] 매체 또는 컴퓨터 프로그램 제품은 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체 또는 컴퓨터 프로그램 제품은 단일 또는 수개 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 것에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체 또는 컴퓨터 프로그램 제품의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체 또는 컴퓨터 프로그램 제품의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.

[138] 이상, 본 개시의 기술적 사상을 바람직한 실시예를 들어 상세하게 설명하였으나, 본 개시의 기술적 사상은 상기 실시예들에 한정되지 않고, 본 개시의 기술적 사상의 범위 내에서 당 분야에서 통상의 지식을 가진 자에 의하여 여러 가지 변형 및 변경이 가능하다.

청구범위

[청구항 1] 소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하는 단계;
상기 리사이징 영상을 미리 훈련된 신경망 모델에 입력하는 단계; 및
상기 신경망 모델에서 출력되는 시프트 맵에 기초하여 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함하는 것을 특징으로 하는 영상 처리 방법.

[청구항 2] 제1항에 있어서,
상기 신경망 모델은,
상기 리사이징 영상을 입력받는 제 1 서브 모델, 상기 제 1 서브 모델로부터 출력되는 특징 맵을 히든 스테이트에 따라 처리하는 컨볼루션 LSTM(long short-term memory) 레이어 및 상기 컨볼루션 LSTM 레이어에서 출력되는 데이터를 처리하여 상기 시프트 맵을 출력하는 제 2 서브 모델을 포함하는 것을 특징으로 하는 영상 처리 방법.

[청구항 3] 제2항에 있어서,
상기 컨볼루션 LSTM 레이어는,
이전 시간대의 원본 영상에 대응하는 리사이징 영상을 처리한 후 업데이트된 상기 히든 스테이트를 이용하여 상기 원본 영상에 대응하는 리사이징 영상을 처리하는 것을 특징으로 하는 영상 처리 방법.

[청구항 4] 제2항에 있어서,
상기 제 1 서브 모델 및 제 2 서브 모델 각각은,
소정의 필터 커널을 이용하여 입력 데이터를 컨볼루션 처리하는 적어도 하나의 컨볼루션 레이어를 포함하는 것을 특징으로 하는 영상 처리 방법.

[청구항 5] 제2항에 있어서,
상기 리사이징 영상의 픽셀들의 제 1 위치 맵 및 제 2 위치 맵이 상기 제 2 서브 모델로 입력되며,
상기 제 1 위치 맵의 첫 번째 행과 마지막 행에는 상기 원본 영상과 상기 리사이징 영상 사이의 세로 크기들 사이의 비율에 대응하는 값이 할당되고,
상기 제 2 위치 맵의 첫 번째 열과 마지막 열에는, 상기 원본 영상과 상기 리사이징 영상 사이의 가로 크기들 사이의 비율에 대응하는 값이 할당되는 것을 특징으로 하는 영상 처리 방법.

[청구항 6] 제5항에 있어서,
상기 제 1 위치 맵의 상기 첫 번째 행으로부터 상기 마지막 행까지 소정 차이를 갖는 값들이 순차적으로 각 행에 할당되고,
상기 제 2 위치 맵의 상기 첫 번째 열로부터 상기 마지막 열까지 소정 차이를 갖는 값들이 순차적으로 각 열에 할당되는 것을 특징으로 하는

영상 처리 방법.

[청구항 7] 제1항에 있어서,
상기 리타겟팅 영상을 획득하는 단계는,
상기 리사이징 영상으로부터 제 1 방향으로의 리타겟팅을 위한 시프트 맵에 기초하여 변환된 영상을, 제 2 방향으로의 리타겟팅을 위한 시프트 맵을 기초로 변환하여 상기 리타겟팅 영상을 획득하는 단계를 포함하는 것을 특징으로 하는 영상 처리 방법.

[청구항 8] 제1항에 있어서,
상기 리타겟팅 영상을 획득하는 단계는,
이전 시간대의 시프트 맵과 현재 시간대의 상기 시프트 맵을 조합하여, 상기 현재 시간대의 시프트 맵을 갱신하는 단계; 및
상기 갱신된 상기 현재 시간대의 시프트 맵에 기초하여, 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 단계를 포함하는 것을 특징으로 하는 영상 처리 방법.

[청구항 9] 제1항에 있어서,
GT(ground truth) 영상과 훈련용 리타겟팅 영상의 차이에 마스크를 적용한 결과에 대응하는 제 1 손실 정보가 감소되도록 상기 신경망 모델이 훈련되는 것을 특징으로 하는 영상 처리 방법.

[청구항 10] 제1항에 있어서,
훈련용 리사이징 영상을 제 1 방향으로의 리타겟팅을 위한 시프트 맵에 따라 변환시켜 획득된 영상과 GT 영상의 차이에 제 1 배경 마스크를 적용한 결과에 대응하는 제 2 손실 정보와,
상기 훈련용 리사이징 영상을 상기 제 1 방향으로의 리타겟팅을 위한 시프트 맵 및 제 2 방향으로의 리타겟팅을 위한 시프트 맵에 따라 변환시켜 획득된 훈련용 리타겟팅 영상과 상기 GT 영상의 차이에 제 1 전경 마스크를 적용한 결과에 대응하는 제 3 손실 정보가 감소되도록 상기 신경망 모델이 훈련되는 것을 특징으로 하는 영상 처리 방법.

[청구항 11] 제10항에 있어서,
상기 제 1 전경 마스크는, 훈련용 원본 영상에 대응하는 제 2 전경 마스크를 패딩하여 획득되며,
상기 제 1 배경 마스크는, 상기 제 2 전경 마스크에 대응하는 제 2 배경 마스크와, 상기 제 2 전경 마스크의 크기를 변경시켜 획득한 제 3 전경 마스크에 대응하는 제 3 배경 마스크를 서로 결합하여 획득되는 것을 특징으로 하는 영상 처리 방법.

