



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0013150
(43) 공개일자 2021년02월03일

(51) 국제특허분류(Int. Cl.)
G06T 15/60 (2006.01) G06N 3/04 (2006.01)
G06T 15/50 (2006.01) G06T 7/507 (2017.01)

(52) CPC특허분류
G06T 15/60 (2013.01)
G06N 3/0454 (2013.01)

(21) 출원번호 10-2020-7036796
(22) 출원일자(국제) 2019년05월13일
심사청구일자 없음
(85) 번역문제출일자 2020년12월21일
(86) 국제출원번호 PCT/US2019/031907
(87) 국제공개번호 WO 2019/226366
국제공개일자 2019년11월28일

(30) 우선권주장
201810510214.1 2018년05월24일 중국(CN)

(71) 출원인
마이크로소프트 테크놀로지 라이선싱, 엘엘씨
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이

(72) 발명자
동 위에
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 마이크로소프트 테크놀로지
라이선싱, 엘엘씨
천 구오준
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 마이크로소프트 테크놀로지
라이선싱, 엘엘씨
통 신
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 마이크로소프트 테크놀로지
라이선싱, 엘엘씨

(74) 대리인
김태홍, 김진희

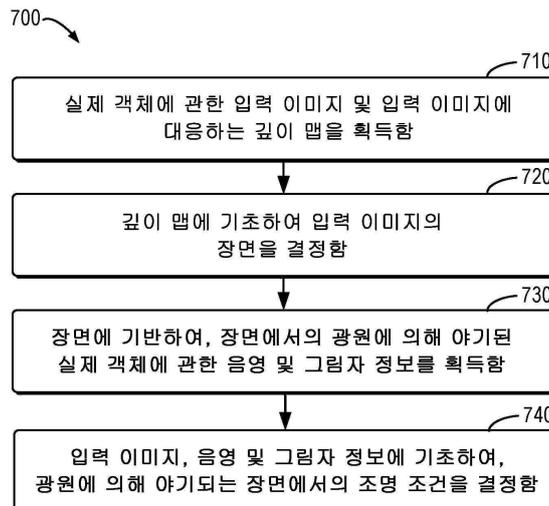
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **조명 추정**

(57) 요약

본 명세서에 설명된 주제의 구현에 따라, 조명 추정의 솔루션이 제공된다. 솔루션에서, 실제 객체에 관한 입력 이미지와 입력 이미지에 해당하는 깊이 맵이 획득된다. 입력 이미지의 장면의 기하학적 구조는 깊이 맵을 기반으로 결정된다. 장면의 광원에 의해 야기되는 실제 객체의 음영 및 그림자 정보는 장면의 결정된 기하학적 구조를 기반으로 결정된다. 그런 다음, 입력 이미지와 음영 및 그림자 정보를 기반으로 광원에 의해 야기되는 장면의 조명 조건이 결정된다. 솔루션에 따라 획득된 조명 조건을 사용하여 렌더링된 가상 객체는 실제 객체와 일치하는 사실적인 효과를 나타낼 수 있다.

대표도 - 도7



(52) CPC특허분류

G06T 15/506 (2013.01)

G06T 7/507 (2017.01)

G06T 2207/10028 (2013.01)

명세서

청구범위

청구항 1

전자 디바이스에 있어서,

처리 유닛; 및

상기 처리 유닛에 결합되고 상기 처리 유닛에 의한 실행을 위한 명령어들을 저장하는 메모리

를 포함하고, 상기 명령어들은 상기 처리 유닛에 의해 실행될 때 상기 디바이스로 하여금 동작(act)들을 수행하게 하고, 상기 동작들은,

실제 객체에 관한 입력 이미지 및 상기 입력 이미지에 대응하는 깊이 맵을 획득하는 동작;

상기 깊이 맵에 기반하여 상기 입력 이미지의 장면을 결정하는 동작;

상기 장면에 기반하여, 상기 장면에서의 광원에 의해 야기되는 실제 객체에 대한 음영 및 그림자 정보를 결정하는 동작; 및

상기 입력 이미지와 상기 음영 및 그림자 정보에 기반하여, 상기 광원에 의해 야기되는 상기 장면에서의 조명 조건을 결정하는 동작

을 포함하는 것인 전자 디바이스.

