(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0057899
(43) 공개일자 2025년04월29일

(51) 국제특허분류(Int. Cl.)
G06T 11/60 (2006.01) *G06F 3/0484* (2022.01)
G06T 15/20 (2011.01) *G06T 5/77* (2024.01)
G06T 7/30 (2017.01) *G06V 40/16* (2022.01)
(52) CPC특허분류
G06T 11/60 (2013.01)
G06F 3/0484 (2022.01)
(21) 출원번호 10-2025-7010962
(22) 출원일자(국제) 2024년10월03일
심사청구일자 2025년04월03일
(85) 번역문제출일자 2025년04월03일
(86) 국제출원번호 PCT/US2024/049845
(87) 국제공개번호 WO 2025/076259
국제공개일자 2025년04월10일
(30) 우선권주장
63/542,283 2023년10월03일 미국(US)

(71) 출원인
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
테넨바움 제이
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
조메트 아사프
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(뒷면에 계속)
(74) 대리인
박장원

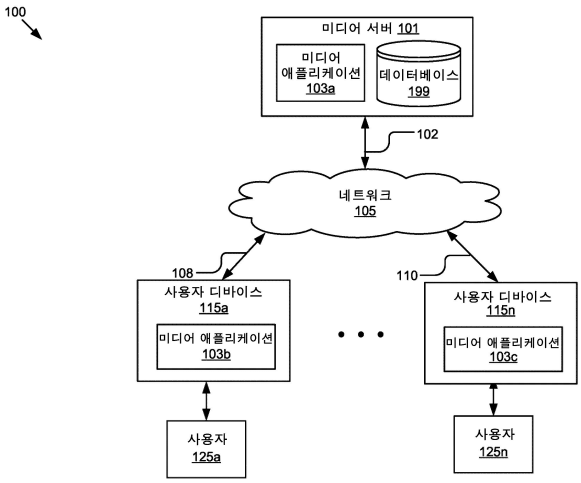
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **머리 포즈 또는 얼굴 영역 개선을 이용한 이미지 생성**

(57) 요 약

미디어 애플리케이션은 소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하며, 소스 이미지 및 타겟 이미지는 적어도 대상체를 포함한다. 미디어 애플리케이션은 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정한다. 머리 편집기를 사용하기로 결정하는 것에 응답하여, 미디어 애플리케이션은, 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하고, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지를 생성한다.

대 표 도



(52) CPC특허분류
G06T 15/20 (2013.01)
G06T 5/77 (2024.01)
G06T 7/30 (2017.01)
G06V 40/16 (2022.01)
G06T 2207/30201 (2013.01)
G06T 2210/12 (2013.01)
(72) 발명자
레비 바렐
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
브로드스키 야론
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
자다 시란
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
크난 야엘 프리치
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
에프라트 아리엘
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
모세리 인바
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
골버트 아브람
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)

명 세 서

청구범위

청구항 1

컴퓨터로 구현되는 방법으로서,

소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하는 단계 – 상기 소스 이미지 및 상기 타겟 이미지는 적어도 대상체를 포함함 –;

상기 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 단계; 및

상기 머리 편집기를 사용하기로 결정하는 것에 응답하여, 합성 이미지를 생성하는 단계로서,

상기 타겟 이미지 내의 상기 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 상기 소스 이미지 내의 상기 대상체의 소스 머리로부터의 머리 픽셀들로 대체하는 단계; 및

상기 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 상기 소스 이미지와 상기 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체하는 단계에 의해, 상기 합성 이미지를 생성하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 상기 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 단계를 더 포함하는, 방법.

청구항 3

제2항에 있어서, 상기 소스 이미지 내의 상기 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 단계는,

초기 포즈에서의 상기 타겟 머리 및 상기 소스 머리를 추출하는 단계;

상기 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 단계;

잠재 공간(latent space)에서, 상기 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 상기 소스 머리를 소스 벡터로서 인코딩하는 단계;

상기 소스 벡터로부터의 하나 이상의 컴포넌트들을 상기 타겟 벡터로 복사하는 단계;

상기 인코딩된 소스 머리로부터 상기 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 단계;

렌더링된 타겟 머리를 상기 초기 포즈에 재정렬하는 단계; 및

상기 재정렬된 타겟 머리를 상기 소스 이미지와 블렌딩(blending)하는 단계를 포함하는, 방법.

청구항 4

제2항에 있어서, 상기 얼굴 편집기를 사용하기로 결정하는 단계는, 상기 타겟 머리의 제1 각도와 상기 소스 머리의 제2 각도 사이의 각도 차이에 기초하는, 방법.

청구항 5

제1항에 있어서, 상기 머리 편집기를 사용하기로 결정하는 단계는, 상기 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 상기 경계 박스와 상기 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초하는, 방법.

청구항 6

제1항에 있어서, 상기 합성 이미지를 생성하는 단계는, 상기 타겟 머리와 연관되고 상기 소스 머리가 아닌 상기 타겟 이미지 내의 나머지 타겟 픽셀들을 식별하는 것에 응답하여, 상기 나머지 타겟 픽셀들을 인페인팅(inpainting)하는 단계를 더 포함하는, 방법.

청구항 7

제1항에 있어서,

상기 타겟 이미지와 상기 소스 이미지의 컬러 히스토그램들(color histograms)에 있어서의 차이를 결정하는 것에 기초하여 상기 타겟 머리의 가림(occlusion) 또는 상기 소스 머리의 가림을 결정하는 단계를 더 포함하며;

상기 머리 편집기를 사용하기로 결정하는 단계는 상기 타겟 머리의 가림 또는 상기 소스 머리의 가림에 기초하는, 방법.

청구항 8

제1항에 있어서, 상기 이미지들의 세트에 기초하여, 상기 하나 이상의 편집기들을 사용할지 여부를 결정하는 단계 전에, 상기 방법은,

카메라를 이용해, 상기 이미지들의 세트를 촬영하는 단계; 및

상기 타겟 이미지, 및 소스 이미지들의 세트 - 상기 소스 이미지들의 세트는 상기 소스 이미지를 포함함 - 로부터 상기 소스 머리를 선택하기 위한 옵션을 포함하는, 사용자 인터페이스를 사용자에게 제공하는 단계; 및

상기 사용자로부터, 상기 소스 이미지의 선택을 수신하는 단계를 더 포함하는, 방법.

청구항 9

제1항에 있어서, 상기 소스 이미지 내의 상기 적어도 하나의 대상체는 사람 또는 동물인, 방법.

청구항 10

시스템으로서,

하나 이상의 프로세서들; 및

상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금 동작들을 수행하게 하는 명령어들이 저장된 하나 이상의 컴퓨터 판독가능 매체들을 포함하며, 상기 동작들은,

소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하는 동작 - 상기 소스 이미지 및 상기 타겟 이미지는 적어도 대상체를 포함함 -;

상기 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 동작; 및

상기 머리 편집기를 사용하기로 결정하는 것에 응답하여, 합성 이미지를 생성하는 동작으로서,

상기 타겟 이미지 내의 상기 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 상기 소스 이미지 내의 상기 대상체의 소스 머리로부터의 머리 픽셀들로 대체하는 동작; 및

상기 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 상기 소스 이미지와 상기 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체하는 동작에 의해, 상기 합성 이미지를 생성하는 동작을 포함하는, 시스템.

청구항 11

제10항에 있어서, 상기 동작들은,

상기 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 상기 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작을 더 포함하는, 시스템.

청구항 12

제11항에 있어서, 상기 소스 이미지 내의 상기 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작은,

초기 포즈에서의 상기 타겟 머리 및 상기 소스 머리를 추출하는 동작;

상기 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 동작;

잠재 공간(latent space)에서, 상기 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 상기 소스 머리를 소스 벡터로서 인코딩하는 동작;

상기 소스 벡터로부터의 하나 이상의 컴포넌트들을 상기 타겟 벡터로 복사하는 동작;

상기 인코딩된 소스 머리로부터 상기 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 동작;

렌더링된 타겟 머리를 상기 초기 포즈에 재정렬하는 동작; 및

상기 재정렬된 타겟 머리를 상기 소스 이미지와 블렌딩(blending)하는 동작을 포함하는, 시스템.

청구항 13

제11항에 있어서, 상기 얼굴 편집기를 사용하기로 결정하는 동작은, 상기 타겟 머리의 제1 각도와 상기 소스 머리의 제2 각도 사이의 각도 차이에 기초하는, 시스템.

청구항 14

제10항에 있어서, 상기 머리 편집기를 사용하기로 결정하는 동작은, 상기 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 상기 경계 박스와 상기 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초하는, 시스템.

청구항 15

제10항에 있어서, 상기 합성 이미지를 생성하는 동작은, 상기 타겟 머리와 연관되고 상기 소스 머리가 아닌 상기 타겟 이미지 내의 나머지 타겟 픽셀들을 식별하는 것에 응답하여, 상기 나머지 타겟 픽셀들을 인페인팅(inpainting)하는 동작을 더 포함하는, 시스템.

청구항 16

하나 이상의 프로세싱 디바이스들에 의한 실행에 응답하여, 상기 하나 이상의 프로세싱 디바이스들로 하여금 동작들을 수행하게 하는, 명령어들이 저장된 비일시적 컴퓨터 판독가능 매체로서, 상기 동작들은,

소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하는 동작 – 상기 소스 이미지 및 상기 타겟 이미지는 적어도 대상체를 포함함 –;

상기 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 동작; 및

상기 머리 편집기를 사용하기로 결정하는 것에 응답하여, 합성 이미지를 생성하는 동작으로서,

상기 타겟 이미지 내의 상기 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 상기 소스 이미지 내의 상기 대상체의 소스 머리로부터의 머리 픽셀들로 대체하는 동작; 및

상기 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 상기 소스 이미지와 상기 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체하는 동작에 의해, 상기 합성 이미지를 생성하는 동작을 포함하는, 비일시적 컴퓨터 판독가능 매체.

청구항 17

제16항에 있어서, 상기 동작들은,

상기 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 상기 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작을 더 포함하

는, 비일시적 컴퓨터 판독가능 매체.

청구항 18

제17항에 있어서, 상기 소스 이미지 내의 상기 소스 얼굴 피처들로부터의 얼굴 픽셀들과 상기 타겟 이미지 내의 상기 타겟 얼굴 피처들에 기초하여 상기 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작은,

초기 포즈에서의 상기 타겟 머리 및 상기 소스 머리를 추출하는 동작;

상기 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 동작;

잠재 공간(latent space)에서, 상기 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 상기 소스 머리를 소스 벡터로서 인코딩하는 동작;

상기 소스 벡터로부터의 하나 이상의 컴포넌트들을 상기 타겟 벡터로 복사하는 동작;

상기 인코딩된 소스 머리로부터 상기 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 동작;

렌더링된 타겟 머리를 상기 초기 포즈에 재정렬하는 동작; 및

상기 재정렬된 타겟 머리를 상기 소스 이미지와 블렌딩(blending)하는 동작을 포함하는, 비일시적 컴퓨터 판독가능 매체.

청구항 19

제17항에 있어서, 상기 얼굴 편집기를 사용하기로 결정하는 동작은, 상기 타겟 머리의 제1 각도와 상기 소스 머리의 제2 각도 사이의 각도 차이에 기초하는, 비일시적 컴퓨터 판독가능 매체.

청구항 20

제16항에 있어서, 상기 머리 편집기를 사용하기로 결정하는 동작은, 상기 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 상기 경계 박스와 상기 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초하는, 비일시적 컴퓨터 판독가능 매체.

발명의 설명

기 술 분 야

[0001] 관련 출원 상호 참조

[0002] 본 출원은 2023년 10월 3일에 출원된 "Generating a Group Photo with Head Pose and Facial Recognition Improvements"라는 명칭의 미국 가 특허 출원 제63/542,283호에 대해 35 U.S. § 119(e)에 의거하여 우선권을 주장하는 정규 특허 출원이며, 가 특허 출원의 내용은 그 전체가 참조로서 본 명세서에 포함된다.

배 경 기 술

[0003] 단체 사진은 이벤트를 기념하는 인기 있는 방법이다. 사람들이 모두 미소를 짓고 있고 카메라를 보고 있는 한 무리의 사람들의 이미지를 획득하는 것은 어려운데, 이는 이미지 내의 사람들이 많을수록, 사람들의 얼굴들 중 적어도 하나가 그들의 최상의 표정이 아닐 가능성이 커지기 때문이다. 예를 들어, 한 사람은 그 입을 벌리고 있을 수 있으며, 다른 사람은 그 눈을 감고 있을 수 있으며, 다른 사람은 카메라를 보고 있지 않을 수 있는 등이다. 또한, 사람은 사진 속 다른 사람들과 다르게, 카메라에 대해 비스듬하게 머리를 기울일 수 있거나, 또는 그 외 질이 좋은 사진을 찍을 수 없는 포즈로 있다.

[0004] 본 명세서에서 제공되는 배경 설명은 본 개시의 컨텍스트를 일반적으로 제시하기 위한 목적이다. 본 배경기술 섹션에서 설명되는 정도로, 현재 명명된 발명자들의 작업뿐만 아니라, 그렇지 않으면 출원 시 종래 기술로서 자격을 갖지 않을 수 있는 설명의 양태들이 본 개시에 대한 종래 기술로서 명시적으로나 암시적으로 자인되는 것이 아니다.

발명의 내용

[0005] 컴퓨터 구현 방법은 소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하는 단계를 포함하며, 소스 이미지 및 타겟 이미지는 적어도 대상체(subject)를 포함한다. 본 방법은, 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 단계를 더 포함한다. 본 방법은, 머리 편집기를 사용하기로 결정하는 것에 응답하여, 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하며, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지를 생성하는 단계를 더 포함한다.

[0006] 일부 실시예들에서, 본 방법은, 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 단계를 더 포함한다. 일부 실시예들에서, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 단계는, 초기 포즈에서의 타겟 머리, 및 소스 머리를 추출하는 단계, 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 단계, 잠재 공간(latent space)에서, 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 소스 머리를 소스 벡터로서 인코딩하는 단계, 소스 벡터로부터의 하나 이상의 컴포넌트들을 타겟 벡터로 복사하는 단계, 인코딩된 소스 머리로부터 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 단계, 렌더링된 타겟 머리를 초기 포즈에 재정렬하는 단계, 및 재정렬된 타겟 머리를 소스 이미지와 블렌딩(blending)하는 단계를 포함한다.

[0007] 일부 실시예들에서, 얼굴 편집기를 사용하기로 결정하는 단계는, 타겟 머리의 제1 각도와 소스 머리의 제2 각도 사이의 각도 차이에 기초한다. 일부 실시예들에서, 머리 편집기를 사용하기로 결정하는 단계는, 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 경계 박스와 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초한다. 일부 실시예들에서, 합성 이미지를 생성하는 단계는, 타겟 머리와 연관되고 소스 머리가 아닌 타겟 이미지 내의 나머지 타겟 픽셀들을 식별하는 것에 응답하여, 나머지 타겟 픽셀들을 인페인팅(inpainting)하는 단계를 더 포함한다. 일부 실시예들에서, 본 방법은, 타겟 이미지와 소스 이미지의 컬러 히스토그램들(color histograms)에 있어서의 차이를 결정하는 것에 기초하여 타겟 머리의 가림(occlusion) 또는 소스 머리의 가림을 결정하는 단계를 더 포함하며, 머리 편집기를 사용하기로 결정하는 단계는 타겟 머리의 가림 또는 소스 머리의 가림에 기초한다.

[0008] 일부 실시예들에서, 본 방법은, 이미지들의 세트에 기초하여, 하나 이상의 편집기들을 사용할지 여부를 결정하는 단계 전에, 본 방법이, 카메라를 이용해, 이미지들의 세트를 촬영하는 단계; 타겟 이미지, 및 소스 이미지들의 세트 - 소스 이미지들의 세트는 소스 이미지를 포함함 - 로부터 소스 머리를 선택하기 위한 옵션을 포함하는, 사용자 인터페이스를 사용자에게 제공하는 단계; 및 사용자로부터, 소스 이미지의 선택을 수신하는 단계를 더 포함하는 것을 포함한다. 일부 실시예들에서, 소스 이미지 내의 적어도 하나의 대상체는 사람 또는 동물이다.