[청구항 12] 제11항에 있어서,
상기 GT 영상은,
상기 훈련용 원본 영상에 상기 제 2 전경 마스크를 적용하여 획득한 전경 오브젝트와, 상기 훈련용 리사이징 영상에 상기 제 3 배경 마스크를

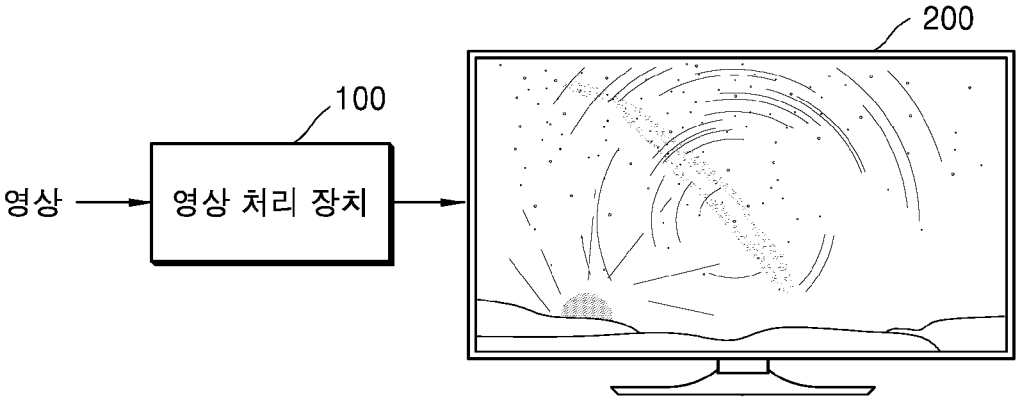
적용하여 획득한 배경 오브젝트를 결합하여 획득되는 것을 특징으로 하는 영상 처리 방법.

[청구항 13] 제1항에 있어서,
이전 시간대의 훈련용 원본 영상에 대응하는 시프트 맵과 현재 시간대의 훈련용 원본 영상에 대응하는 시프트 맵 사이의 관계에 기초하여 도출되는 제 4 손실 정보가 감소되도록 상기 신경망 모델이 훈련되는 것을 특징으로 하는 영상 처리 방법.

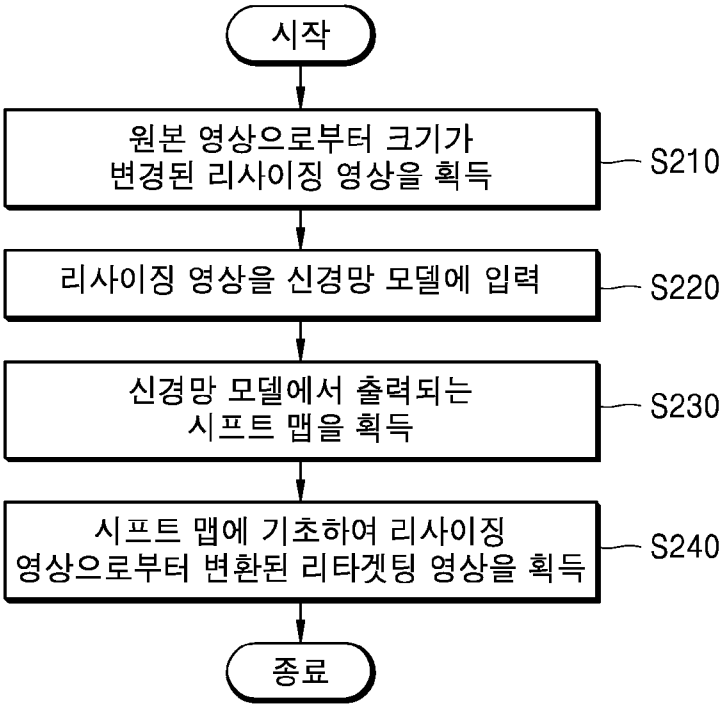
[청구항 14] 제1항 내지 제13항 중 어느 하나의 항의 영상 처리 방법을 실행하기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

[청구항 15] 프로세서; 및
미리 훈련된 신경망 모델 및 적어도 하나의 프로그램을 저장하는 메모리를 포함하되,
상기 프로세서는, 상기 메모리에 저장된 상기 적어도 하나의 프로그램이 실행됨에 따라,
소정 종횡비에 따라 원본 영상으로부터 크기가 변경된 리사이징 영상을 획득하고,
상기 리사이징 영상을 상기 신경망 모델에 입력하고,
상기 신경망 모델에서 출력되는 시프트 맵에 기초하여 상기 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득하는 것을 특징으로 하는 영상 처리 장치.

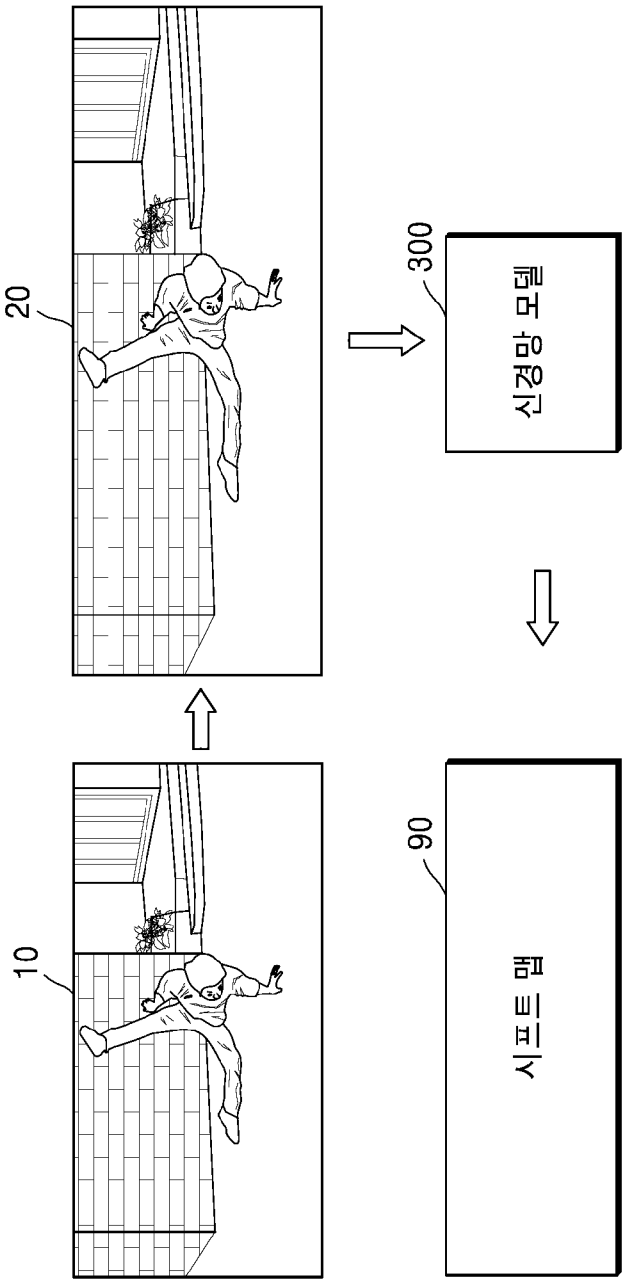
[도1]