청구항 2

제1항에 있어서,

상기 동작들은,

상기 장면 및 조명 조건에 기반하여, 실제 객체 및 가상 객체를 포함하는 출력 이미지를 생성하는 동작

을 더 포함하고, 상기 출력 이미지는 상기 가상 객체를 상기 입력 이미지의 장면 내로 합성하는 것인 전자 디바이스.

청구항 3

제1항에 있어서,

상기 장면은 상기 실제 객체의 배치를 위한 표면을 포함하고,

상기 장면을 결정하는 동작은,

상기 깊이 맵을 포인트 클라우드로 변환하는 동작;

상기 포인트 클라우드 내의 복수의 포인트 - 상기 복수의 포인트는 표면을 나타냄 - 를 평면 내로 피팅하는 동작; 및

상기 평면 위에 위치된 상기 포인트 클라우드 내의 복수의 포인트를 3 차원 메시로 재구성하는 동작

을 포함하고,

상기 평면 및 상기 3차원 메시는 상기 입력 이미지의 장면을 공동으로 나타내는 것인 전자 디바이스.

청구항 4

제1항에 있어서,

상기 광원은 미리 결정된 모양의 이미지로 표현되고, 상기 이미지는 상기 광원의 복수의 부분에 대응하는 복수의 픽셀을 포함하고, 상기 복수의 픽셀의 각각의 픽셀 값은 상기 광원의 복수의 부분의 각각의 방사 조도 세기

를 나타내고,

상기 음영 및 그림자 정보를 결정하는 동작은,

상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 미리 결정된 값으로 초기화하는 동작; 및

상기 장면 및 초기화된 이미지에 기반하여 상기 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵 - 방사 조도 맵 각각은 광원의 대응 부분에 의해 야기되는 실제 객체에 대한 대응하는 음영 및 그림자 정보를 나타냄 - 을 생성하는 동작

을 포함하고;

상기 장면의 조명 조건을 결정하는 동작은,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작

을 포함하는 것인 전자 디바이스.

청구항 5

제4항에 있어서,

상기 광원을 나타내는 이미지는 크로스 모양의 이미지인 것인 전자 디바이스.

청구항 6

제4항에 있어서,

상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 신경망을 사용하여 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작

을 포함하는 것인 전자 디바이스.

청구항 7

제4항에 있어서,

상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 복수의 트레이닝된 신경망을 사용하여 점진적으로 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작

을 포함하는 것인 전자 디바이스.

청구항 8

제7항에 있어서,

상기 복수의 트레이닝된 신경망은 적어도 제1 신경망 및 제2 신경망을 포함하고,

상기 이미지 내의 상기 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 복수의 픽셀 내의 제1 픽셀의 제1 픽셀 값 및 상기 제1 신경망을 사용하여 상기 제1 픽셀에 인접한 적어도 하나의 픽셀의 픽셀 값들을 결정하는 동작;

상기 제1 픽셀, 상기 적어도 하나의 픽셀 및 상기 장면에 기반하여 제2 세트의 방사 조도 맵 - 상기 제2 세트의 방사 조도 맵은 상기 제1 픽셀의 복수의 서브-픽셀, 상기 적어도 하나의 픽셀 및 상기 복수의 픽셀 중 상기 제1 픽셀과 상기 적어도 하나의 픽셀이 아닌 픽셀들의 조합에 각각 대응함 - 을 생성하는 동작;

상기 입력 이미지 및 상기 제2 세트의 방사 조도 맵에 기반하여, 상기 제1 픽셀이 상기 제2 신경망을 사용하여 분할되어야 하는지 여부를 결정하는 동작; 및

상기 제1 픽셀이 분할되지 않아야 한다는 결정에 응답하여, 상기 제1 픽셀 값을 상기 제1 픽셀의 최종 픽셀 값

으로서 사용하는 동작
을 포함하는 것인 전자 디바이스.