[0009] 시스템은 하나 이상의 프로세서들 및 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금 동작들을 수행하게 하는 명령어들이 저장된 하나 이상의 컴퓨터 판독가능 매체들을 포함한다. 동작들은, 소스 이미지 및 타겟 이미지 - 소스 이미지 및 타겟 이미지는 적어도 대상체를 포함함 - 를 포함하는 이미지들의 세트를 수신하는 동작; 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 동작; 및 머리 편집기를 사용하기로 결정하는 것에 응답하여, 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하며, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지를 생성하는 동작을 포함한다.

[0010] 일부 실시예들에서, 동작들은, 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작을 더 포함한다. 일부 실시예들에서, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작은, 초기 포즈에서의 타겟 머리, 및 소스 머리를 추출하는 동작; 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 동작; 잠재 공간(latent space)에서, 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 소스 머리를 소스 벡터로서 인코딩하는 동작; 소스 벡터로부터의 하나 이상의 컴포넌트들을 타겟 벡터로 복사하는 동작; 인코딩된 소스 머리로부터 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 동작; 렌더링된 타겟 머리를 초기 포즈에 재정렬하는 동작; 및 재정렬된

타겟 머리를 소스 이미지와 블렌딩(blending)하는 동작을 포함한다. 일부 실시예들에서, 얼굴 편집기를 사용하기로 결정하는 동작은, 타겟 머리의 제1 각도와 소스 머리의 제2 각도 사이의 각도 차이에 기초한다. 일부 실시예들에서, 머리 편집기를 사용하기로 결정하는 동작은, 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 경계 박스와 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초한다. 일부 실시예들에서, 합성 이미지를 생성하는 동작은, 타겟 머리와 연관되고 소스 머리가 아닌 타겟 이미지 내의 나머지 타겟 픽셀들을 식별하는 것에 응답하여, 나머지 타겟 픽셀들을 인페인팅(inpainting)하는 동작을 더 포함한다.

[0011] 명령어들이 저장된 비일시적 컴퓨터 판독가능 매체로서, 하나 이상의 프로세싱 디바이스들에 의한 실행에 응답하여, 하나 이상의 프로세싱 디바이스들로 하여금 동작들을 수행하게 한다. 동작들은, 소스 이미지 및 타겟 이미지 – 소스 이미지 및 타겟 이미지는 적어도 대상체를 포함함 – 를 포함하는 이미지들의 세트를 수신하는 동작; 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정하는 동작; 및 머리 편집기를 사용하기로 결정하는 것에 응답하여, 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하며, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지를 생성하는 동작을 포함한다.

[0012] 일부 실시예들에서, 동작들은, 얼굴 편집기를 사용하기로 결정하는 것에 응답하여, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작을 더 포함한다. 일부 실시예들에서, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 동작은, 초기 포즈에서의 타겟 머리, 및 소스 머리를 추출하는 동작; 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 동작; 잠재 공간(latent space)에서, 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 소스 머리를 소스 벡터로서 인코딩하는 동작; 소스 벡터로부터의 하나 이상의 컴포넌트들을 타겟 벡터로 복사하는 동작; 인코딩된 소스 머리로부터 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 동작; 렌더링된 타겟 머리를 초기 포즈에 재정렬하는 동작; 및 재정렬된 타겟 머리를 소스 이미지와 블렌딩(blending)하는 동작을 포함한다. 일부 실시예들에서, 얼굴 편집기를 사용하기로 결정하는 동작은, 타겟 머리의 제1 각도와 소스 머리의 제2 각도 사이의 각도 차이에 기초한다. 일부 실시예들에서, 머리 편집기를 사용하기로 결정하는 동작은, 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 경계 박스와 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리에 기초한다.

도면의 간단한 설명

[0013] 도 1은 본 명세서에서 설명되는 일부 실시예들에 따른, 예시적인 네트워크 환경의 블록도이다.

도 2는 본 명세서에서 설명되는 일부 실시예들에 따른, 예시적인 컴퓨팅 디바이스의 블록도이다.

도 3a는 본 명세서에서 설명되는 일부 실시예들에 따른, 사용자가 이미지들의 세트로부터 타겟 이미지를 지정하기 위한 옵션들을 포함하는 예시적인 사용자 인터페이스이다.

도 3b는 본 명세서에서 설명되는 일부 실시예들에 따른, 사용자가 이미지들의 세트로부터 하나 이상의 소스 이미지들을 지정하기 위한 옵션들을 포함하는 예시적인 사용자 인터페이스이다.

도 4a는 본 명세서에서 설명되는 일부 실시예들에 따른, 머리 편집기에 대한 예시적인 타겟 이미지이다.

도 4b는 본 명세서에서 설명되는 일부 실시예들에 따른, 머리 편집기에 의해 생성된 예시적인 합성 이미지이다.

도 5는 본 명세서에서 설명되는 일부 실시예들에 따른, 예시적인 타겟 이미지이다.

도 6a는 본 명세서에서 설명되는 일부 실시예들에 따른, 타겟 이미지로부터 분리된 예시적인 제1 타겟 머리이다.

도 6b는 본 명세서에서 설명되는 일부 실시예들에 따른, 타겟 이미지로부터 분리된 예시적인 제2 타겟 머리이다.

도 7a는 본 명세서에서 설명되는 일부 실시예들에 따른, 소스 이미지로부터 분리된 예시적인 제1 소스 머리이다.

도 7b는 본 명세서에서 설명되는 일부 실시예들에 따른, 소스 이미지로부터 분리된 예시적인 제2 소스

머리이다.

도 8a는 본 명세서에서 설명되는 일부 실시예들에 따른, 제1 타겟 머리와 정렬되는 예시적인 제1 소스 머리이다.

도 8b는 본 명세서에서 설명되는 일부 실시예들에 따른, 제2 타겟 머리와 정렬되는 예시적인 제2 소스 머리이다.

도 9는 본 명세서에서 설명되는 일부 실시예들에 따른, 인페인팅을 위해 분석되는 예시적인 합성 이미지이다.

도 10은 본 명세서에서 설명되는 일부 실시예들에 따른, 소스 머리들이 도 5의 타겟 머리들을 대체하는 예시적인 합성 이미지이다.

도 11은 본 명세서에서 설명되는 일부 실시예들에 따른, 합성 이미지를 생성하는 기계 학습 모델의 예시적인 블록도이다.

도 12a는 본 명세서에서 설명되는 일부 실시예들에 따른, 얼굴 편집기에 대한 예시적인 타겟 이미지이다.

도 12b는 본 명세서에서 설명되는 일부 실시예들에 따른, 얼굴 편집기에 의해 생성된 예시적인 합성 이미지이다.

도 13은 본 명세서에서 설명되는 일부 실시예들에 따른, 베스트 테이크를 선택하기 위한 예시적인 사용자 인터페이스들을 예시한다.

도 14는 본 명세서에서 설명되는 일부 실시예들에 따른, 합성 이미지를 생성하기 위한 예시적인 방법의 흐름도를 예시한다.

발명을 실시하기 위한 구체적인 내용

[0014] **개요**

[0015] 미디어 애플리케이션은, 합성 이미지 내의 대상체들 중 하나 이상이 소스 이미지들로부터의 머리들 및/또는 얼굴들을 갖는, 합성 이미지를 생성한다. 이미지들의 부분들을 조합하려는 이전의 시도들은, 끊김들이 가시적이며, 대체 객체가 원래 객체와 정렬되지 않는 경우 원래 객체로부터의 픽셀들이 가시적이며, 객체들이 가려지면 아티팩트를 초래하는 등 비사실적인 합성 이미지들을 초래할 수 있다.

[0016] 일부 실시예들에서, 미디어 애플리케이션은 소스 이미지 및 타겟 이미지를 수신하고, 합성 이미지를 생성하기 위해 머리 편집기를 사용할지 및/또는 얼굴 편집기를 사용할지를 결정한다. 예를 들어, 미디어 애플리케이션은, 타겟 이미지 내의 두 대상체들의 머리들이 충분히 떨어져 있거나 머리가 객체에 의해 가려지지 않은 것에 기초하여, 머리 편집기를 선택할 수 있다. 다른 예에서, 미디어 애플리케이션은, 소스 이미지와 타겟 이미지 사이의 얼굴의 각도의 포즈의 차이가 소스 이미지의 부분들이 타겟 이미지에 추가될 수 있을 정도로 충분히 가까운 것에 기초하여, 얼굴 편집기를 선택할 수 있다.

[0017] 머리 편집기는 타겟 이미지 내의 머리를 소스 이미지 내의 머리로 대체할 수 있다. 얼굴 편집기는 소스 이미지로부터의 얼굴의 부분들에 기초하여 타겟 이미지로부터의 눈들 및 입들과 같은, 얼굴의 부분들을 조정할 수 있다. 예를 들어, 얼굴 편집기는 타겟 이미지 내의 얼굴의 부분들에 대한 픽셀 값들을 컴퓨트하기 위한 임베딩을 사용할 수 있다. 미디어 애플리케이션은 타겟 이미지와 소스 이미지의 조합들로부터 합성 이미지를 생성한다.

[0018] 합성 이미지를 생성하는 것은, 이미지 데이터를 분석하여 가림을 결정하는 것 및 이미지 데이터를 변환하는 것을 포함하고, 및 합성 이미지를 비사실적으로 보이게 할 수 있는 가시적인 끊김들 또는 다른 결함들을 회피하기 위해 인페인팅을 수행할 수 있다. 달리 말하면, 하나 이상의 소스 이미지들로부터 합성 이미지를 생성하기 위한 본 명세서에서 설명되는 기법들은 하나 이상의 소스 이미지들의 사실감을 끊김없이 보존할 수 있다.

[0019] **예시적인 환경**

[0020] 도 1은 예시적인 환경(100)의 블록도를 예시한다. 일부 실시예들에서, 환경(100)은 네트워크(105)에 결합된 미디어 서버(101), 사용자 디바이스(115a), 및 사용자 디바이스(115n)를 포함한다. 사용자들(125a, 125n)은 개별 사용자 디바이스들(115a, 115n)과 연관될 수 있다. 일부 실시예들에서, 환경(100)은 도 1에서 도시되지 않은 다른 서버들 또는 디바이스들을 포함할 수 있다. 도 1 및 나머지 도면들에서, 참조 번호 뒤의 문자, 예를 들어, "115a"는 특정 참조 번호를 갖는 엘리먼트에 대한 참조를 나타낸다. 이어지는 문자가 없는 본문에서의 참조 번

호, 예를 들어, "115"는 해당 참조 번호를 갖는 엘리먼트의 실시예들에 대한 일반적인 참조를 나타낸다.

[0021] 미디어 서버(101)는 프로세서, 메모리, 및 네트워크 통신 하드웨어를 포함할 수 있다. 일부 실시예들에서, 미디어 서버(101)는 하드웨어 서버이다. 미디어 서버(101)는 신호 라인(102)을 통해 네트워크(105)에 통신가능하게 결합된다. 신호 라인(102)은 이더넷, 동축 케이블, 광섬유 케이블 등과 같은 유선 연결, 또는 Wi-Fi®, Bluetooth®, 또는 다른 무선 기술과 같은 무선 연결일 수 있다. 일부 실시예들에서, 미디어 서버(101)는 네트워크(105)를 통해 사용자 디바이스들(115a, 115n) 중 하나 이상으로 그리고 이로부터 데이터를 송수신한다. 미디어 서버(101)는 미디어 애플리케이션(103a) 및 데이터베이스(199)를 포함할 수 있다.

[0022] 데이터베이스(199)는 기계 학습 모델, 트레이닝 데이터 세트, 이미지 등을 저장할 수 있다. 데이터베이스(199)는 또한, 사용자들(125)과 연관된 소셜 네트워크 데이터, 사용자들(125)에 대한 사용자 선호도 등을 저장할 수 있다.

[0023] 사용자 디바이스(115)는 하드웨어 프로세서에 결합된 메모리를 포함하는 컴퓨팅 디바이스일 수 있다. 예를 들어, 사용자 디바이스(115)는 모바일 디바이스, 태블릿 컴퓨터, 모바일 전화, 웨어러블 디바이스, 머리 장착형 디스플레이, 모바일 이메일 디바이스, 휴대용 게임 플레이어, 휴대용 음악 플레이어, 리더기 디바이스, 또는 네트워크(105)에 액세스할 수 있는 다른 전자 디바이스를 포함할 수 있다.

[0024] 예시된 구현예에서, 사용자 디바이스(115a)는 신호 라인(108)을 통해 네트워크(105)에 결합되고, 사용자 디바이스(115n)는 신호 라인(110)을 통해 네트워크(105)에 결합된다. 미디어 애플리케이션(103)은 사용자 디바이스(115a) 상에 미디어 애플리케이션(103b)으로서 및/또는 사용자 디바이스(115n) 상에 미디어 애플리케이션(103c)으로서 저장될 수 있다. 신호 라인들(108 및 110)은 이더넷, 동축 케이블, 광섬유 케이블 등과 같은 유선 연결, 또는 Wi-Fi®, Bluetooth®, 또는 다른 무선 기술과 같은 무선 연결일 수 있다. 사용자 디바이스들(115a, 115n)은 각각, 사용자들(125a, 125n)에 의해 액세스된다. 도 1에서의 사용자 디바이스들(115a, 115n)은 예로서 사용된다. 도 1은 두 개의 사용자 디바이스들(115a 및 115n)을 예시하지만, 본 개시는 하나 이상의 사용자 디바이스들(115)을 갖는 시스템 아키텍처에 적용된다.

[0025] 미디어 애플리케이션(103)은 미디어 서버(101) 또는 사용자 디바이스(115) 상에 저장될 수 있다. 일부 실시예들에서, 본 명세서에서 설명되는 동작들은 미디어 서버(101) 또는 사용자 디바이스(115) 상에서 수행된다. 일부 실시예들에서, 일부 동작들은 미디어 서버(101) 상에서 수행될 수 있고, 일부는 사용자 디바이스(115) 상에서 수행될 수 있다. 동작들의 수행은 사용자 설정에 따른다. 예를 들어, 사용자(125a)는 동작들이 미디어 서버(101) 상이 아닌 자신의 개별 디바이스(115a) 상에서 수행되어야 한다는 설정을 지정할 수 있다. 이러한 설정으로, 본 명세서에서 설명되는 동작들은 전적으로 사용자 디바이스(115a) 상에서 수행되고, 미디어 서버(101) 상에서는 어떠한 동작도 수행되지 않는다. 또한, 사용자(125a)는 사용자의 이미지들 및/또는 다른 데이터가 미디어 서버(101) 상이 아닌 사용자 디바이스(115a) 상에 로컬로만 저장되도록 지정할 수 있다. 이러한 설정으로, 어떠한 사용자 데이터도 미디어 서버(101)로 송신되거나 미디어 서버(101) 상에 저장되지 않는다. 미디어 서버(101)로의 사용자 데이터의 송신, 미디어 서버(101)에 의한 이러한 데이터의 임의의 일시적 또는 영구적 저장, 및 미디어 서버(101)에 의한 이러한 데이터에 대한 동작들의 수행은 사용자가 미디어 서버(101)에 의한 송신, 저장, 및 동작들의 수행에 동의한 경우에만 수행된다. 사용자들에게는 예를 들어, 사용자들이 미디어 서버(101)의 사용을 인에이블하거나 디스에이블할 수 있도록, 언제든지 설정을 변경하기 위한 옵션들이 제공된다.

[0026] 기계 학습 모델들(예를 들어, 신경망 또는 다른 타입의 모델)은 하나 이상의 동작들에 이용되는 경우, 특정 사용자 권한으로, 사용자 디바이스(115) 상에 로컬로 저장되고 이용된다. 서버 측 모델들은 사용자에 의해 허용된 경우에만 사용된다. 또한, 사용자 디바이스(115) 상에서 사용하기 위해 트레이닝된 모델이 제공될 수 있다. 이러한 사용 동안, 사용자(125)에 의해 허용되는 경우, 모델의 온-디바이스 트레이닝이 수행될 수 있다. 업데이트된 모델 파라미터들은 예를 들어, 연합 학습을 가능하게 하기 위해, 사용자(125)에 의해 허용되는 경우 미디어 서버(101)에 송신될 수 있다. 모델 파라미터들은 어떠한 사용자 데이터도 포함하지 않는다.