200
100
영상
영상 처리 장치

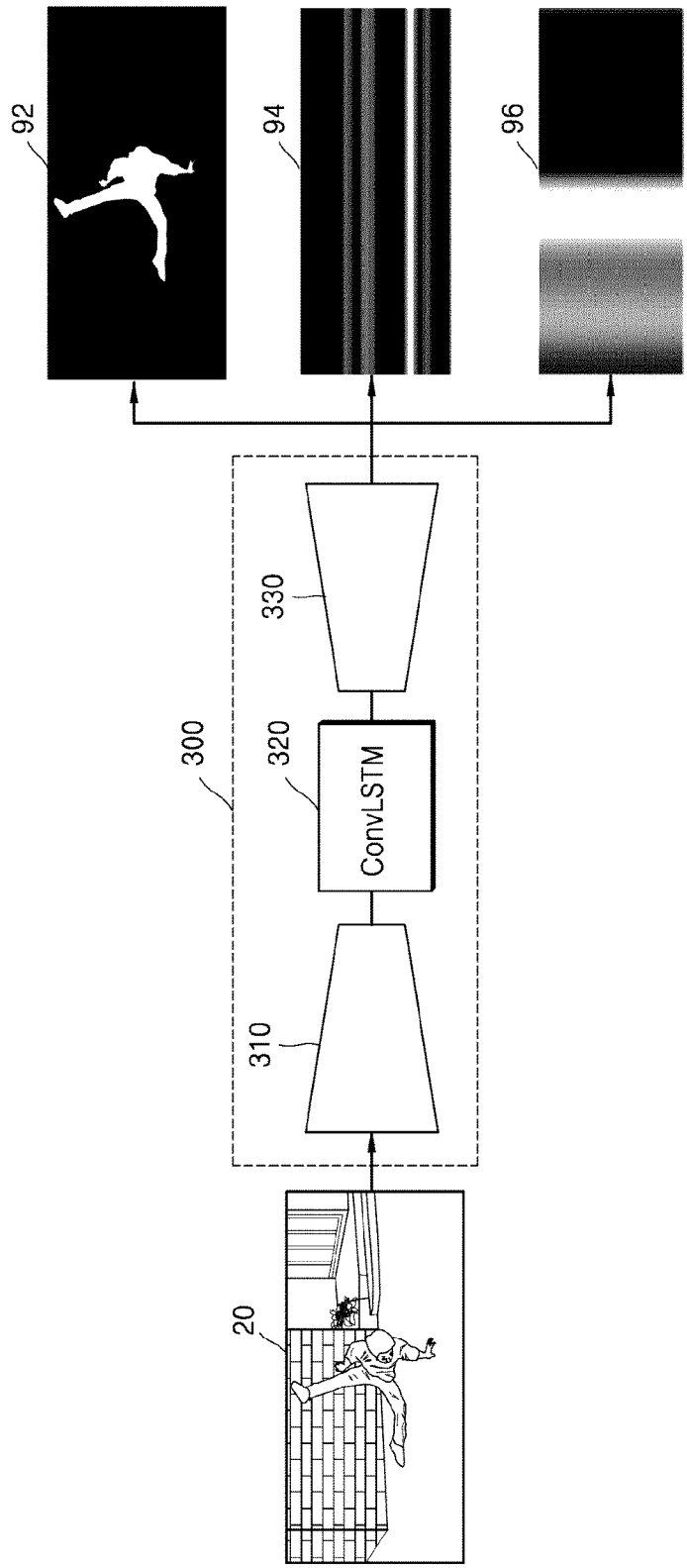
[도2]


시작
원본 영상으로부터 크기가 변경된 리사이징 영상을 획득
S210
리사이징 영상을 신경망 모델에 입력
S220
신경망 모델에서 출력되는 시프트 맵을 획득
S230
시프트 맵에 기초하여 리사이징 영상으로부터 변환된 리타겟팅 영상을 획득
S240
종료