청구항 9

제8항에 있어서,
상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은,
상기 제1 픽셀이 분할되어야 한다고 결정하는 것에 응답하여, 상기 제1 픽셀을 복수의 서브-픽셀로 분할하는 동작;
상기 복수의 서브-픽셀의 각각의 서브-픽셀 값을 결정하는 동작
을 포함하는 것인 전자 디바이스.

청구항 10

컴퓨터 구현 방법에 있어서,
실제 객체에 관한 입력 이미지 및 상기 입력 이미지에 대응하는 깊이 맵을 획득하는 단계;
상기 깊이 맵에 기반하여 상기 입력 이미지의 장면을 결정하는 단계;
상기 장면에 기반하여, 상기 장면의 광원에 의해 야기되는 실제 객체에 대한 음영 및 그림자 정보를 결정하는 단계; 및
상기 입력 이미지와 상기 음영 및 그림자 정보에 기반하여, 상기 광원에 의해 야기되는 상기 장면의 조명 조건을 결정하는 단계
를 포함하는 컴퓨터 구현 방법.

청구항 11

제10항에 있어서,
상기 장면 및 조명 조건에 기반하여, 실제 객체 및 가상 객체를 포함하는 출력 이미지를 생성하는 단계
를 더 포함하고, 상기 출력 이미지는 상기 가상 객체를 상기 입력 이미지의 장면 내로 합성하는 것인 컴퓨터 구현 방법.

청구항 12

제10항에 있어서,
상기 장면은 상기 실제 객체의 배치를 위한 표면을 포함하고,
상기 장면을 결정하는 단계는,
상기 깊이 맵을 포인트 클라우드로 변환하는 단계;
상기 포인트 클라우드 내의 복수의 포인트 - 상기 복수의 포인트는 표면을 나타냄 - 를 평면 내로 피팅하는 단계; 및
상기 평면 위에 위치된 상기 포인트 클라우드 내의 복수의 포인트를 3 차원 메시로 재구성하는 단계
를 포함하고,
상기 평면 및 상기 3차원 메시는 상기 입력 이미지의 장면을 공동으로 나타내는 것인 컴퓨터 구현 방법.

청구항 13

제10항에 있어서,
상기 광원은 미리 결정된 모양의 이미지로 표현되고, 상기 이미지는 상기 광원의 복수의 부분에 대응하는 복수의 픽셀을 포함하고, 상기 복수의 픽셀의 각각의 픽셀 값은 상기 광원의 복수의 부분의 각각의 방사 조도 세기

를 나타내고,

상기 음영 및 그림자 정보를 결정하는 단계는,

상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 미리 결정된 값으로 초기화하는 단계; 및

상기 장면 및 초기화된 이미지에 기반하여 상기 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵 - 방사 조도 맵 각각은 광원의 대응 부분에 의해 야기되는 실제 객체에 대한 대응하는 음영 및 그림자 정보를 나타냄 - 을 생성하는 단계

를 포함하고;

상기 장면의 조명 조건을 결정하는 단계는,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계

를 포함하는 것인 컴퓨터 구현 방법.

청구항 14

제13항에 있어서,

상기 광원을 나타내는 이미지는 크로스 모양의 이미지인 것인 컴퓨터 구현 방법.

청구항 15

제13항에 있어서,

상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는,

상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 신경망을 사용하여 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계

를 포함하는 것인 컴퓨터 구현 방법.

발명의 설명

기술 분야

[0001] 본 발명은 조명 추정에 관한 것이다.