[0027] 미디어 애플리케이션(103)은 소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하며, 소스 이미지 및 타겟 이미지는 적어도 대상체를 포함한다. 미디어 애플리케이션(103)은 이미지들의 세트에 기초하여, 머리 편집기를 사용할지 및/또는 얼굴 편집기를 사용할지를 결정한다. 머리 편집기를 사용하기로 결정하는 것에 응답하여, 미디어 애플리케이션(103)은 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하고, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지를 생성한다.

[0028] 일부 실시예들에서, 미디어 애플리케이션(103)은 중앙 프로세싱 유닛(CPU), 필드 프로그래밍가능 게이트 어레이(FPGA), 애플리케이션 특정 집적 회로(ASIC), 기계 학습 프로세서/공동 프로세서, 임의의 다른 타입의 프로세서, 또는 이들의 조합을 포함하는 하드웨어를 사용하여 구현될 수 있다. 일부 실시예들에서, 미디어 애플리케이션(103a)은 하드웨어와 소프트웨어의 조합을 사용하여 구현될 수 있다.

[0029] **예시적인 컴퓨팅 디바이스**

[0030] 도 2는 본 명세서에서 설명되는 하나 이상의 피처들을 구현하기 위해 사용될 수 있는 예시적인 컴퓨팅 디바이스(200)의 블록도이다. 컴퓨팅 디바이스(200)는 임의의 적합한 컴퓨터 시스템, 서버, 또는 다른 전자 또는 하드웨어 디바이스일 수 있다. 일 예에서, 컴퓨팅 디바이스(200)는 미디어 애플리케이션(103a)을 구현하기 위해 사용되는 미디어 서버(101)이다. 다른 예에서, 컴퓨팅 디바이스(200)는 사용자 디바이스(115)이다.

[0031] 일부 실시예들에서, 컴퓨팅 디바이스(200)는 프로세서(235), 메모리(237), 입력/출력(I/O) 인터페이스(239), 디스플레이(241), 카메라(243), 및 저장 디바이스(245)를 포함하며, 모두 버스(218)를 통해 결합된다. 프로세서(235)는 신호 라인(222)을 통해 버스(218)에 결합될 수 있고, 메모리(237)는 신호 라인(224)을 통해 버스(218)에 결합될 수 있고, I/O 인터페이스(239)는 신호 라인(226)을 통해 버스(218)에 결합될 수 있고, 디스플레이(241)는 신호 라인(228)을 통해 버스(218)에 결합될 수 있고, 카메라(243)는 신호 라인(230)을 통해 버스(218)에 결합될 수 있으며, 저장 디바이스(245)는 신호 라인(232)을 통해 버스(218)에 결합될 수 있다.

[0032] 프로세서(235)는 컴퓨팅 디바이스(200)의 기본 동작들을 제어하고 프로그램 코드를 실행하기 위한 하나 이상의 프로세서들 및/또는 프로세싱 회로들일 수 있다. "프로세서"는 데이터, 신호들 또는 다른 정보를 프로세싱하는 임의의 적합한 하드웨어 시스템, 메커니즘 또는 컴포넌트를 포함한다. 프로세서는 (예를 들어, 단일 코어, 듀얼 코어, 또는 멀티 코어 구성의) 하나 이상의 코어들을 갖는 범용 중앙 프로세싱 유닛(CPU), (예를 들어, 멀티 프로세서 구성의) 다수의 프로세싱 유닛들, 그래픽 프로세싱 유닛(GPU), 필드 프로그래밍가능 게이트 어레이(FPGA), 애플리케이션 특정 집적 회로(ASIC), 복합 프로그래밍가능 로직 디바이스(CPLD), 기능을 달성하기 위한 전용 회로부, 신경망 모델 기반 프로세싱을 구현하기 위한 특수 목적 프로세서, 신경망 회로들, 행렬 컴퓨테이션(예를 들어, 행렬 곱셈)에 최적화된 프로세서들을 갖는 시스템, 또는 다른 시스템들을 포함할 수 있다. 일부 실시예들에서, 프로세서(235)는 신경망 프로세싱을 구현하는 하나 이상의 공동 프로세서들을 포함할 수 있다. 일부 실시예들에서, 프로세서(235)는 확률적 출력을 생성하기 위해 데이터를 프로세싱하는 프로세서일 수 있으며, 예를 들어, 프로세서(235)에 의해 생성되는 출력은 부정밀할 수 있거나 예상 출력으로부터의 범위 내에서 정확할 수 있다. 프로세싱은 특정 지리적 위치로 제한될 필요가 없거나 시간적 제한을 가질 필요가 없다. 예를 들어, 프로세서는 그 기능들을 실시간, 오프라인, 일괄 모드 등으로 수행할 수 있다. 프로세싱의 부분들은 상이한(또는 동일한) 프로세싱 시스템에 의해 상이한 시간에 그리고 상이한 위치에서 수행될 수 있다. 컴퓨터는 메모리와 통신하는 임의의 프로세서일 수 있다.

[0033] 메모리(237)는 프로세서(235)에 의한 액세스를 위해 전형적으로 컴퓨팅 디바이스(200)에서 제공되고, 프로세서 또는 프로세서들의 세트들에 의한 실행을 위한 명령어들을 저장하기에 적합하고, 프로세서(235)로부터 분리되어 있고/있거나 이와 통합되어 있는, 랜덤 액세스 메모리(RAM), 판독 전용 메모리(ROM), 전기 소거가능 판독 전용 메모리(EEPROM), 플래시 메모리 등과 같은, 임의의 적합한 프로세서 판독가능 저장 매체일 수 있다. 메모리(237)는 미디어 애플리케이션(103)을 포함하여, 프로세서(235)에 의해 컴퓨팅 디바이스(200) 상에서 동작하는 소프트웨어를 저장할 수 있다.

[0034] 메모리(237)는 운영 체제(262), 기타 애플리케이션들(264), 및 애플리케이션 데이터(266)를 포함할 수 있다. 기타 애플리케이션들(264)은 예를 들어, 이미지 라이브러리 애플리케이션, 이미지 관리 애플리케이션, 이미지 갤러리 애플리케이션, 통신 애플리케이션, 웹 호스팅 엔진 또는 애플리케이션, 미디어 공유 애플리케이션 등을 포함할 수 있다. 본 명세서에서 개시되는 하나 이상의 방법들은 여러 환경들 및 플랫폼들에서, 예를 들어, 임의의 타입의 컴퓨팅 디바이스 상에서 실행될 수 있는 독립형 컴퓨터 프로그램으로서, 웹 페이지를 갖는 웹 애플리케이션으로서, 모바일 컴퓨팅 디바이스 상에서 실행되는 모바일 애플리케이션("앱")으로서 동작할 수 있다.

[0035] 애플리케이션 데이터(266)는 컴퓨팅 디바이스(200)의 하드웨어 또는 기타 애플리케이션들(264)에 의해 생성되는 데이터일 수 있다. 예를 들어, 애플리케이션 데이터(266)는 이미지 라이브러리 애플리케이션에 의해 사용되는 이미지들 및 기타 애플리케이션들(264)(예를 들어, 소셜 네트워킹 애플리케이션)에 의해 식별되는 사용자 액션들 등을 포함할 수 있다.

[0036] I/O 인터페이스(239)는 컴퓨팅 디바이스(200)를 다른 시스템들 및 디바이스들과 인터페이싱할 수 있게 하는 기

능들을 제공할 수 있다. 인터페이싱된 디바이스들은 컴퓨팅 디바이스(200)의 일부로서 포함될 수 있거나, 분리되어 컴퓨팅 디바이스(200)와 통신할 수 있다. 예를 들어, 네트워크 통신 디바이스들, 저장 디바이스들(예를 들어, 메모리(237) 및/또는 저장 디바이스(245)), 및 입력/출력 디바이스들은 I/O 인터페이스(239)를 통해 통신할 수 있다. 일부 실시예들에서, I/O 인터페이스(239)는 입력 디바이스들(키보드, 포인팅 디바이스, 터치스크린, 마이크로폰, 스캐너, 센서 등) 및/또는 출력 디바이스들(디스플레이 디바이스, 스피커 디바이스, 프린터, 모니터 등)과 같은 인터페이스 디바이스들에 연결될 수 있다.

[0037] I/O 인터페이스(239)에 연결될 수 있는 인터페이싱된 디바이스들의 일부 예들은 본 명세서에서 설명되는 바와 같은 출력 애플리케이션의 콘텐츠, 예를 들어, 이미지, 비디오, 및/또는 사용자 인터페이스를 디스플레이하고, 사용자로부터 터치(또는 제스처) 입력을 수신하기 위해 사용될 수 있는 디스플레이(241)를 포함할 수 있다. 예를 들어, 디스플레이(241)는 뷰파인더 상에 그래픽 가이드를 포함하는 사용자 인터페이스를 디스플레이하기 위해 이용될 수 있다. 디스플레이(241)는 LCD(liquid crystal display), LED(light emitting diode), 또는 플라즈마 디스플레이 스크린, CRT(cathode ray tube), 텔레비전, 모니터, 터치스크린, 3차원 디스플레이 스크린, 또는 다른 시각적 디스플레이 디바이스와 같은 임의의 적합한 디스플레이 디바이스를 포함할 수 있다. 예를 들어, 디스플레이(241)는 모바일 디바이스 상에 제공되는 평면 디스플레이 스크린, 안경 폼 팩터 또는 헤드셋 디바이스에 임베딩된 다중 디스플레이 스크린, 또는 컴퓨터 디바이스용 모니터 스크린일 수 있다.

[0038] 카메라(243)는 이미지 및/또는 비디오를 촬영(capture)할 수 있는 임의의 타입의 이미지 촬영 디바이스일 수 있다. 일부 실시예들에서, 카메라(243)는 I/O 인터페이스(239)가 미디어 애플리케이션(103)에 송신하는 이미지 또는 비디오를 촬영한다.

[0039] 저장 디바이스(245)는 미디어 애플리케이션(103)과 관련된 데이터를 저장한다. 예를 들어, 저장 디바이스(245)는 라벨링된 이미지, 기계 학습 모델, 기계 학습 모델로부터의 출력 등을 포함하는 트레이닝 데이터 세트를 저장할 수 있다.

[0040] 도 2는 메모리(237)에 저장된, 예시적인 미디어 애플리케이션(103)을 예시하며, 이는 이미지 모듈(202), 머리 편집기(204), 및 얼굴 편집기(206)를 포함한다.

[0041] 이미지 모듈(202)은 이미지들의 세트를 포함하는 사용자 인터페이스를 디스플레이하기 위한 그래픽 데이터를 생성한다. 이미지들의 세트는 I/O 인터페이스(239)를 통해 컴퓨팅 디바이스(200)의 카메라(243)로부터 그리고/또는 미디어 서버(101)로부터 수신될 수 있다. 예를 들어, 이미지들의 세트는 카메라(243)에 의해 촬영된 연사 이미지들로부터의 이미지들을 포함할 수 있다. 연사 이미지들은 짧은 시구간에 걸쳐 연속으로 빠르게 촬영된 다수의 사진들을 포함할 수 있다. 이미지들의 세트는 하나 이상의 소스 이미지들, 및 하나 이상의 대상체들을 포함하는 타겟 이미지를 포함한다. 하나 이상의 대상체들은 사람, 동물 등일 수 있다.

[0042] 이미지 모듈(202)은 이미지들의 세트에서의 임의의 이미지를 수정하기 위한 권한을 사용자로부터 획득한다. 사용자는 본 명세서에서 설명된 시스템들, 프로그램들, 또는 피처들이 사용자 정보(예를 들어, 이미지 내의 사용자의 식별, 사용자의 선호도, 또는 사용자의 현재 위치)의 수집 또는 사용을 가능하게 할 수 있는 경우와 그 때, 그리고 사용자가 서버로부터 콘텐츠 또는 통신을 전송받는 경우 둘 모두에 관해 사용자가 선출을 할 수 있게 하는 제어를 제공받을 수 있다. 또한, 특정 데이터는 개인 식별가능 정보가 제거되도록, 저장되거나 사용되기 전에 하나 이상의 방식들로 처리될 수 있다. 예를 들어, 사용자에 대해 어떠한 개인 식별가능 정보도 결정될 수 없도록 사용자의 신원이 처리될 수 있거나, 사용자의 특정 위치가 결정될 수 없도록 사용자의 지리적 위치가 (도시, 우편 번호, 또는 주 레벨과 같은) 위치 정보가 획득되는 곳에서 일반화될 수 있다. 이에 따라, 사용자는 사용자에 관해 어떤 정보가 수집되는지, 그 정보가 어떻게 사용되는지, 및 사용자에게 어떤 정보가 제공되는지에 대한 제어를 가질 수 있다.

[0043] 일부 실시예들에서, 이미지 모듈(202)은 합성 이미지를 생성하기 위해 사용될 하나 이상의 소스 이미지들 및 타겟 이미지를 선택한다. 예를 들어, 이미지 모듈(202)은 이미지 내의 타입들, 대상체들의 위치, 조명 등과 같은, 합성 인자들에 기초하여 이미지들의 세트에서의 각 이미지에 대한 합성 점수를 자동으로 생성할 수 있다. 예를 들어, 타겟 이미지는 그룹 내의 대상체들이 잘 위치되는 것뿐만 아니라 최상의 경관 합성을 갖는 것에 기초하여서도 선택될 수 있다. 다른 예에서, 타겟 이미지는, 가장 많은 수의 대상체들이 카메라를 보는 것, 미소를 짓는 것, 눈을 뜨고 있는 것 등에 기초하여 선택될 수 있다. 종합적으로 높은 합성 점수를 갖는 이미지가 타겟 이미지에 사용될 수 있다. 일부 실시예들에서, 사용자는 이미지들의 세트 내의 특정 이미지를 타겟 이미지로서 선택할 수 있고, 이미지들의 세트 내의 하나 이상의 다른 이미지들은 소스 이미지들로서 제시될 수 있다.

[0044] 이미지 모듈(202)은 이미지 내의 대상체들에 기초하여 하나 이상의 소스 이미지들을 선택할 수 있다. 예를 들어, 이미지 모듈(202)은 품질 메트릭들에 기초하여 이미지 내의 각 대상체에 대한 얼굴 점수들을 생성할 수 있다. 예를 들어, 각 얼굴은 얼굴이 미소를 특징으로 하는지, 눈을 뜨고 있는지, 다른 사람들 또는 객체들에 의해 사진에서 완전히 또는 부분적으로 가려지는지, 및/또는 움직임이 없는지(예를 들어, 흐릿하지 않은지)에 기초하여 점수가 매겨질 수 있다. 이미지 모듈(202)은, (예를 들어, 카메라로부터 기울어진, 떨어져 회전된 등) 다른 각도들 대신에 수직 평면 내에 있는(예를 들어, 카메라 정면을 향해 있는) 머리들에 대한 더 높은 머리 점수와 같이, 대상체의 머리의 위치에 기초하여 머리 점수들을 생성할 수 있다. 이러한 점수 매기기는, 사진 내의 얼굴을 검출하고 품질 점수를 매기는 기준들을 적용할 수 있는 트레이닝된 기계 학습 모델들과 같은, 이미지 검출 및 인식 기법들을 사용하여 수행될 수 있다.

[0045] 점수들에 기초하여, 이미지 모듈(202)은 이미지들 내의 얼굴들을 순위화할 수 있다. 예를 들어, 이미지 모듈(202)은 이미지들의 세트의 상이한 이미지들 내의 다수의 특정 얼굴들이 동일한 대상체(사람, 동물 등)에 속한다고 결정할 수 있고, 이러한 각 대상체는 이미지들로부터의 얼굴들의 순위화된 리스트와 연관될 수 있다. 이미지 모듈(202)은 타겟 이미지 및 하나 이상의 소스 이미지들을 자동으로 선택할 수 있거나, 또는 이미지 모듈(202)은 사용자가 특정 이미지의 선택을 타겟 이미지로서 확정하기 위한 사용자에 대한 제안으로서 최상위권 이미지들을 제시할 수 있다.