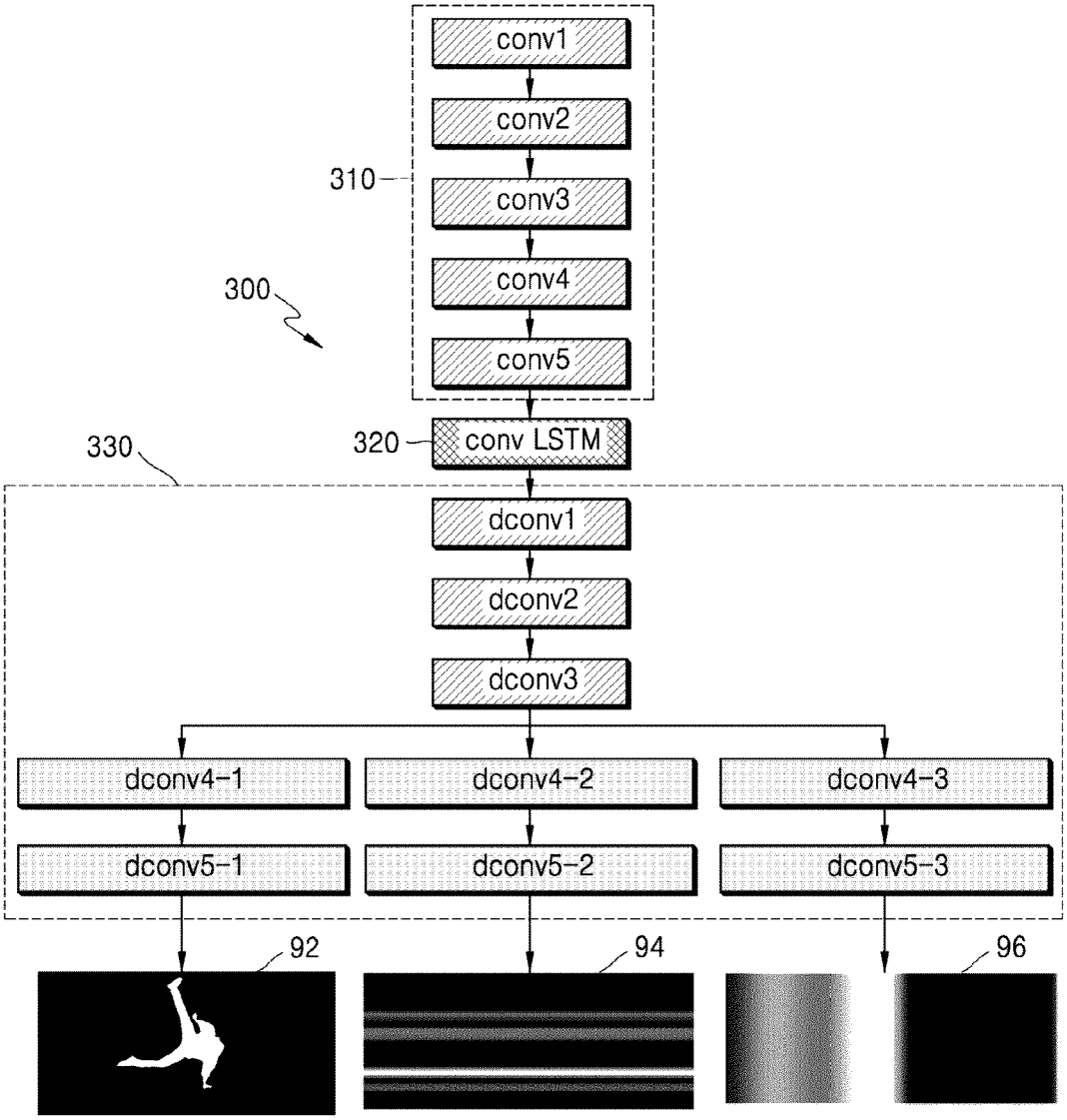
[도3]



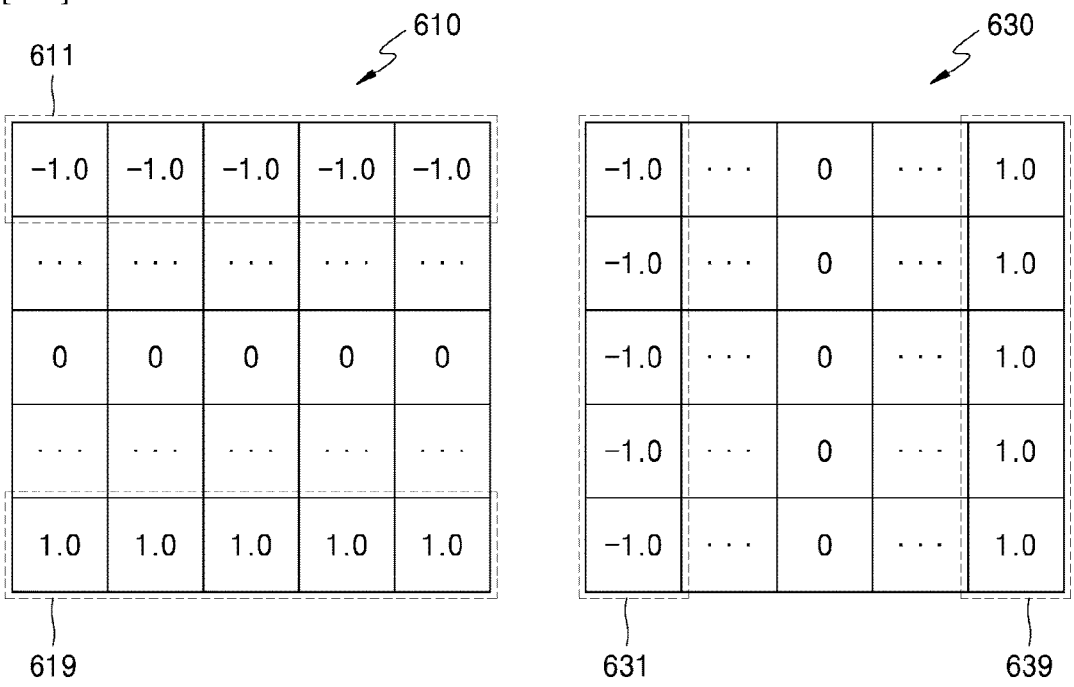
[도4]