배경 기술

[0002] 3 차원 가상 객체(virtual object)를 실제 장면에 관한 이미지로 합성하는 것은 증강 현실, 이미지 편집, 영화 제작 등과 같은 많은 애플리케이션에서 중요한 작업이다. 실사(photorealistic) 효과를 얻기 위해서는 합성 이미지 내의 가상 객체가 장면에서 실제 객체(real object)의 것과 일치하는 음영(shading) 및 그림자(shadow)를 나타낼 수 있게 하고, 합성 이미지가 가상 객체와 실제 객체 사이의 투영 및/또는 폐색을 올바르게 표시할 수 있게 하기 위하여 실제 장면의 조명 조건(lighting condition)을 추정할 필요가 생긴다. 최근 깊이 카메라의 발전으로 실제 장면의 3 차원 기하학적 구조를 쉽게 획득할 수 있다. 그러나, 실제 장면에 관한 단일 이미지로부터 장면의 조명 조건을 추정하는 것은 여전히 어려운 작업이다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0003] 본 명세서에 설명된 주제(subject matter)의 구현에 따라, 조명 추정의 솔루션이 제공된다. 이 솔루션에서는 실제 객체에 대한 입력 이미지와 입력 이미지에 해당하는 깊이 맵을 얻는다. 입력 이미지의 장면의 기하학적 구조

는 깊이 맵을 기반으로 결정된다. 장면의 광원에 의해 발생하는 실제 객체의 음영 및 그림자 정보는 장면의 결정된 기하학적 구조를 기반으로 결정된다. 그런 다음, 입력 이미지와 음영 및 그림자 정보를 기반으로 광원에 의해 야기되는 장면의 조명 조건이 결정된다.

[0004] 본 명세서에 기술된 주제에 따른 조명 추정 솔루션은 장면에서 실제 객체의 재료 특성에 대한 어떠한 가정도 하지 않고 입력 이미지 및 입력 이미지에 대한 음영 및 그림자 정보를 기반으로 여러 레벨의 신경망을 사용하여 장면의 조명 조건을 점진적으로 추정할 수 있다. 표준 광 베이스(canonical light base) 세트의 선형 조합으로 장면의 광원을 모델링하고 표준 광 베이스 세트로부터 렌더링된 방사 조도 맵(irradiance map) 세트를 신경망의 입력으로서 활용함으로써, 솔루션은 조명 추정의 복잡성을 크게 단순화하고, 결과의 정확도를 향상시킬 수 있다. 또한 멀티 트리 기반의 점진적 추정을 활용하여, 이 솔루션이 신경망의 각 레벨의 트레이닝 및 계산 비용을 절감하고 견고성(robustness)을 향상시킬 수 있다. 솔루션에 따라 얻은 조명 조건을 사용하여 렌더링된 가상 객체는 실제 객체와 일치하는 사실적인 효과를 나타낼 수 있다.

[0005] 이 요약은 아래의 상세한 설명에서 더 설명되는 단순화된 형태로 개념의 선택을 소개하기 위해 제공된다. 이 요약은 청구된 주제의 주요 특징 또는 필수 특징을 식별하기 위한 것이 아니며 본 명세서에 설명된 주제의 범위를 제한하는 데 사용되는 의도도 아니다.