[0046] 일부 실시예들에서, 이미지 모듈(202)은, 사용자가 이미지들의 세트로부터 타겟 이미지 및 하나 이상의 소스 이미지들을 선택하기 위한 옵션들을 제공하는 사용자 인터페이스를 디스플레이하기 위한 그래픽 데이터를 생성한다. 사용자 인터페이스에 포함되는 이미지들의 세트는, 촬영된 연사 이미지들, 특정 시구간 동안 촬영된 이미지들(예를 들어, 지난 24시간, 특정 장소에서 촬영된 이미지들 등)로부터 온 것일 수 있다.

[0047] 도 3a는 사용자가 이미지들의 세트로부터 타겟 이미지를 지정하기 위한 옵션들을 포함하는 예시적인 사용자 인터페이스(300)이다. 이러한 예에서, 사용자는 타겟 이미지로서 사용될 이미지(305)를 선택한다. 일부 실시예들에서, 이미지들의 세트로부터의 특정 이미지는, 예를 들어, (예를 들어, 테두리, 아이콘 등으로) 하이라이트되어, 타겟 이미지로서 제시될 수 있다. 사용자는 제시된 이미지 또는 임의의 다른 이미지를 타겟 이미지로서 선택할 수 있다.

[0048] 도 3b는 사용자가 이미지들의 세트로부터 하나 이상의 소스 이미지들을 지정하기 위한 옵션들을 포함하는 예시적인 사용자 인터페이스(325)이다. 이러한 예에서, 사용자는 제1 대상체에 대한 소스 이미지로서 사용될 제1 이미지(330) 및 제2 대상체에 대한 소스 이미지로서 사용될 제2 이미지(335)를 선택한다. 일부 실시예들에서, 이미지 모듈(202)은 이미지 내의 특정 대상체를 선택하기 위한 옵션들을 포함하는 사용자 인터페이스를 생성할 수 있다. 예를 들어, 컴퓨팅 디바이스(200)는 대상체의 더블 클릭, 대상체의 머리 주위의 원 등의 형태의 사용자 입력을 수신할 수 있다.

[0049] 일부 실시예들에서, 단일 이미지가 둘 이상의 대상체들에 대한 소스 이미지로서 선택될 수 있다. 일부 실시예들에서, 각 대상체는 상이한 소스 이미지와 연관될 수 있다. 일부 실시예들에서, 소스 이미지의 제안이 하나 이상의 대상체들에 대해 제공될 수 있다. 예를 들어, 타겟 이미지가 기울어진 머리를 갖는 제1 대상체, 및 눈을 감은 제2 대상체를 갖는 경우, 제1 대상체가 기울어지지 않고 카메라를 향하는 머리를 갖는 제1 소스 이미지, 및 제2 대상체가 열린 눈을 뜨고 있는 제2 소스 이미지가 각각의 소스 이미지들로서 제시될 수 있다. 일부 실시예들에서, 조명, 타겟 이미지와 특정 소스 이미지들의 촬영 사이의 지속 시간, 소스 이미지와 타겟 이미지 내의 대상체의 위치 사이의 거리 등과 같은 다른 인자들이 추천될 특정 소스 이미지들을 선택하는 데 사용될 수 있다.

[0050] 일부 실시예들에서, 사용자 인터페이스는 다른 소스 이미지를 검색하기 위한 옵션을 포함할 수 있다. 예를 들어, 사용자 인터페이스는, 사용자가 카메라 롤에서 이미지들을 스크롤하고(또는 그 외 브라우징하고) 소스 이미지를 선택하기 위한 옵션을 포함할 수 있다. 일부 실시예들에서, 카메라 롤로부터의 소스 이미지가 상이한 조명 조건들, 상이한 음영들 등으로 촬영되었을 수 있기 때문에, 이미지 모듈(202)은 소스 이미지를 타겟 이미지의 대응하는 속성들과 일치하는 색상(및/또는 밝기, 화이트 밸런스, 콘트라스트 등과 같은 다른 이미지 속성들)을 갖도록 수정할 수 있다.

[0051] 이미지 모듈(202)은 합성 이미지를 생성하기 위해 머리 편집기(204)를 사용할지, 얼굴 편집기(206)를 사용할지, 또는 둘 다 사용할지를 결정한다. 이미지 모듈(202)은 타겟 이미지 또는 소스 이미지 내의 머리들의 근접도, 타겟 이미지 또는 소스 이미지 내의 머리들의 각도, 및 타겟 이미지 또는 소스 이미지 내의 머리의 가림과 같은,

상이한 기준들을 사용하여 결정을 내릴 수 있다.

[0052] 일부 실시예들에서, 이미지 모듈(202)은 상이한 분류기들에 의해 평가되는 상이한 인자들에 기초하여 머리 편집기(204)를 사용할지 또는 얼굴 편집기(206)를 사용할지를 결정한다. 일부 실시예들에서, 이미지 모듈(202)은 결정이 하나의 단일 인자가 아닌 인자들의 총체에 기초하여 이루어지도록 상이한 인자들의 가중 합 또는 로지스틱 회귀(logistic regression)를 수행한다. 일부 실시예들에서, 인자들 중 일부는 머리들 중 하나가 객체에 의해 70% 가려진 경우와 같이, 결정적일 수 있다.

[0053] 일부 실시예들에서, 분류기들은 이미지들의 트레이닝 세트에서 머리 편집기(204)로부터의 머리 편집본들 및 얼굴 편집기(206)로부터의 얼굴 편집본들의 등급들로부터 트레이닝된다. 예를 들어, 트레이닝 데이터는, 타겟 이미지, 소스 이미지, 머리 편집기(204) 및/또는 얼굴 편집기(206)에 의해 생성된 합성 이미지, 및 사람 또는 품질 알고리즘에 의해 대응하는 등급이 제공될 수 있는 경우 합성 이미지에 대한 대응하는 등급을 포함할 수 있다. 등급들을 수신한 결과로서, 이미지 모듈(202)은 머리 편집기(204) 및 얼굴 편집기(206)에 의해 생성되는 합성 이미지들의 품질을 개선하기 위해 분류기 가중치들을 재계산할 수 있다.

[0054] 일부 실시예들에서, 이미지 모듈(202)은 타겟 이미지 내의 각 타겟 머리/타겟 얼굴 주위에 경계 박스들을 생성하고, 타겟 이미지 내의 경계 박스들 사이의 거리에 기초하여 타겟 머리를 소스 머리로 대체하기 위해 머리 편집기(204)를 사용할지 여부를 결정한다. 예를 들어, 경계 박스들 사이의 오버랩이 있어, 대상체들의 머리들이 서로 가까이 있다는 것을 나타내는 경우, 다수의 대상체들이 함께 가까이 있을 때 머리 위치들의 변동이 합성 이미지에서 인페이팅되는 배경의 더 큰 섹션을 초래하는 것 및 대상체들에서의 오버랩 가능성이 타겟 이미지와 하나 이상의 소스 이미지들 내의 대상체들을 정렬시키는 것을 복잡하게 한다는 것에 기인하여, 머리 편집기(204)가 사용되지 않을 수 있다. 일부 실시예들에서, 이미지 모듈(202)은 연속 함수에 기초하여 머리 편집기(204)를 사용할지 여부를 결정한다. 일부 실시예들에서, 연속 함수는 경계 박스들 사이의 가중 거리 및/또는 경계 박스들 사이의 오버랩의 가중 퍼센티지에 기초하며, 여기서 가중치들의 값은 이미지 모듈(202)에 의해 사용되는 분류기의 트레이닝 동안 학습될 수 있다.

[0055] 일부 실시예들에서, 이미지 모듈(202)은 타겟 머리의 제1 각도와 소스 머리의 제2 각도 사이의 각도 차이에 기초하여 머리 편집기(204)를 사용할지 여부를 결정한다. 얼굴 편집기(206)는 소스 이미지의 머리의 각도가 타겟 이미지의 머리의 각도로부터 임계 차이보다 더 큰 경우 불만족스럽거나 낮은 품질의 합성 이미지들을 생성할 수 있다.

[0056] 이미지 모듈(202)은 제1 각도와 제2 각도 사이의 차이에 비선형 함수를 적용하는 것에 기초하여 얼굴 편집기(206)를 사용하기로 결정할 수 있다. 일부 실시예들에서, 이미지 모듈(202)은 제1 각도에 코사인을 적용하고, 제2 각도에 코사인을 적용하며, 제1 각도와 제2 각도 사이의 차이를 결정한다. 일부 실시예들에서, 이미지 모듈(202)은, 각도들 사이의 차이가 임계 각도 차이를 초과하는 경우, 이미지 모듈(202)이 얼굴 편집기(206)가 아닌 머리 편집기(204)를 사용하기로 결정하도록 얼굴 편집기(206)를 사용할지 여부를 결정하는 데 임계 각도 차이를 사용한다.

[0057] 일부 실시예들에서, 이미지 모듈(202)은 타겟 머리의 가림 또는 소스 머리의 가림에 기초하여 머리 편집기(204)를 사용할지 여부를 결정한다. 이미지 모듈(202)이 가려진 타겟 이미지 또는 소스 이미지 중 어느 하나를 수신할 때, 결과적인 합성 이미지는 소스 이미지가 가려질 때 더 높은 실패율을 갖고/거나 낮은 품질을 갖는다. 일부 실시예들에서, 이미지 모듈(202)은 소스 머리 또는 타겟 머리가 가려지는지 여부를 결정하기 위해 이미지들의 컬러 히스토그램을 사용할 수 있다. 예를 들어, 이미지 모듈(202)은 가림을 식별하기 위해 히스토그램에서의 특정 채널의 평균을 사용할 수 있다.

[0058] 일부 실시예들에서, 이미지 모듈(202)은 대상체로부터 이미지 경계들까지의 거리에 기초하여 머리 편집기(204)를 사용할지, 얼굴 편집기(206)를 사용할지, 또는 머리 편집기(204)와 얼굴 편집기(206) 둘 다를 사용할지를 결정한다. 예를 들어, 머리가 이미지 경계에 가깝고 머리 편집기(204)가 이미지 경계에 가까운 머리의 포즈를 변경하는 경우, 이는 머리의 일부가 이미지 경계로부터 크롭되는 것을 초래할 수 있다(이에 따라, 대상체 머리가 이미지 내에 완전히 있지 않으므로, 만족스럽지 않은 합성 이미지를 제공함).

[0059] 머리 편집기(204)는 타겟 머리를 소스 머리로 대체한다. 일부 실시예들에서, 머리 편집기(204)는 타겟 이미지 내의 타겟 머리와 연관된 타겟 픽셀들을 소스 이미지 내의 소스 머리로부터의 소스 픽셀들로 대체한다. 머리 편집기(204)는 대상체의 모발을 포함하는 머리 마스크를 생성하고, 소스 이미지로부터 머리 마스크를 세그먼트화하며, 머리 마스크 내의 픽셀들을 타겟 이미지에 적용할 수 있다. 머리 편집기(204)는 소스 머리의 위치 및 스

케일을 타겟 머리의 치수들과 일치하도록 수정할 수 있다.

[0060] 머리 편집기(204)는 타겟 머리를 소스 머리로 대체하는 것이 타겟 머리와 소스 머리가 오버랩되지 않는 부분들을 초래하는 경우들에 인페인팅을 수행한다. 이는 타겟 머리 및 소스 머리가 상이한 각도들과 연관되는 경우에 발생할 수 있다. 이는 또한, 대체 전에, 타겟 및 소스 이미지들에서의 타겟 머리와 소스 머리의 각도가 상이하고 머리 편집기(204)가 타겟 머리의 각도를 소스 머리의 각도와 정렬시킬 때 발생할 수 있다. 타겟 머리를 소스 머리와 정렬시키는 것은 소스 머리가 오버랩되지 않는 타겟 머리의 부분들을 식별할 수 있다. 머리 편집기(204)는 소스 머리가 아닌 타겟 머리와 연관된 타겟 이미지 내의 나머지 픽셀들을 식별하고, 나머지 타겟 픽셀들을 대체함으로써 나머지 타겟 픽셀들의 인페인팅을 수행할 수 있다. 픽셀들을 인페인팅하는 것은 소스 픽셀들과 나머지 픽셀들 사이의 거리를 결정하는 것, 및 소스 픽셀들과의 유사도 및 소스 픽셀들까지의 및 거리에 기초하는 대체 픽셀들을 생성하는 것을 포함할 수 있다.

[0061] 타겟 머리와 소스 머리가 상이한 각도들로 있기 때문에, 대상체의 목 영역은 두 개의 이미지들에서 상이하게 위치될 수 있다. 머리 편집기(204)는 목과 어깨 영역의 보간된 영역을 생성함으로써 타겟 몸통과 소스 머리 사이의 매끄러운 전이를 렌더링한다. 일부 실시예들에서, 머리 편집기(204)는 타겟 머리와 타겟 몸통 사이의 영역(예를 들어, 타겟 목 및 타겟 어깨로서 설명되는 영역)을 포함하는 보간된 영역을 생성하고, 타겟 목 및 타겟 어깨와 연관된 타겟 픽셀들을 보간된 영역 – 여기서 보간된 영역은 대응하는 목 및 어깨 영역들에 대한 타겟 픽셀들 및 소스 픽셀들의 보간임 – 으로 대체한다. 일부 실시예들에서, 머리 편집기(204)는 카메라(243)에 의해 촬영된 연사 이미지들로부터 생성된 이미지 프레임들의 세트와 같은, 다수의 이미지 프레임들을 사용하고, 다수의 이미지 프레임들로부터 보간된 영역을 생성한다.

[0062] 본 명세서에서 설명되는 일부 실시예들에 따른, 머리 편집기에 대한 예시적인 타겟 이미지(400)인, 도 4a를 참조한다. 우측에 있는 대상체(405)는 위쪽을 보고 있고 매끄러운 어깨를 갖는다.

[0063] 도 4b는 머리 편집기(204)에 의해 생성된 예시적인 합성 이미지(425)이다. 머리 편집기(204)는 우측에 있는 대상체(430)에 대해 타겟 머리를 소스 머리로 대체하였다. 합성 이미지(425)에서, 목 및 어깨 상단에 대한 보간된 영역(430)은, 타겟 이미지(400) 내의 대상체(405) 상의 어깨와 비교해 어깨가 더 주름진 직물을 포함하고 목이 상이한 방향으로 있는 결과이다.

[0064] 일부 실시예들에서, 머리 편집기(204)는, 타겟 이미지 및 하나 이상의 소스 이미지들을 입력으로서 수신하고 합성 이미지를 출력하는 기계 학습 모델을 포함한다. 트레이닝된 기계 학습 모델은 하나 이상의 모델 형태들 또는 구조들을 포함할 수 있다. 예를 들어, 모델 형태들 또는 구조들은 임의의 타입의 신경망, 이를테면 선형 네트워크, 복수의 계층들(예를 들어, 각 계층이 선형 네트워크인 입력 계층과 출력 계층 사이의 "은닉 계층들")을 구현하는 심층 학습 신경망, 컨볼루션 신경망(예를 들어, 입력 데이터를 다수의 부분들 또는 타일들로 분해하거나 파티셔닝하고, 하나 이상의 신경망 계층들을 사용하여 각 타일을 별개로 처리하며, 각 타일의 처리로부터 결과들을 조합하는 네트워크), 시퀀스-시퀀스 신경망(예를 들어, 문장 내의 단어들, 비디오 내의 프레임들 등과 같은 순차적 데이터를 입력으로서 수신하고 결과 시퀀스를 출력으로서 생성하는 네트워크)을 포함할 수 있다.

[0065] 모델 형태 또는 구조는 다양한 노드들 간의 연결성 및 노드들의 계층들로의 조직화를 지정할 수 있다. 예를 들어, 제1 계층(예를 들어, 입력 계층)의 노드들은 입력 데이터 또는 애플리케이션 데이터로서 데이터를 수신할 수 있다. 이러한 데이터는 예를 들어, 트레이닝된 모델이 예를 들어, 타겟 이미지 및 하나 이상의 소스 이미지들의 분석을 위해 사용될 때, 노드당 하나 이상의 픽셀들을 포함할 수 있다. 후속 중간 계층들은 모델 형태 또는 구조로 지정된 연결성에 따라 이전 계층의 노드들의 출력을 입력으로서 수신할 수 있다. 이러한 계층들은 또한, 은닉 계층들로도 지칭될 수 있다. 최종 계층(예를 들어, 출력 계층)은 기계 학습 모델의 출력을 생성한다. 예를 들어, 출력 계층은 합성 이미지를 출력할 수 있다. 일부 실시예들에서, 모델 형태 또는 구조는 또한, 각 계층에서의 노드의 개수 및/또는 타입을 지정한다.