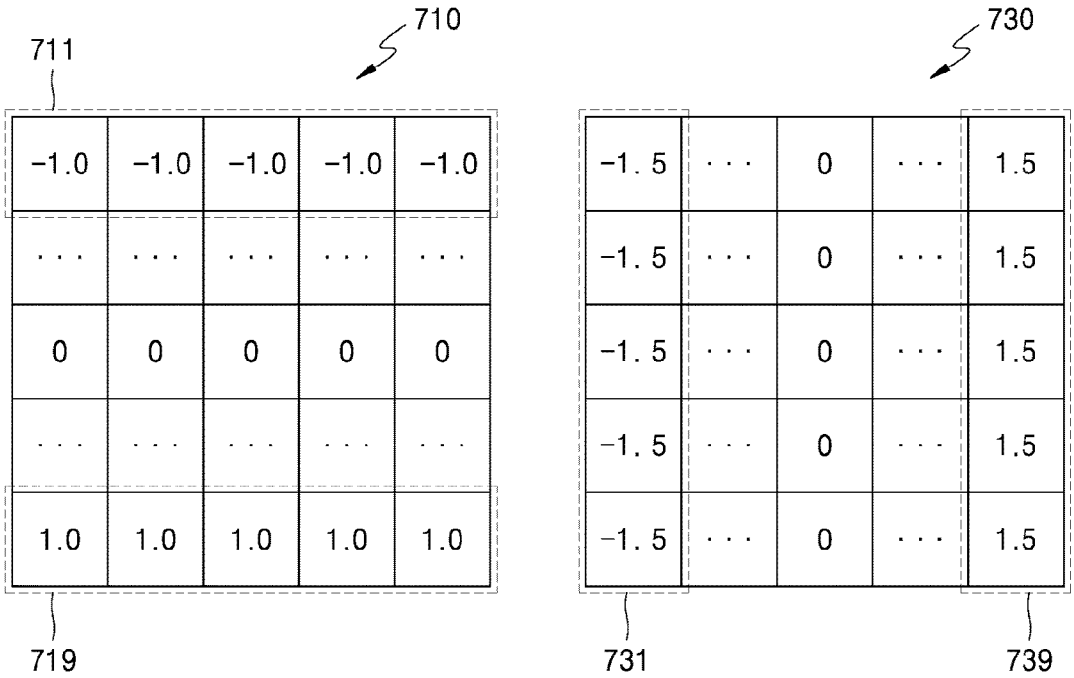
[도5]


conv1
conv2
conv3
conv4
conv5
310
300
320
conv LSTM
330
dconv1
dconv2
dconv3
dconv4-1
dconv4-2
dconv4-3
dconv5-1
dconv5-2
dconv5-3
92
94
96

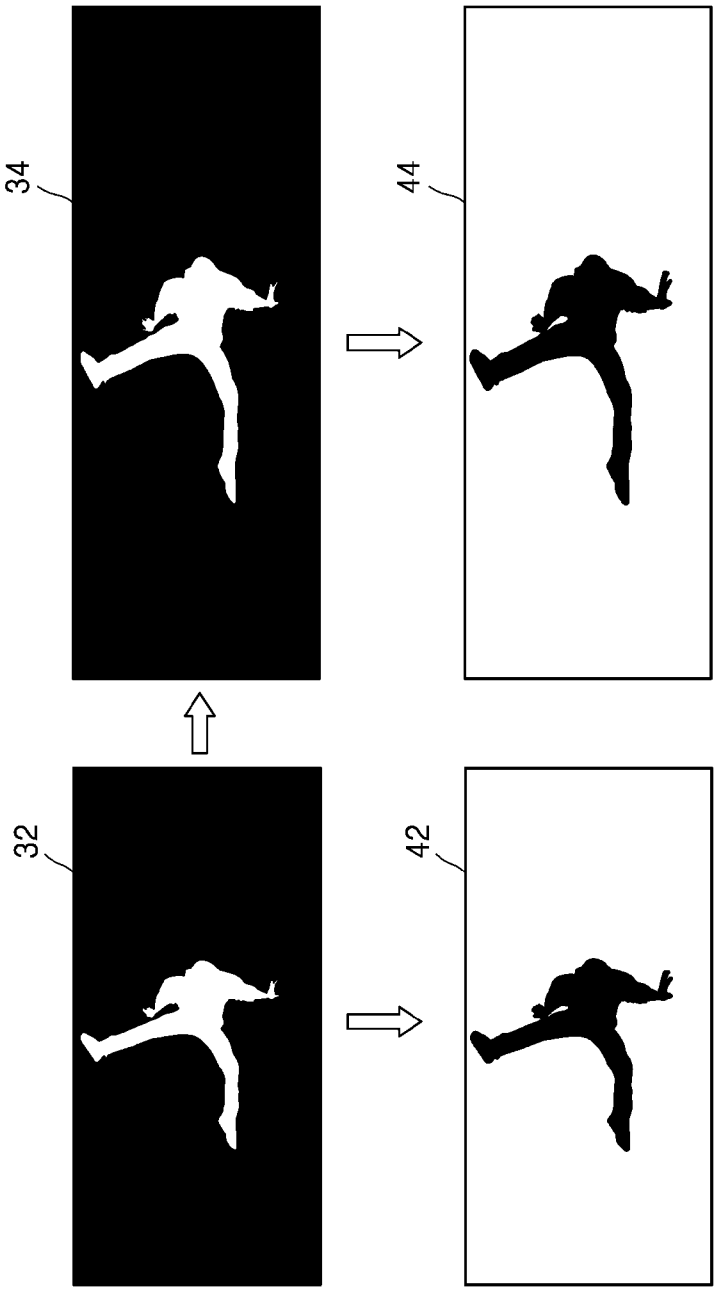
[도6]


610
611
630
-1.0 -1.0 -1.0 -1.0 -1.0
0 0 0 0 0
1.0 1.0 1.0 1.0 1.0
-1.0 0 1.0
619
631
639