[0066] 상이한 실시예들에서, 트레이닝된 모델은 하나 이상의 모델들을 포함할 수 있다. 모델들 중 하나 이상은, 모델 구조 또는 형태에 따라 계층들로 배열되는, 복수의 노드들을 포함할 수 있다. 일부 실시예들에서, 노드들은, 예를 들어, 입력의 하나의 유닛을 프로세싱하여 출력의 하나의 유닛을 생성하도록 구성된, 메모리가 없는 컴퓨테이션 노드들일 수 있다. 노드에 의해 수행되는 컴퓨테이션은 예를 들어, 복수의 노드 입력들 각각에 가중치를 곱하는 것, 가중된 합을 획득하는 것, 및 가중된 합을 바이어스 또는 인터셉트 값으로 조정하여 노드 출력을 생성하는 것을 포함할 수 있다. 일부 실시예들에서, 노드에 의해 수행되는 컴퓨테이션은 또한, 조정된 가중 합에 계단/활성화 함수를 적용하는 것을 포함할 수 있다. 일부 실시예들에서, 계단/활성화 함수는 비선형 함수일 수 있다. 다양한 실시예들에서, 이러한 컴퓨테이션은 행렬 곱셈과 같은 연산들을 포함할 수 있다. 일부 실시예들에

서, 복수의 노드들에 의한 컴퓨테이션은 예를 들어, 멀티코어 프로세서의 다수의 프로세서 코어들을 사용하여, 그래픽 프로세싱 유닛(GPU), 또는 특수 목적 뉴럴 회로부의 개별 프로세싱 유닛들을 사용하여 병렬로 수행될 수 있다. 일부 실시예들에서, 노드들은 메모리를 포함할 수 있으며, 예를 들어, 후속 입력을 프로세싱 시 하나 이상의 이전 입력들을 저장하고 사용할 수 있다. 예를 들어, 메모리를 갖는 노드들은 장단기 메모리(long short-term memory, LSTM) 노드들을 포함할 수 있다. LSTM 노드들은 노드가 유한 상태 기계(finite state machine, FSM)처럼 작용할 수 있게 하는 "상태"를 유지하기 위해 메모리를 사용할 수 있다.

[0067] 일부 실시예들에서, 트레이닝된 모델은 개별 노드에 대한 가중치들 또는 임베딩들을 포함할 수 있다. 예를 들어, 모델은 모델 형태 또는 구조에 의해 지정된 바와 같이 계층들로 조직화된 복수의 노드들로서 개시될 수 있다. 초기화 시, 모델 형태에 따라 연결된 노드들의 각 쌍, 예를 들어, 신경망의 연속적인 계층들에서의 노드들 간의 연결에 개별 가중치가 적용될 수 있다. 예를 들어, 개별 가중치들은 랜덤하게 부여될 수 있거나, 디폴트 값들로 초기화될 수 있다. 그런 다음, 모델은 결과를 생성하기 위해 예를 들어, 트레이닝 데이터를 사용하여, 트레이닝될 수 있다.

[0068] 트레이닝은 지도 학습(supervised learning) 기법들을 적용하는 것을 포함할 수 있다. 지도 학습에서, 트레이닝 데이터는 복수의 입력들(예를 들어, 타겟 이미지들 및 소스 이미지들, 세그먼트화 마스크들 등) 및 각 입력에 대한 대응하는 정답 출력(예를 들어, 각 이미지, 합성 이미지 등에서 대상체의 얼굴과 같은 대상체의 일부분을 정확하게 식별하는 정답 마스크)을 포함할 수 있다. 모델의 출력과 정답 출력의 비교에 기초하여, 예를 들어, 모델이 합성 이미지에 대한 정답 출력을 생성할 확률을 증가시키는 방식으로, 가중치들의 값들이 자동으로 조정된다.

[0069] 다양한 실시예들에서, 트레이닝된 모델은 모델 구조에 대응하는 가중치들 또는 임베딩들의 세트를 포함한다. 일부 실시예들에서, 트레이닝된 모델은 고정된, 예를 들어, 가중치들을 제공하는 서버로부터 다운로드되는, 가중치들의 세트를 포함할 수 있다. 다양한 실시예들에서, 트레이닝된 모델은 모델 구조에 대응하는 가중치들 또는 임베딩들의 세트를 포함한다. 데이터가 생략되는 실시예들에서, 머리 편집기(204)는 예를 들어, 머리 편집기(204)의 개발자, 서드 파티 등에 의해, 사전 트레이닝에 기초한 트레이닝된 모델을 생성할 수 있다. 일부 실시예들에서, 트레이닝된 모델은 고정된, 예를 들어, 가중치들을 제공하는 서버로부터 다운로드되는, 가중치들의 세트를 포함할 수 있다.

[0070] 일부 실시예들에서, 트레이닝된 기계 학습 모델은 하나 이상의 대상체들을 갖는 하나 이상의 소스 이미지들 및 타겟 이미지를 수신한다. 기계 학습 모델은, 하나 이상의 소스 이미지들 내의 모발을 포함하는 하나 이상의 머리들에 대응하는 하나 이상의 소스 이미지들 내의 픽셀들을 식별하는 하나 이상의 세그먼트화 마스크들을 생성할 수 있다. 각 대상체에 대해, 기계 학습 모델은 타겟 이미지로부터의 머리 픽셀들을, 소스 이미지에 따르는 대응하는 위치, 스케일, 및 각도를 갖는 하나 이상의 소스 이미지들로부터의 머리 픽셀들로 대체한다. 기계 학습 모델은 머리의 가장자리를 따르는 머리 픽셀들을 블렌딩한다. 일부 실시예들에서, 소스 머리와 타겟 머리 사이의 차이들의 결과로서 드러났던 배경 픽셀들이 인페인팅된다. 일부 실시예들에서, 기계 학습 모델은 보간된 영역을, 대상체의 목 및 어깨에 대응하는 타겟 이미지로 블렌딩한다. 일부 실시예들에서, 트레이닝된 기계 학습 모델은 이러한 변화들을 통합하는 합성 이미지를 출력한다.

[0071] 일부 실시예들에서, 기계 학습 모델은 트레이닝된 기계 학습 모델에 의해 출력된 각 합성 이미지에 대한 신뢰도 값을 출력한다. 신뢰도 값은 퍼센티지, 0 내지 1의 수 등으로서 표현될 수 있다. 예를 들어, 기계 학습 모델은, 합성 이미지가 타겟 머리를 소스 머리로 정확하게 대체하였고 다른 사람 또는 객체로부터의 픽셀들을 포함하지 않는다는 신뢰도에 대한 85%의 신뢰도 값을 출력한다. 일부 실시예들에서, 신뢰도 값이 임계 신뢰도 값을 초과하는 경우 합성 이미지가 사용자에게 제공된다. 일부 실시예들에서, 신뢰도 값은 미리 컴퓨트되고, 신뢰도 값이 임계 신뢰도 값을 초과하지 않는 한 합성 이미지가 생성되지 않는다.

[0072] 일부 실시예들에서, 머리 편집기(204)는, 합성 이미지를 생성하기 위해 사용되는 단계들에서의 상이한 기능들을 수행하는 다수의 기계 학습 모델들을 포함한다. 예를 들어, 제1 기계 학습 모델은 타겟 머리와 연관된 타겟 픽셀들을 소스 머리로부터의 소스 픽셀들로 대체할 수 있고, 제2 기계 학습 모델은 소스 머리를 타겟 머리와 정렬시키며, 제3 기계 학습 모델은 타겟 목 및 타겟 어깨와 연관된 타겟 픽셀들을 대체하는 보간된 영역을 생성한다.

[0073] 도 5는 예시적인 타겟 이미지(500)이다. 타겟 이미지(500) 내의 소년의 머리(505)와 소녀의 머리(510)가 소스 이미지로부터의 머리들로 대체될 것이다.

[0074] 도 6a는 도 5의 타겟 이미지(500)로부터 분리된 예시적인 제1 타겟 머리(600)이다. 도 6b는 도 5의 타겟 이미지(500)로부터 분리된 예시적인 제2 타겟 머리(650)이다. 제1 타겟 머리(600)는 소년의 코가 위로 향하고 그의 입이 벌려 있기 때문에 대체를 위해 분리 및 선택될 수 있다. 제2 타겟 머리(650)는 소녀가 카메라 정면을 보고 있지 않기 때문에 대체를 위해 분리 및 선택될 수 있다.

[0075] 도 7a는 소스 이미지로부터 분리된 예시적인 제1 소스 머리(700)이다. 도 7b는 제1 소스 머리(700)를 포함하는 동일한 소스 이미지로부터 또는 상이한 소스 이미지로부터 분리된 예시적인 제2 소스 머리(750)이다. 머리 편집기(204)는 제1 타겟 머리(600)와 연관된 타겟 픽셀들의 적어도 일부분 및 제2 타겟 머리(650)와 연관된 타겟 픽셀들의 적어도 일부분을 각각, 제1 소스 머리(700) 및 제2 소스 머리(750)로부터의 소스 픽셀들로 대체한다.

[0076] 도 8a는 제1 타겟 머리와 정렬되는 예시적인 제1 소스 머리(800)이다. 도 8b는 제2 타겟 머리와 정렬되는 예시적인 제2 소스 머리(850)이다. 이미지들이 정렬되어 있지만, 소년의 캡(805)의 위치와 같은, 이미지들이 오버랩되지 않는 몇몇 곳들이 있다.

[0077] 도 9는 제1 타겟 머리에 의해 이전에 커버되었던 배경의 영역들을 노출시키는 제1 소스 머리의 정렬에 기초하여 인페인팅을 위해 분석되는 예시적인 합성 이미지(900)이다. 머리 편집기(204)가 타겟 머리들의 인페인팅을 수행한 후, 합성 이미지는 소스 머리들이 가시적인 끊김 또는 다른 결함 없이 합성 이미지와 맞는 외관을 갖는다.

[0078] 도 10은 소스 머리들이 타겟 머리들을 대체한 예시적인 합성 이미지(1000)이다. 도 5에서의 소년의 머리(505)는 도 10에서의 소년의 머리(1005)로 대체되고, 도 5에서의 소녀의 머리(510)는 도 10에서의 소녀의 머리(1010)로 대체된다.

[0079] 얼굴 편집기(206)는 얼굴 피처들(예를 들어, 미소, 뜬 눈, 입 형상들 등)을 소스 이미지로부터 타겟 이미지로 전달한다. 얼굴 편집기(206)는 타겟 얼굴의 포즈를 변경하지 않는다. 일부 실시예들에서, 얼굴 편집기(206)는 소스 이미지 내의 소스 얼굴 피처들로부터의 소스 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들과 연관된 타겟 픽셀 값들의 적어도 일부분을 조정한다.

[0080] 일부 실시예들에서, 얼굴 편집기(206)는 특정 사용자 권한으로 구현되는, 얼굴 매칭 알고리즘을 사용하여 동일한 대상체의 타겟 얼굴 및 소스 얼굴을 추출한다. 얼굴 편집기(206)는 초기 포즈로부터의 타겟 얼굴을 기준 포즈에 정렬시키며, 여기서 기준 포즈는 기계 학습 모델이 사용하도록 트레이닝되는 포즈들 중 하나이다.

[0081] 일부 실시예들에서, 얼굴 편집기(206)는 머리 편집기(204)를 참조하여 위에서 설명된 모델 타입들 중 하나로부터 선택되는 기계 학습 모델을 사용한다. 예를 들어, 기계 학습 모델은 인코더 및 컨볼루션 신경망(convolutional neural network, CNN)을 포함할 수 있다.

[0082] 일부 실시예들에서, 얼굴의 일부 컴포넌트들이 각도 및 포즈의 작은 변화들에 더 취약할 수 있기 때문에, 타겟 얼굴 전부가 소스로 대체되는 것은 아니다. 예를 들어, 사람의 코는 위로 젖혀진 머리로부터의 코가 정면을 보고 있는 얼굴에 추가되는 경우 어울리지 않아 보일 수 있다. 그 결과, 기계 학습 모델의 인코더는 소스 머리의 컴포넌트들을 소스 벡터로서 인코딩하기 위해 얼굴 피처 마스크(예를 들어, 눈/입 마스크)를 사용한다. 인코더는 또한, 타겟 머리를 타겟 벡터로서 인코딩한다. 기계 학습 모델은 대응하는 컴포넌트들을 소스 벡터로부터 타겟 벡터로 복사한다.

[0083] 일부 실시예들에서, CNN은 소스 벡터 컴포넌트들을 포함하는 타겟 벡터를 렌더링함으로써 합성 이미지의 버전을 각각 생성하는 다수의 계층들을 포함한다. CNN은 각 계층에서 점점 더 높은 해상도 레벨을 갖는 합성 이미지를 생성한다. 고해상도 출력들은 최종 합성 이미지가 출력될 때까지 기존의 더 낮은 해상도 계층들과 혼합된다. 일부 실시예들에서, 기계 학습 모델은 렌더링된 타겟 머리를 다시 초기 포즈로 재정렬하고, 재정렬된 타겟 머리를 합성 이미지로서 출력되는 소스 이미지와 블렌딩한다.

[0084] 도 11은 합성 이미지를 생성하는 온-디바이스 기계 학습 모델의 예시적인 블록도(1100)이다. 소스 이미지(1105) 및 타겟 이미지(1110)는, 도 1에서 예시된 사용자 디바이스(115) 상에 저장된 미디어 애플리케이션(103)과 같은, 오프라인 모듈(1115)과 온-디바이스 모듈(1120) 둘 다에 의해 수신된다. 온-디바이스 모듈(1120)은 예측 이미지(1130)를 출력하여서, 소스 이미지(1105)로부터의 얼굴 표정 및/또는 머리 포즈가 타겟 이미지(1110)로 융합된다. 온-디바이스 모듈(1120)은 레이턴시 및 메모리 사용량의 버짓들 내의 사용자 디바이스들, 예를 들어, 기계 학습 회로부를 갖거나 갖지 않는 스마트폰들 상에서 실행될 수 있도록 구현된다.

[0085] 오프라인 모듈(1115)은 온-디바이스 모델 트레이닝에 대한 감독을 제공할 수 있으며, 이는 바람직하게는, 경량 온-디바이스 모듈(1120)이 합성 이미지(1125)를 생성할 수 있게 한다. 오프라인 모듈(1115)이 레이턴시 및 메모

리 사용량에 대해 더 큰 버짓을 갖기 때문에 합성 이미지(1125)는 예측 이미지(1130)보다 더 높은 품질 버전일 수 있다.

[0086] 일부 실시예들에서, 온-디바이스 모듈(1120)은 휨 및 복원을 사용하고 – 여기서 조밀한 얼굴 메쉬가 소스 이미지(1105)를 타겟 위치로 휘어지게 함 –, 아티팩트를 제거하기 위한 신경망(또는 다른 적합한 기법)을 사용할 수 있다. 두 개의 얼굴 이미지들 및 이들의 얼굴 메쉬들이 주어지면, 온-디바이스 모듈(1120)은 소스 얼굴의 포즈를 타겟 얼굴 메쉬에 정렬시키고, 그런 다음 소스 얼굴을 타겟 이미지로 휘어지게 한다. 휘어진 얼굴은 소스 표정을 갖지만, 정렬 및 휨에 의해 야기되는 아티팩트를 가지며, 이는 저하로서 고려될 수 있다. 온-디바이스 모듈(1120)은 저하를 제거 또는 교정할 수 있는 오프라인 모듈(1115)로부터의 감독을 갖는 복원 모델을 포함할 수 있다.

[0087] 일부 실시예들에서, 온-디바이스 모듈(1120)은 지정된 표정을 반영하도록 얼굴들을 수정할 수 있다. 온-디바이스 모듈(1120)은 두 개의 인코더들, 즉, 소스 이미지로부터 표정 코드를 추출하기 위한 소스 인코더 및 타겟 이미지로부터 머리 포즈 및 외관 정보를 추출하기 위한 타겟 인코더를 포함할 수 있다. 온-디바이스 모듈(1120)은 디코더의 하나 이상의 레벨들에서(예를 들어, 모든 레벨에서) 표정 코드를 삽입할 수 있다. 온-디바이스 모듈(1120)이 이미지의 다른 부분들이 아닌 얼굴을 편집하는 것을 보장하기 위해 타겟 인코더로부터의 스킵 연결들이 또한 추가될 수 있다.