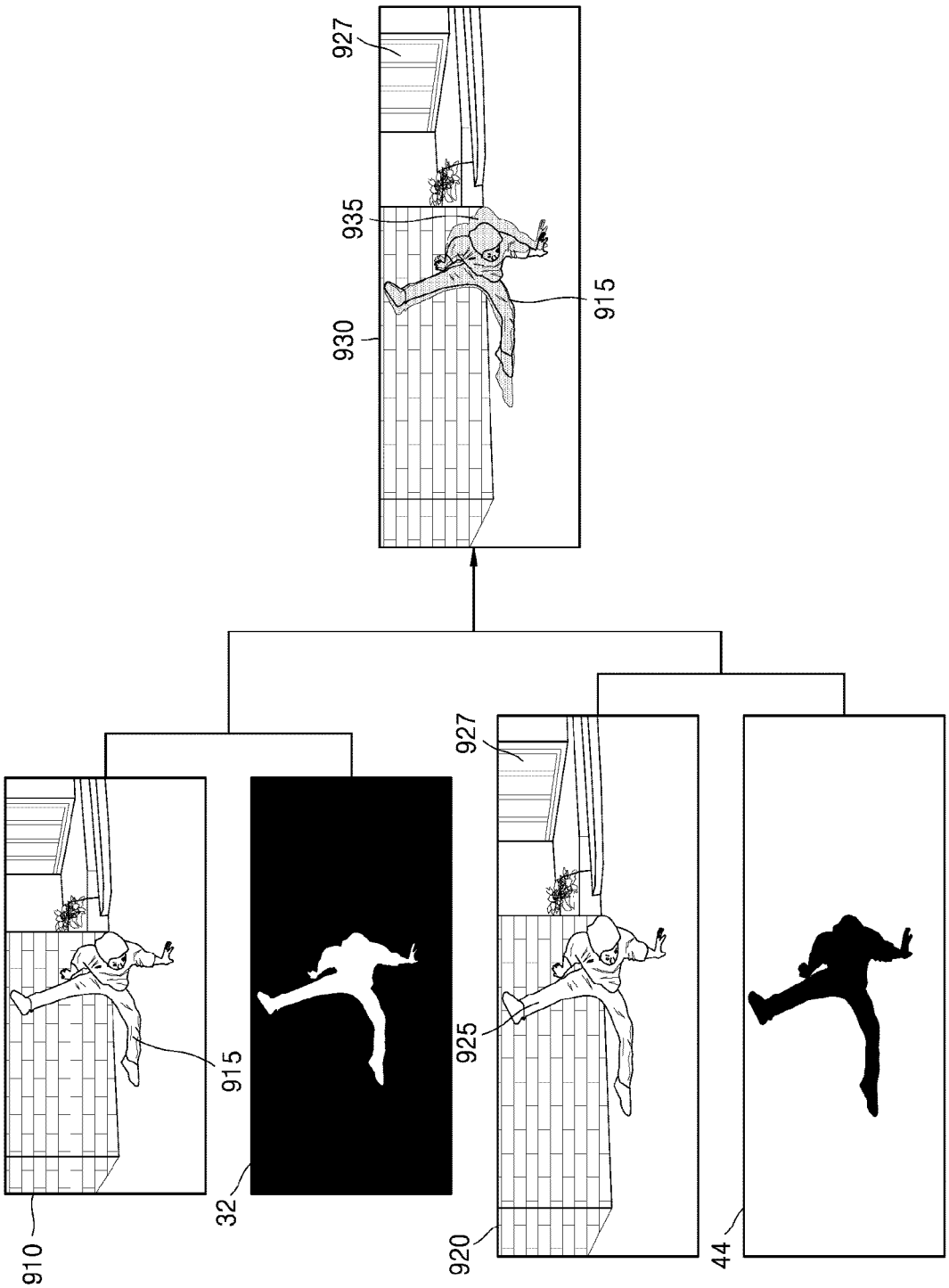
[도7]


710
711
730
-1.0 -1.0 -1.0 -1.0 -1.0
0 0 0 0 0
1.0 1.0 1.0 1.0 1.0
719
-1.5 0 1.5
-1.5 0 1.5
-1.5 0 1.5
-1.5 0 1.5
-1.5 0 1.5
731
739

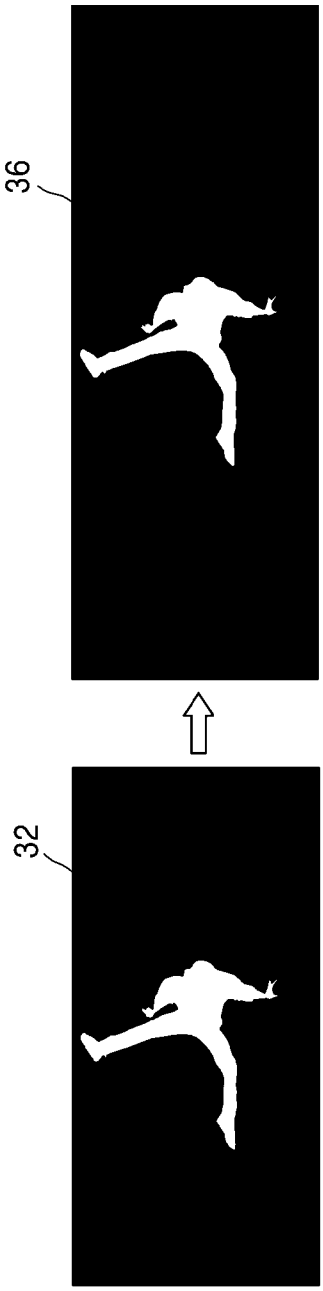
[도8]