[0088] 도 12a는 제2 대상체(1205)가 얼굴 편집기(206)에 의해 대체되는 표정을 갖는 예시적인 타겟 이미지(1200)이다.

[0089] 도 12b는 위에서 설명된 기법들에 따라 얼굴 편집기(206)에 의해 생성된 예시적인 합성 이미지(1250)이다. 볼 수 있는 바와 같이, 타겟 이미지(1200) 내의 우측에 있는 대상체(1205)의 입을 벌려 이빨을 드러낸 미소가 합성 이미지(1250) 내의 대상체(1255)의 입을 닫은 웃음으로 대체된다.

[0090] 도 13은 본 명세서에서 설명되는 일부 실시예들에 따른, 베스트 테이크를 선택하기 위한 예시적인 사용자 인터페이스들(1300, 1350, 1375)을 예시한다. 제1 사용자 인터페이스(1300)는 두 대상체들(1306, 1307)의 타겟 이미지(1305)를 포함한다. 사용자는 합성 이미지를 획득하기 위한 프로세스를 시작하기 위해 "베스트 테이크" 버튼(1310)을 선택할 수 있다.

[0091] 제2 사용자 인터페이스(1350)는 제1 대상체 및 제2 대상체의 타겟 이미지(1355) 및 각 아이콘들(1357, 1359)을 포함한다. 사용자는 대상체에 대한 상이한 얼굴을 선택하기 위해 아이콘들(1357, 1359) 중 하나를 선택할 수 있다. 이러한 예에서, 사용자는 제2 대상체의 아이콘(1359)을 선택하였다. 사용자 인터페이스(1350)는 제2 대상체에 대한 세 개의 옵션들(1361, 1363, 및 1365)을 포함한다. 제1 옵션(1361) 및 제3 옵션(1365)은 소스 이미지들로부터 온 것이고, 제2 옵션(1363)은 타겟 이미지(1355)로부터 온 것이다. 사용자가 세 개의 옵션들(1361, 1363, 및 1365) 중 하나를 선택하면, 사용자는 합성 이미지를 보기 위해 완료 버튼(1367) 타겟 이미지(1355) 상의 선택된 옵션을 선택하거나, 프로세스를 재시작하기 위해 리셋 버튼(1369)을 선택할 수 있다. 이러한 예에서, 사용자는 제1 옵션(1361)을 선택하고 완료 버튼(1367)을 선택한다.

[0092] 제3 사용자 인터페이스(1375)는, 제2 사용자 인터페이스(1350)의 타겟 이미지(1355) 및 소스 얼굴(1379)으로부터 생성된 합성 이미지(1377)를 포함한다. 사용자는 이미지의 복사본을 저장하기 위해 복사본 저장 버튼(1385)을 누를 수 있다.

[0093] **예시적인 흐름도**

[0094] 도 14는 본 명세서에서 설명되는 일부 실시예들에 따른, 이미지를 세그먼트화하기 위한 예시적인 방법(1400)의 흐름도를 예시한다. 방법(1400)은 도 2에서의 컴퓨팅 디바이스(200)에 의해 수행될 수 있다. 일부 실시예들에서, 방법(1400)은 사용자 디바이스(115)에 의해, 미디어 서버(101)에 의해, 또는 부분적으로 사용자 디바이스(115) 상에서 그리고 부분적으로 미디어 서버(101) 상에서 수행된다.

[0095] 도 14의 방법(1400)은 블록(1402)에서 시작할 수 있다. 블록(1402)에서, 소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트가 수신되며, 소스 이미지 및 타겟 이미지는 적어도 대상체를 포함한다. 일부 실시예들에서, 블록(1402) 전에, 방법(1400)은, 카메라를 이용해 이미지들의 세트를 촬영하는 단계, 타겟 이미지, 및 소스 이미지들의 세트 – 소스 이미지들의 세트는 소스 이미지를 포함함 – 로부터 소스 머리를 선택하기 위한 옵션을 포함하는, 사용자 인터페이스를 사용자에게 제공하는 단계, 및 사용자로부터 소스 이미지의 선택을 수신하는 단계를 더 포함한다. 대상체들은 사람 또는 동물을 포함할 수 있다. 블록(1402)에는 블록(1404)이 뒤따를 수 있다.

[0096] 블록(1404)에서, 소스 이미지 및 타겟 이미지를 수정하기 위한 권한이 획득되는지 여부가 결정된다. 권한이 획득되지 않은 경우, 블록(1404)에, 방법(1400)이 종료되는 블록(1406)이 뒤따를 수 있다. 권한이 획득되는 경우, 블록(1404)에 블록(1408)이 뒤따를 수 있다.

[0097] 블록(1408)에서, 이미지들의 세트에 기초하여, 머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부가 결정된다. 얼굴 편집기를 사용하기로 결정하는 것은 타겟 머리의 제1 각도와 소스 머리의 제2 각도 사이의 각도 차이; 타겟 머리 또는 타겟 얼굴을 둘러싸는 경계 박스, 및 경계 박스와 타겟 이미지 내의 하나 이상의 다른 대상체들과 연관된 경계 박스들 사이의 거리; 타겟 머리의 가림 또는 소스 머리의 가림에 기초할 수 있다. 타겟 머리의 가림 또는 소스 머리의 가림을 결정하는 것은 컬러 히스토그램들(color histograms)에 있어서의 차이를 결정하는 것에 기초할 수 있다.

[0098] 머리 편집기가 선택되는 경우, 블록(1408)에 블록(1410)이 뒤따를 수 있다. 블록(1410)에서, 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을, 소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체하고, 타겟 머리와 타겟 몸통 사이의 영역을 포함하는 타겟 목과 연관된 목 픽셀들 및 타겟 어깨와 연관된 어깨 픽셀들을, 소스 이미지와 타겟 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함으로써, 합성 이미지가 생성된다. 일부 실시예들에서, 머리 편집기는 타겟 이미지 내의 타겟 머리와 연관된 타겟 픽셀들의 적어도 일부분을 대체하기 전에 타겟 머리의 각도를 소스 머리의 각도와 정렬한다. 얼굴 편집기가 선택되는 경우, 블록(1408)에 블록(1412)이 뒤따를 수 있다.

[0099] 블록(1412)에서, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정함으로써, 합성 이미지가 생성된다. 일부 실시예들에서, 소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴 피처들의 적어도 일부분을 조정하는 것은, 초기 포즈에서의 타겟 머리, 및 소스 머리를 추출하는 것, 타겟 머리를 기준 포즈(canonical pose)에 정렬하는 것, 잠재 공간(latent space)에서, 정렬된 타겟 머리를 타겟 벡터로서 인코딩하고 소스 머리를 소스 벡터로서 인코딩하는 것, 소스 벡터로부터의 하나 이상의 컴포넌트들을 타겟 벡터로 복사하는 것, 인코딩된 소스 머리로부터 하나 이상의 컴포넌트들을 포함하는 수정된 타겟 벡터를 렌더링하는 것, 렌더링된 타겟 머리를 초기 포즈에 재정렬하는 것, 및 재정렬된 타겟 머리를 소스 이미지와 블렌딩(blending)하는 것을 포함한다.

[0100] 상기한 설명에서, 설명을 목적으로, 본 명세서에 대한 완전한 이해를 제공하기 위해 많은 특정 세부사항들이 제시된다. 그러나, 본 개시는 이러한 특정 세부사항들 없이 실시될 수 있다는 것이 당업자에게 명백할 것이다. 일부 경우들에서, 구조들 및 디바이스들은 설명을 모호하게 하는 것을 피하기 위해 블록도 형태로 도시된다. 예를 들어, 실시예들은 주로 사용자 인터페이스들 및 특정 하드웨어를 참조하여 위에서 설명될 수 있다. 그러나, 실시예들은 데이터 및 커맨드들을 수신할 수 있는 임의의 타입의 컴퓨팅 디바이스, 및 서비스들을 제공하는 임의의 주변 디바이스들에 적용될 수 있다.

[0101] 본 명세서에서 "일부 실시예들" 또는 "일부 경우들"에 대한 언급은 실시예들 또는 경우들과 관련하여 설명되는 특정 피처, 구조, 또는 특성이 설명의 적어도 하나의 구현예에 포함될 수 있음을 의미한다. 본 명세서의 다양한 위치들에서의 "일부 실시예들에서"라는 문구의 출현들은 반드시 모두 동일한 실시예들을 지칭하는 것은 아니다.

[0102] 상기한 상세한 설명의 일부 부분들은 컴퓨터 메모리 내의 데이터 비트들에 대한 동작들의 알고리즘 및 부호식 표현의 관점에서 제시된다. 이러한 알고리즘 설명 및 표현은 데이터 프로세싱 분야의 숙련자가 자신의 작업의 내용을 당업자에게 가장 효과적으로 전달하기 위해 사용하는 수단이다. 여기서는 알고리즘이 있고, 일반적으로, 원하는 결과를 초래하는 단계들의 자기 일관적인 시퀀스인 것으로 구상된다. 단계들은 물리적 양들의 물리적 조작을 필요로 하는 것들이다. 일반적으로, 반드시 그런 것은 아니지만, 이러한 양들은 저장, 전송, 조합, 비교, 및 다른 방식으로 조작될 수 있는 전기적 또는 자기 데이터의 형태를 취한다. 이러한 데이터를 비트, 값, 엘리먼트, 부호, 문자, 용어, 숫자 등으로서 지칭하는 것은 주로 통상적인 사용상의 이유로 때때로 편리한 것으로 입증되었다.

[0103] 그러나, 이러한 용어들 및 유사한 용어들 모두는 적절한 물리적 양들과 연관되어야 하고, 단지 이러한 양들에 적용되는 편리한 라벨들일 뿐이라는 것을 유념해야 한다. 다음의 논의로부터 명백하게 달리 명시되지 않는 한, 본 설명 전반에 걸쳐, "프로세싱" 또는 "컴퓨팅" 또는 "계산" 또는 "결정" 또는 "디스플레이" 등을 포함하는 용어들을 이용하는 논의는, 컴퓨터 시스템의 레지스터들 및 메모리들 내의 물리적(전자) 양들로서 표현되는 데이터를 컴퓨터 시스템 메모리들 또는 레지스터들 또는 다른 이러한 정보 저장, 송신, 또는 디스플레이 디바이스들 내의 물리적 양들로서 유사하게 표현되는 다른 데이터로 조작 및 변환하는 컴퓨터 시스템, 또는 유사한 전자 컴

퓨팅 디바이스의 액션 및 프로세스들을 지칭하는 것으로 이해된다.

[0104] 본 명세서의 실시예들은 또한, 위에서 설명된 방법들의 하나 이상의 단계들을 수행하기 위한 프로세서에 관한 것일 수 있다. 프로세서는 컴퓨터에 저장된 컴퓨터 프로그램에 의해 선택적으로 활성화되거나 재구성된 특수 목적 프로세서일 수 있다. 이러한 컴퓨터 프로그램은, 각각 컴퓨터 시스템 버스에 결합되는 광학 디스크, ROM, CD-ROM, 자기 디스크, RAM, EPROM, EEPROM, 자기 또는 광학 카드, 비휘발성 메모리를 갖는 USB 키를 포함하는 플래시 메모리, 또는 전자 명령어를 저장하기에 적합한 임의의 타입의 매체를 포함하지만 이에 제한되지 않는 비일시적 컴퓨터 판독가능 저장 매체에 저장될 수 있다.

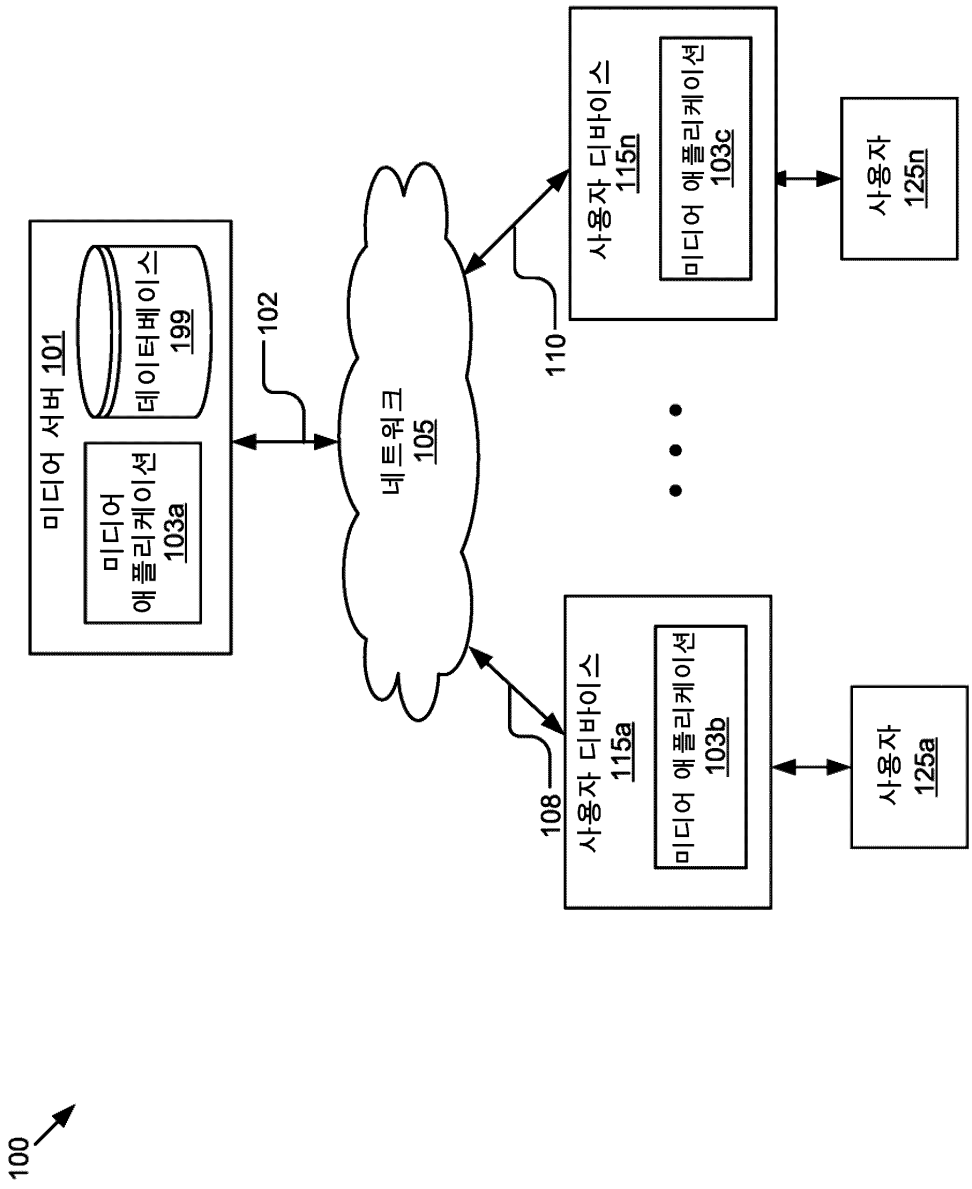
[0105] 본 명세서는 일부 완전히 하드웨어 실시예들, 일부 완전히 소프트웨어 실시예들, 또는 하드웨어와 소프트웨어 엘리먼트들 둘 모두를 포함하는 일부 실시예들의 형태를 취할 수 있다. 일부 실시예들에서, 본 명세서는 펌웨어, 상주 소프트웨어, 마이크로코드 등을 포함하지만 이에 제한되지 않는 소프트웨어로 구현된다.

[0106] 또한, 본 설명은 컴퓨터 또는 임의의 명령어 실행 시스템에 의해 또는 이와 관련하여 사용하기 위한 프로그램 코드를 제공하는 컴퓨터 사용가능 또는 컴퓨터 판독가능 매체로부터 액세스가능한 컴퓨터 프로그램 제품의 형태를 취할 수 있다. 본 설명의 목적들을 위해, 컴퓨터 사용가능 또는 컴퓨터 판독가능 매체는 명령어 실행 시스템, 장치, 또는 디바이스에 의해 또는 이와 관련하여 사용하기 위해 프로그램을 포함, 저장, 통신, 전파, 또는 전송할 수 있는 임의의 장치일 수 있다.

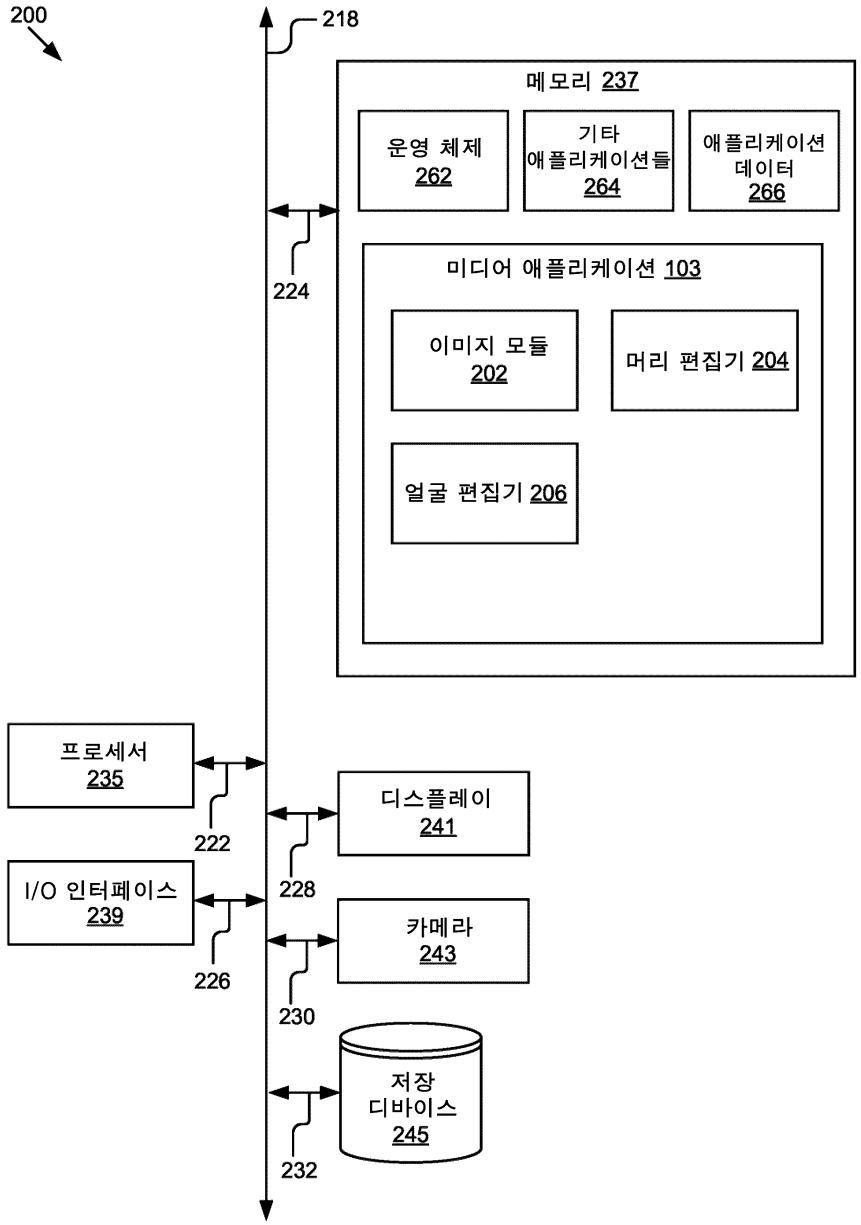
[0107] 프로그램 코드를 저장하거나 실행하기에 적합한 데이터 프로세싱 시스템은 시스템 버스를 통해 메모리 엘리먼트들에 직접적으로 또는 간접적으로 결합된 적어도 하나의 프로세서를 포함할 것이다. 메모리 엘리먼트들은 프로그램 코드의 실제 실행 동안 채용된 로컬 메모리, 벌크 스토리지, 및 실행 동안 벌크 스토리지로부터 코드가 검색되어야 하는 횟수를 감소시키기 위해 적어도 일부 프로그램 코드의 임시 스토리지를 제공하는 캐시 메모리들을 포함할 수 있다.

도면

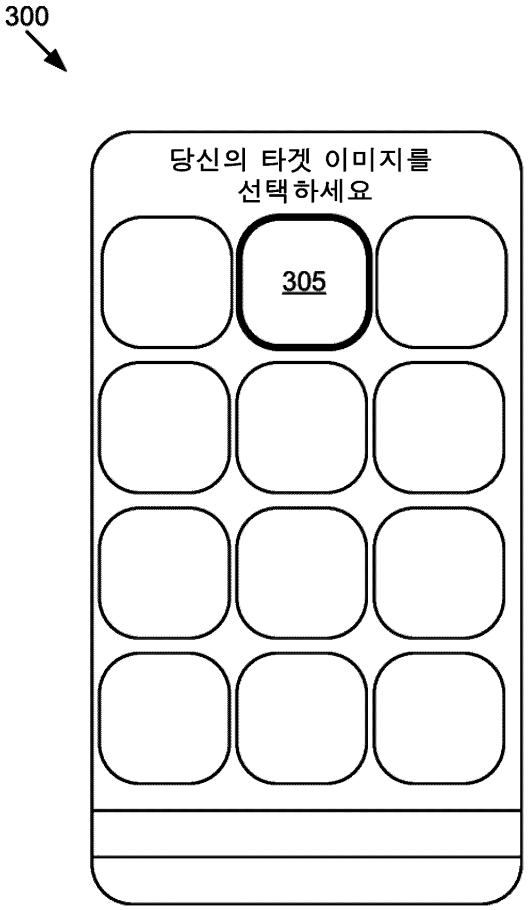
도면1


100
미디어 서버 101
미디어 애플리케이션 103a
데이터베이스 199
102
네트워크 105
110
사용자 디바이스 115n
미디어 애플리케이션 103c
사용자 125n
108
사용자 디바이스 115a
미디어 애플리케이션 103b
사용자 125a

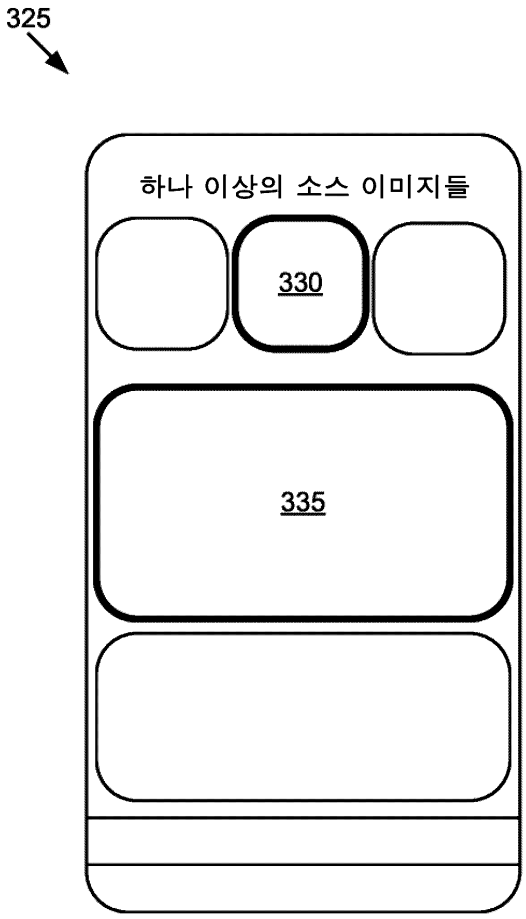
도면2


200
218
메모리 237
운영 체제 262
기타 애플리케이션들 264
애플리케이션 데이터 266
224
미디어 애플리케이션 103
이미지 모듈 202
머리 편집기 204
얼굴 편집기 206
프로세서 235
222
디스플레이 241
228
I/O 인터페이스 239
226
카메라 243
230
저장 디바이스 245
232