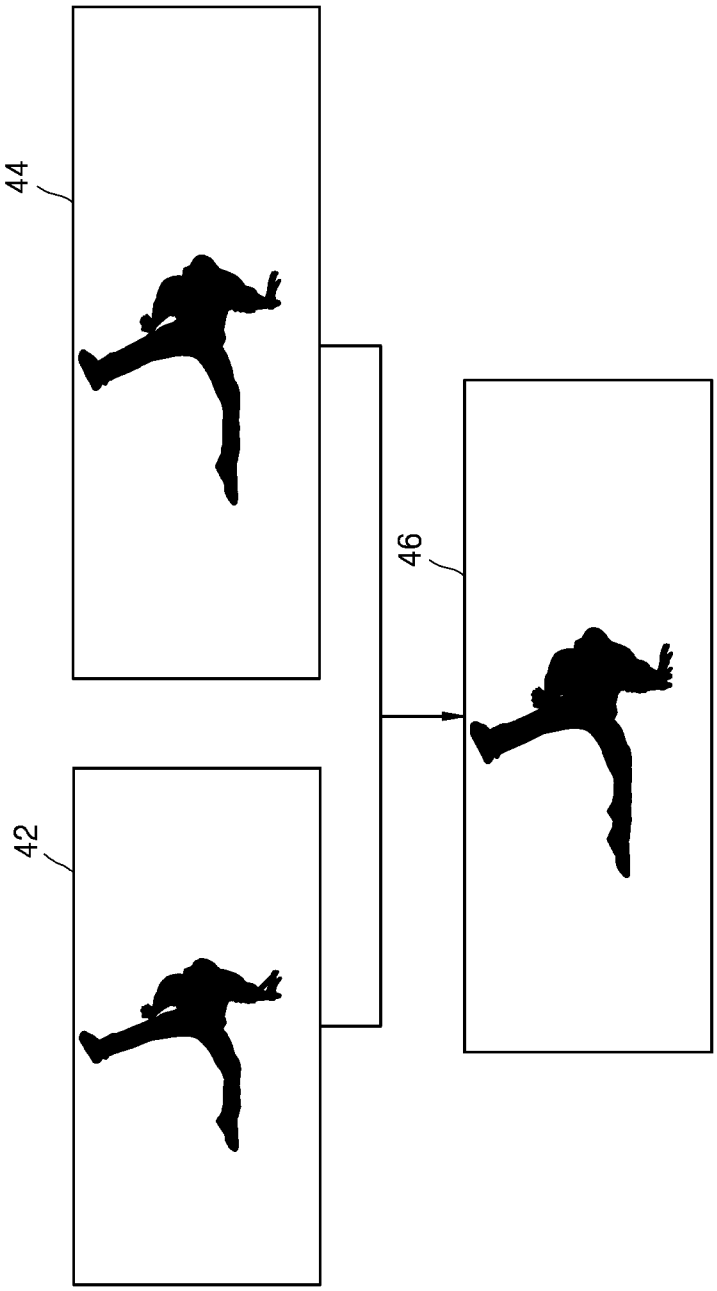
[도9]



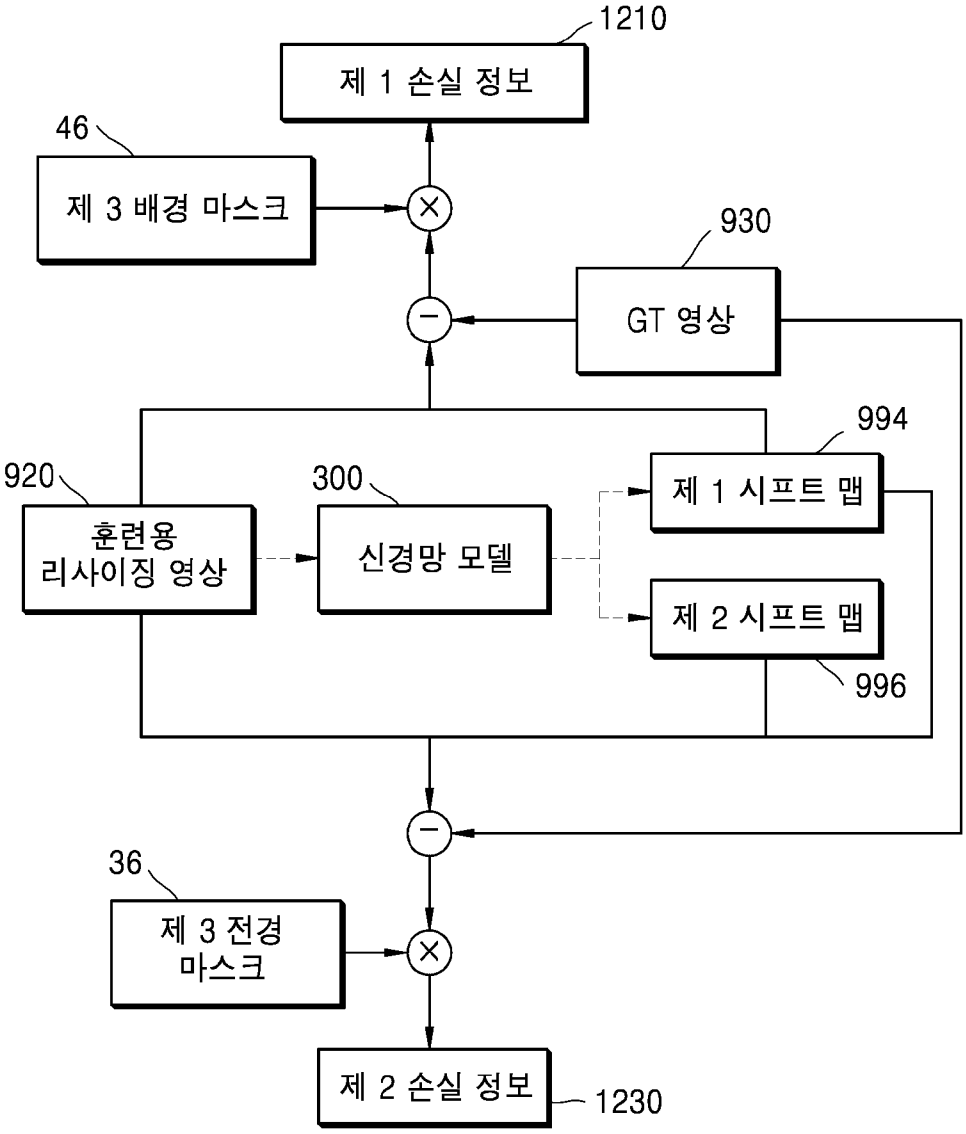
[도10]



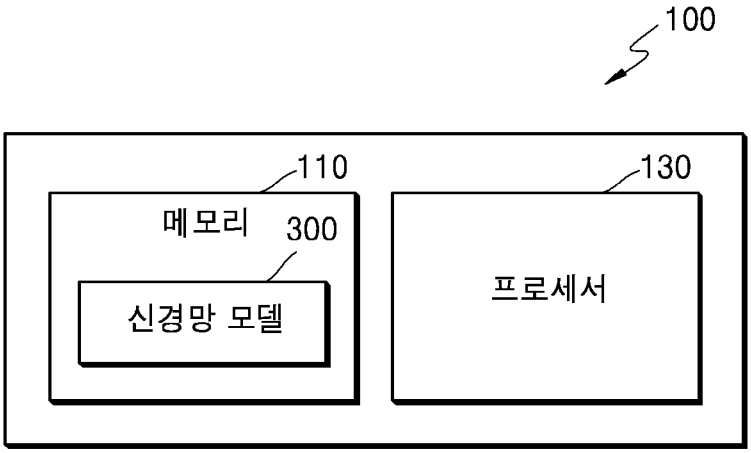
[도11]