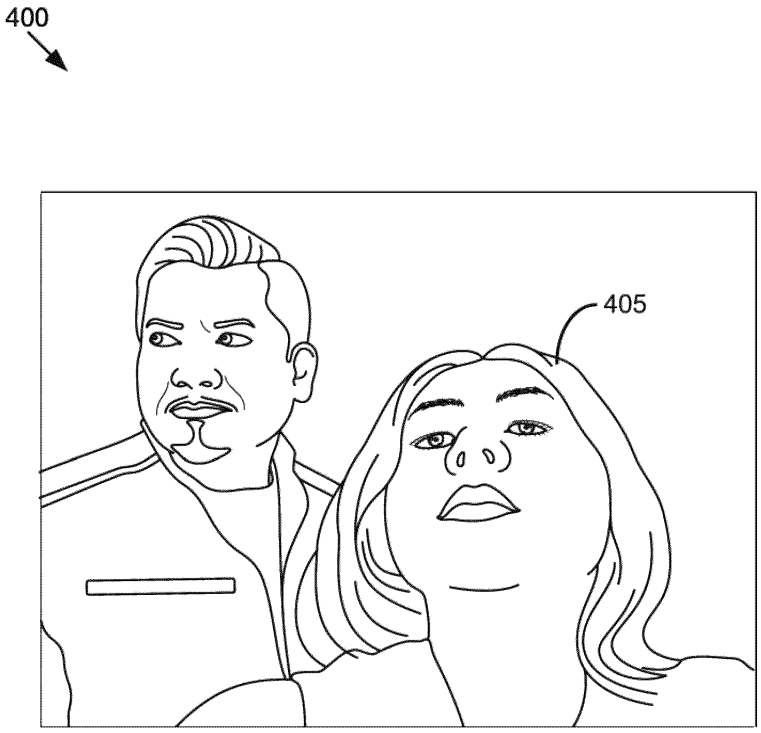
도면3a



도면3b



도면4a



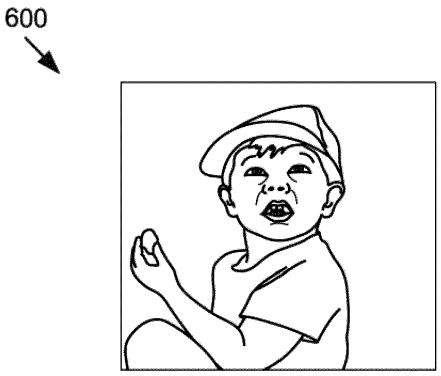
도면4b



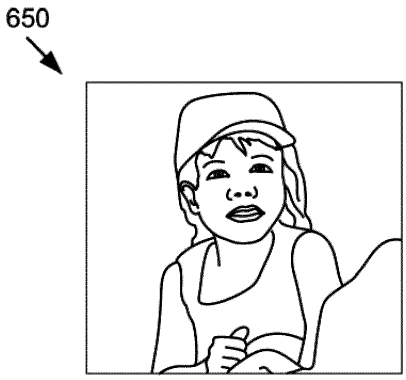
도면5



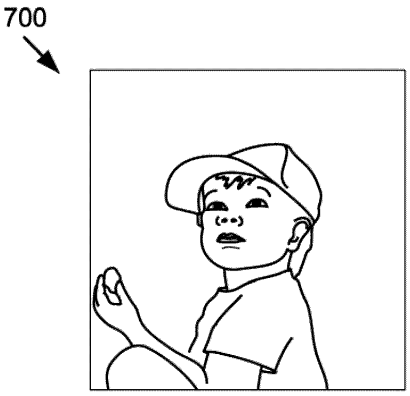
도면6a



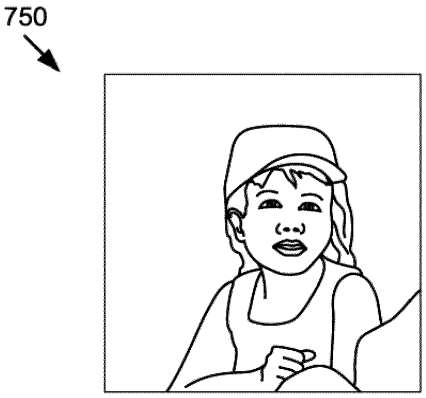
도면6b



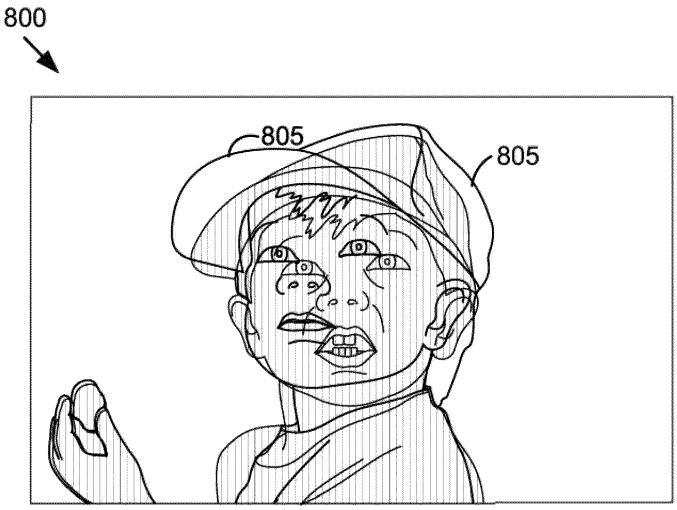
도면7a



도면7b



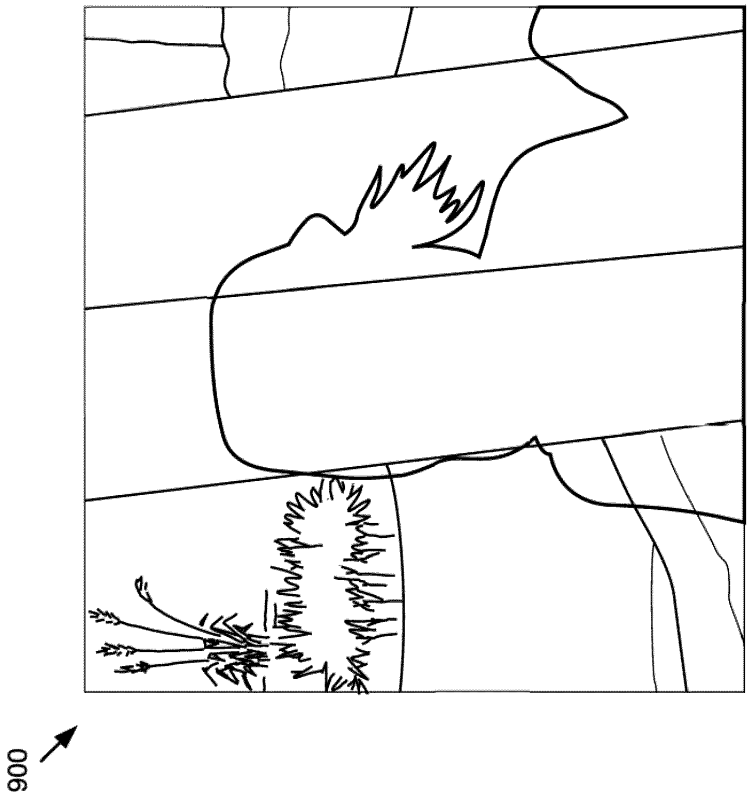
도면8a



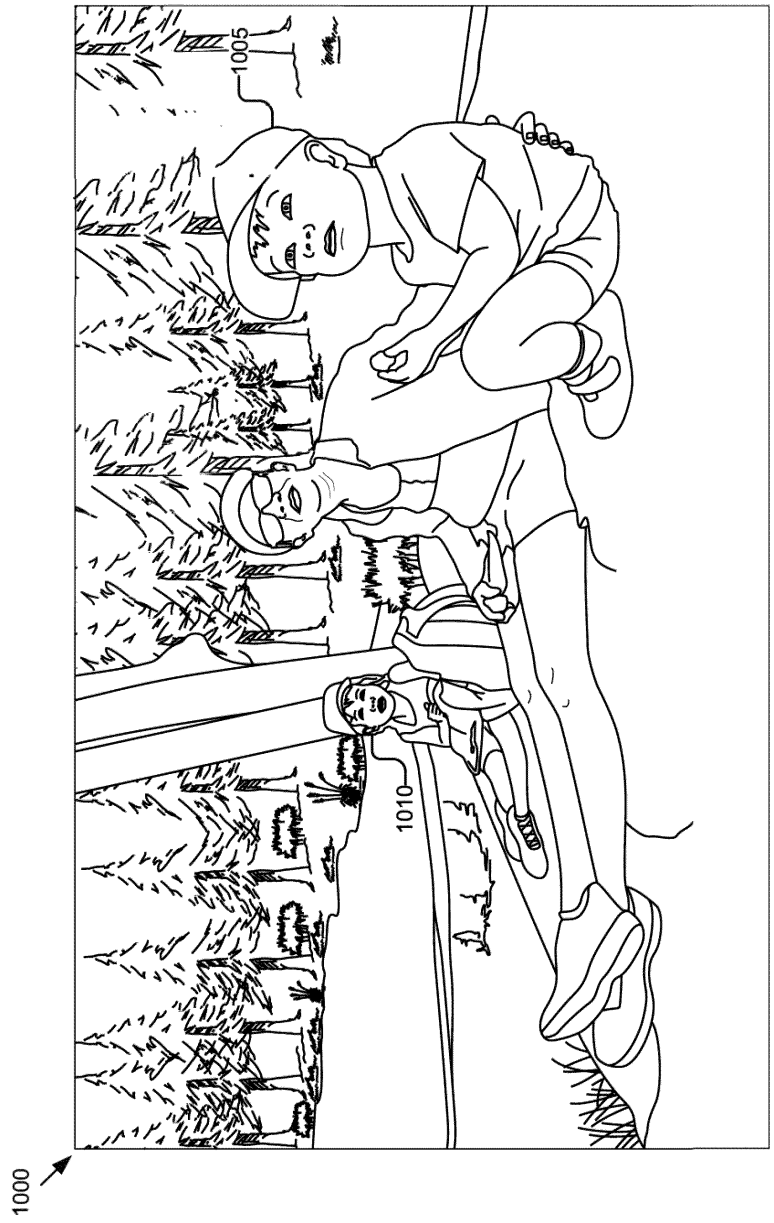
도면8b



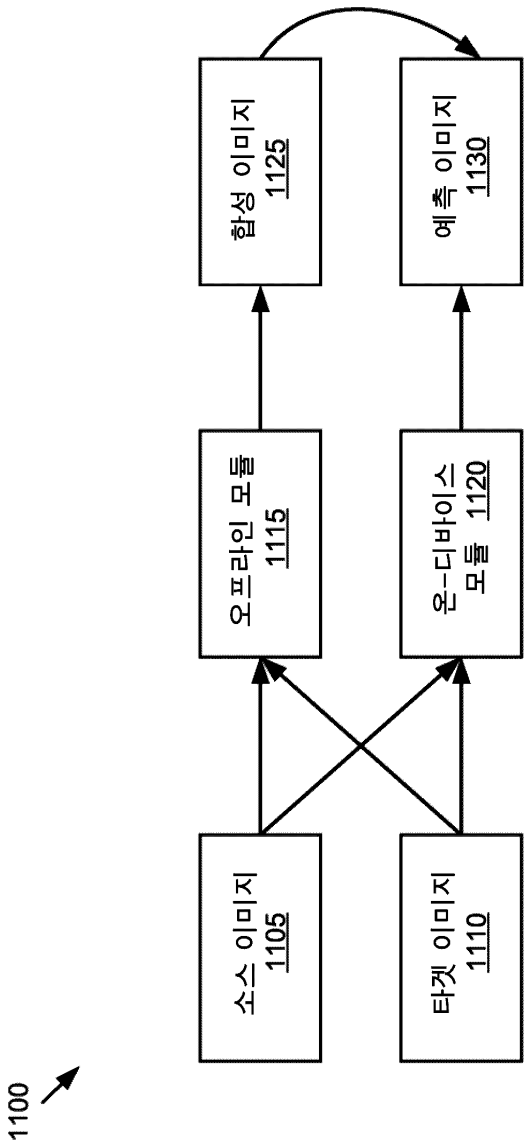
도면9



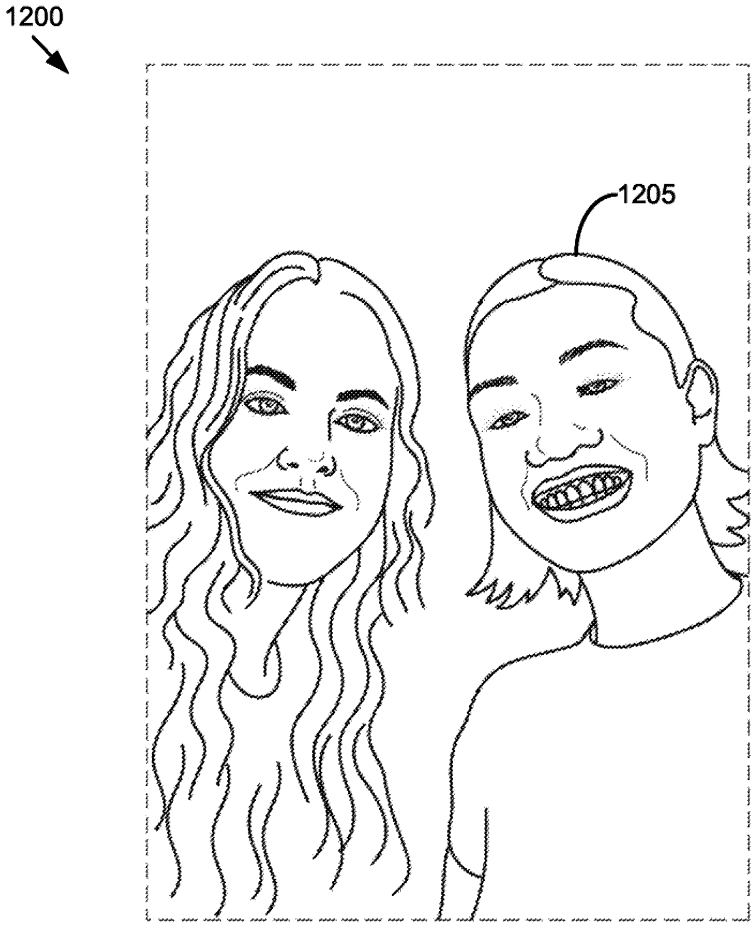
도면10



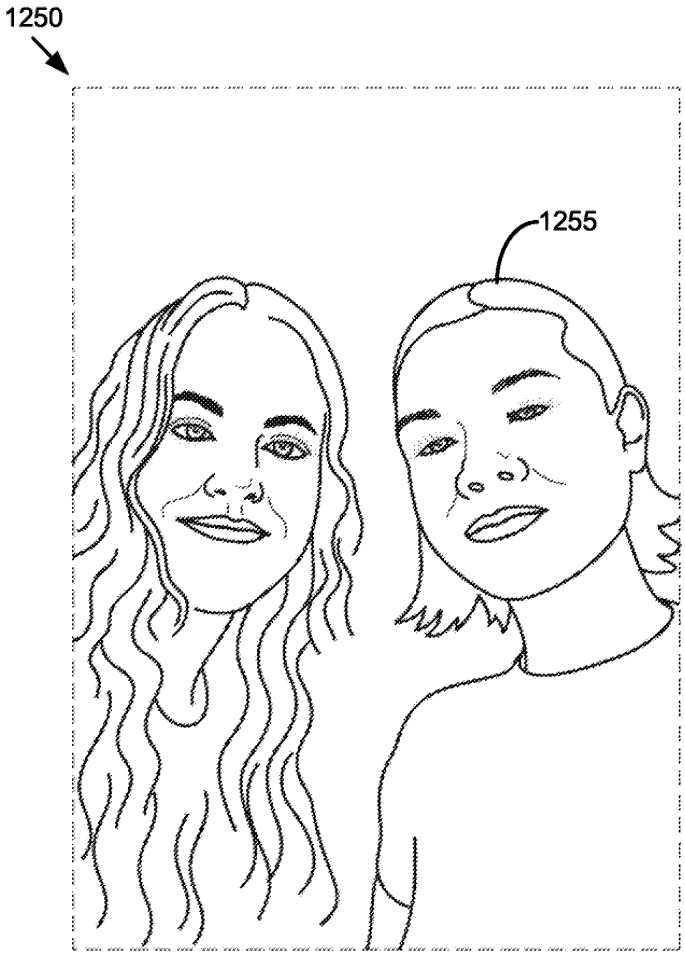
도면11


1100
소스 이미지
1105
타겟 이미지
1110
오프라인 모듈
1115
온-디바이스
모듈 1120
합성 이미지
1125
예측 이미지
1130

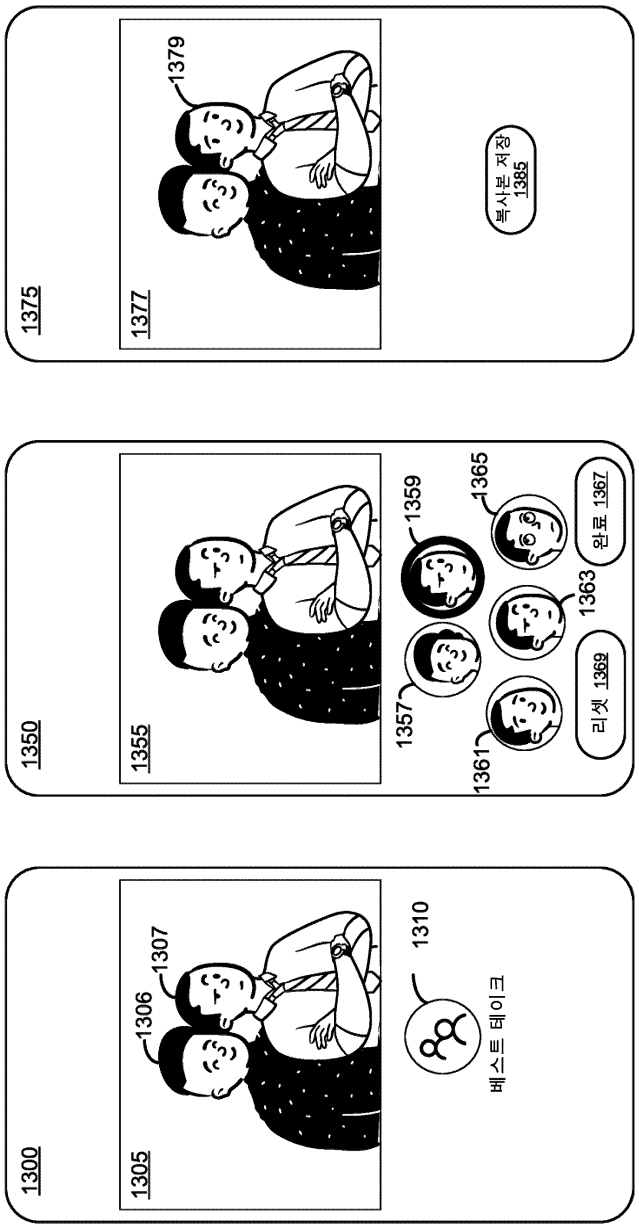
도면12a



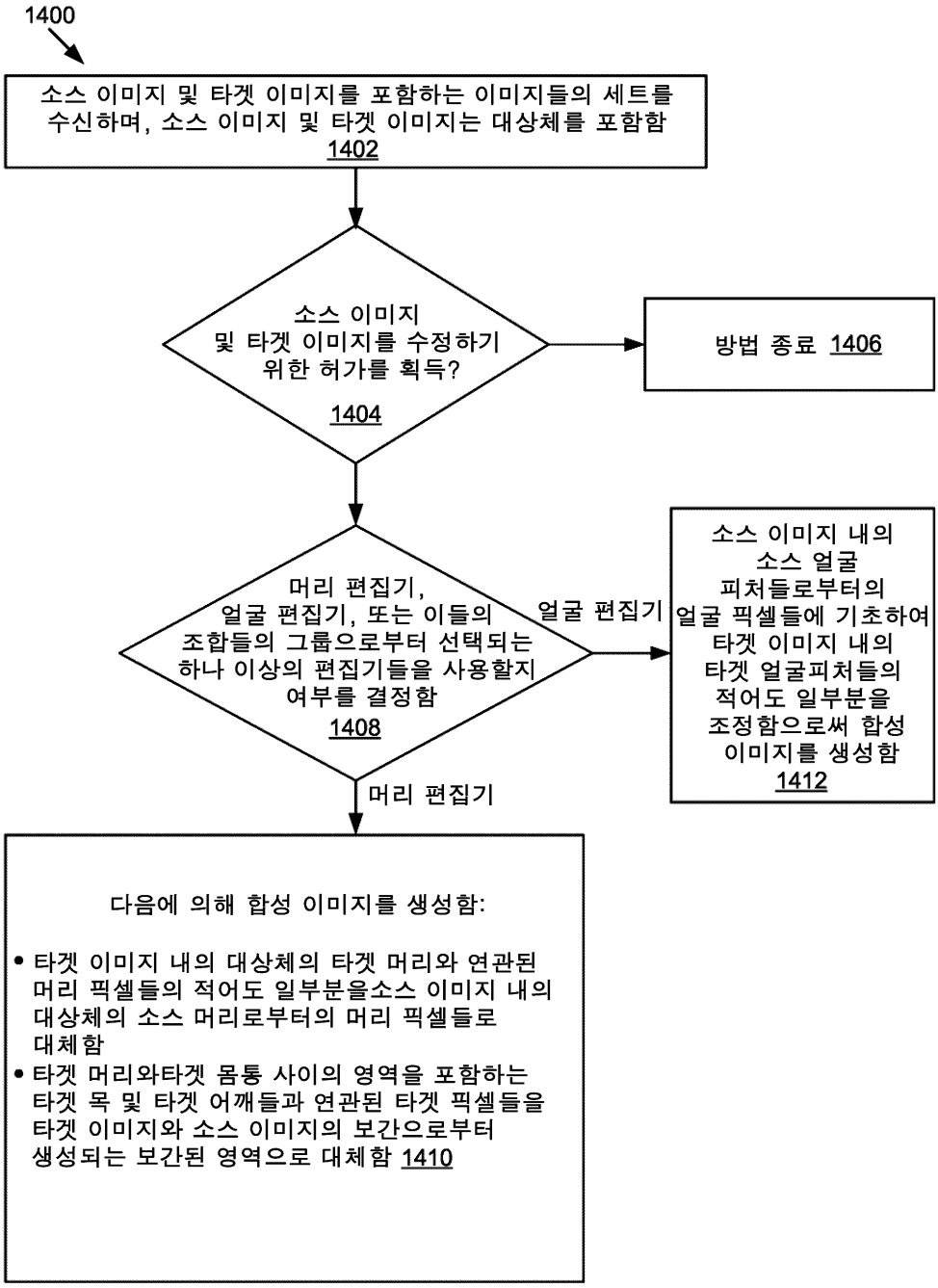
도면12b



도면13


1300
1305
1306
1307
1310
베스트 테이크
1350
1355
1357
1359
1361
1363
1365
리셋 1369
완료 1367
1375
1377
1379
복사본 저장
1385

도면14


1400
소스 이미지 및 타겟 이미지를 포함하는 이미지들의 세트를 수신하며, 소스 이미지 및 타겟 이미지는 대상체를 포함함
1402
소스 이미지 및 타겟 이미지를 수정하기 위한 허가를 획득?
1404
방법 종료 1406
머리 편집기, 얼굴 편집기, 또는 이들의 조합들의 그룹으로부터 선택되는 하나 이상의 편집기들을 사용할지 여부를 결정함
1408
얼굴 편집기
소스 이미지 내의 소스 얼굴 피처들로부터의 얼굴 픽셀들에 기초하여 타겟 이미지 내의 타겟 얼굴피처들의 적어도 일부분을 조정함으로써 합성 이미지를 생성함
1412
머리 편집기
다음에 의해 합성 이미지를 생성함:
• 타겟 이미지 내의 대상체의 타겟 머리와 연관된 머리 픽셀들의 적어도 일부분을소스 이미지 내의 대상체의 소스 머리로부터의 머리 픽셀들로 대체함
• 타겟 머리와타겟 몸통 사이의 영역을 포함하는 타겟 목 및 타겟 어깨들과 연관된 타겟 픽셀들을 타겟 이미지와 소스 이미지의 보간으로부터 생성되는 보간된 영역으로 대체함 1410