[도12]


1210
제 1 손실 정보
46
제 3 배경 마스크
930
GT 영상
994
920
300
제 1 시프트 맵
훈련용
리사이징 영상
신경망 모델
제 2 시프트 맵
996
36
제 3 전경
마스크
제 2 손실 정보
1230

[도13]


100
110
130
메모리
300
신경망 모델
프로세서

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2019/015668

A. CLASSIFICATION OF SUBJECT MATTER

G06T 3/40(2006.01)i, G06T 5/20(2006.01)i, G06N 3/02(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T 3/40; G06K 9/00; G06K 9/62; H04N 19/117; H04N 19/177; G06T 5/20; G06N 3/02

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: shift-map, retargeting, resizing, neural model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PRITCH, Yael et al. Shift-Map Image Editing. 2009 IEEE 12th International Conference on Computer Vision. pages 151-158, 02 October 2009, [Retrieved on 10 February 2020], Retrieved from [https://www.cse.huji.ac.il/~peleg/papers/iccv09-shiftmap.pdf] See pages 151-154, 157.	1-10,13-15
A		11-12
Y	US 2018-0114071 A1 (NOKIA TECHNOLOGIES OY.) 26 April 2018 See paragraphs [0035], [0042]; and claims 1, 4.	1-10,13-15
Y	YAMASHITA, Rikiya et al. Convolutional neural networks: an overview and application in radiology. Insights into Imaging. pages 611-629, 22 June 2018, [Retrieved on 12 February 2020], Retrieved from [https://link.springer.com/content/pdf/10.1007/s13244-018-0639-9.pdf] See page 617.	9-10,13
A	LI, Yue et al. Learning a Convolutional Neural Network for Image Compact-Resolution. IEEE Transactions on Image Processing. pages 1-16, 28 September 2018, [Retrieved on 12 February 2020], Retrieved from [https://sci-hub.se/https://ieeexplore.ieee.org/abstract/document/8476610] See pages 1-14.	1-15

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 FEBRUARY 2020 (27.02.2020)	**27 FEBRUARY 2020 (27.02.2020)**
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea Facsimile No. +82-42-481-8578	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2019/015668

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2016-132148 A1 (MAGIC PONY TECHNOLOGY LIMITED) 25 August 2016 See pages 11-12; claims 1-16; and figures 1-7.	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2019/015668

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2018-0114071 A1	26/04/2018	GB 2555136 A	25/04/2018
WO 2016-132148 A1	25/08/2016	EP 3298776 A1	28/03/2018
		EP 3493149 A1	05/06/2019
		GB 2543958 A	03/05/2017
		GB 2548749 A	27/09/2017
		US 2018-0131953 A1	10/05/2018
		US 2018-0139458 A1	17/05/2018
		WO 2017-158363 A1	21/09/2017

국 제 조 사 보 고 서

국제출원번호
PCT/KR2019/015668

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06T 3/40(2006.01)i, G06T 5/20(2006.01)i, G06N 3/02(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06T 3/40; G06K 9/00; G06K 9/62; H04N 19/117; H04N 19/177; G06T 5/20; G06N 3/02

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 시프트맵(shift-map), 리타겟팅(retargeting), 리사이징(resizing), 신경망모델(neural model),

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	YAEL PRITCH 등, Shift-Map Image Editing, 2009 IEEE 12th International Conference on Computer Vision, 페이지 151-158, 2009.10.02 [검색일: 2020-02-10], 출처 [https://www.cse.huji.ac.il/~peleg/papers/iccv09-shiftmap.pdf] 페이지 151-154, 157	1-10,13-15
A		11-12
Y	US 2018-0114071 A1 (NOKIA TECHNOLOGIES OY) 2018.04.26 단락 [0035], [0042]; 및 청구항 1, 4	1-10,13-15
Y	RIKIYA YAMASHITA 등, Convolutional neural networks: an overview and application in radiology, Insights into Imaging, 페이지 611-629, 2018.06.22 [검색일: 2020-02-12], 출처 [https://link.springer.com/content/pdf/10.1007/s13244-018-0639-9.pdf] 페이지 617	9-10,13
A	YUE LI 등, Learning a Convolutional Neural Network for Image Compact-Resolution, IEEE Transactions on Image Processing, 페이지 1-16, 2018.09.28 [검색일: 2020-02-12], 출처 [https://sci-hub.se/https://ieeexplore.ieee.org/abstract/document/8476610] 페이지 1-14	1-15

☒ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

"D" 본 국제출원에서 출원인이 인용한 문헌

"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

"&" 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일	국제조사보고서 발송일
2020년 02월 27일 (27.02.2020)	2020년 02월 27일 (27.02.2020)

ISA/KR의 명칭 및 우편주소
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)
팩스 번호 +82-42-481-8578

심사관
김성희
전화번호 +82-42-481-3516

국 제 조 사 보 고 서

국제출원번호
PCT/KR2019/015668

C (계속). 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	WO 2016-132148 A1 (MAGIC PONY TECHNOLOGY LIMITED) 2016.08.25 페이지 11-12; 청구항 1-16; 및 도면 1-7	1-15

국제조사보고서
대응특허에 관한 정보

국제출원번호
PCT/KR2019/015668

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2018-0114071 A1	2018/04/26	GB 2555136 A	2018/04/25
WO 2016-132148 A1	2016/08/25	EP 3298776 A1	2018/03/28
		EP 3493149 A1	2019/06/05
		GB 2543958 A	2017/05/03
		GB 2548749 A	2017/09/27
		US 2018-0131953 A1	2018/05/10
		US 2018-0139458 A1	2018/05/17
		WO 2017-158363 A1	2017/09/21