도면의 간단한 설명

- [0006] 도 1은 본 명세서에 설명된 주제의 복수의 구현을 구현할 수 있는 컴퓨팅 디바이스(100)의 블록도를 도시한다.
 - 도 2는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 시스템(200)을 도시한다.
 - 도 3은 본 명세서에 설명된 주제의 일부 구현에 따른 시스템(200)의 조명 조건 결정 서브 시스템의 다이어그램을 예시한다.
 - 도 4는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 초기화 네트워크의 동작 다이어그램을 도시한다.
 - 도 5는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 리파인(refinement) 네트워크의 동작 다이어그램을 도시한다.
 - 도 6은 본 명세서에 설명된 주제의 일부 구현에 따라 다중 레벨의 신경망을 사용하여 장면의 조명 조건을 점진적으로 결정하는 다이어그램을 도시한다.
 - 도 7은 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 프로세스의 흐름도를 도시한다.
- 도면 전체에 걸쳐 동일하거나 유사한 참조 부호는 동일하거나 유사한 요소를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0007] 본 명세서에 설명된 주제는 이제 몇 가지 예시적인 구현을 참조하여 논의될 것이다. 이러한 구현은 당업자가 주제의 범위에 대한 어떠한 제한을 제안하기보다는 본 명세서에 설명된 주제를 더 잘 이해하고 이에 따라 구현할 수 있도록 하기 위한 목적으로만 논의된다는 것을 이해해야 한다.
- [0008] 본 명세서에 사용된 용어 "포함한다" 및 그 변형은 "포함하지만 이에 제한되지 않는다"를 의미하는 개방형 용어로 해석되어야 한다. "기반하여"라는 용어는 "적어도 부분적으로 기반하여"로 읽혀져야 한다. 용어 "하나의 구현" 및 "구현"은 "적어도 하나의 구현"으로 읽혀져야 한다. 용어 "다른 구현"은 "적어도 하나의 다른 구현"으로 읽혀져야 한다. 용어 "제1", "제2" 등은 상이한 또는 동일한 객체를 지칭할 수 있다. 명시적 및 암시적 다른 정의가 아래에 포함될 수 있다.
- [0009] 예시적 환경
- [0010] 본 명세서에 설명된 주제의 기본 원리 및 다양한 예시적인 구현이 이제 도면을 참조하여 설명될 것이다. 도 1은 본 명세서에 설명된 주제의 구현이 구현될 수 있는 컴퓨팅 환경(100)의 블록도를 예시한다. 도 1에 도시된 컴퓨팅 디바이스(100)는 임의의 방식으로 본 명세서에 설명된 주제의 구현의 기능 및 범위에 대한 어떠한 제한도 제안하지 않고 단지 예시적인 것임을 이해해야 한다. 도 1에 도시된 바와 같이, 컴퓨팅 디바이스(100)는 일반적인 컴퓨터 디바이스 형태의 컴퓨팅 디바이스(100)를 포함한다. 컴퓨팅 디바이스(100)의 컴포넌트는 하나 이상의 프로세서 또는 처리 유닛(110), 메모리(120), 저장 디바이스(130), 하나 이상의 통신 유닛(140), 하나 이상의 입

력 디바이스(150) 및 하나 이상의 출력 디바이스(160)를 포함하지만, 이에 제한되지 않는다.

- [0011] 일부 구현에서, 컴퓨팅 디바이스(100)는 컴퓨팅 능력을 갖는 다양한 사용자 단말 또는 서비스 단말로서 구현될 수 있다. 서비스 단말은 다양한 서비스 제공자가 제공하는 서버, 대규모 컴퓨팅 디바이스 등이 될 수 있다. 예를 들어, 사용자 단말은 이동 전화, 스테이션, 유닛, 디바이스, 멀티미디어 컴퓨터, 멀티미디어 태블릿, 인터넷 노드, 커뮤니케이터, 데스크톱 컴퓨터, 랩톱 컴퓨터, 노트북 컴퓨터, 넷북 컴퓨터, 태블릿 컴퓨터, PCS(Personal Communication System) 디바이스, 개인용 내비게이션 디바이스, PDA(Personal Digital Assistant), 오디오/비디오 플레이어, 디지털 카메라/비디오 플레이어, 포지셔닝 디바이스, 텔레비전 수신기, 라디오 방송 수신기, 전자 책 디바이스, 게임 디바이스 또는 이들의 임의의 조합을 포함하거나, 이들 디바이스의 액세서리 및 주변 장치를 포함하거나, 이들의 임의의 조합을 포함한 임의의 타입의 모바일 단말, 고정 단말 또는 휴대용 단말이다. 컴퓨팅 디바이스(100)가 사용자를 위한 임의의 타입의 인터페이스(예를 들어, "웨어러블" 회로 등)를 지원할 수 있다는 것이 추가로 예측될 수 있다.
- [0012] 처리 유닛(110)은 물리적 또는 가상 프로세서일 수 있으며 메모리(120)에 저장된 프로그램을 기반으로 다양한 프로세스를 실행할 수 있다. 다중 프로세서 시스템에서, 다중 처리 유닛은 컴퓨팅 디바이스(100)의 병렬 처리 능력을 개선하기 위해 컴퓨터 실행 가능한 명령어들을 병렬로 실행한다. 처리 유닛(110)은 또한 중앙 처리 장치(central processing unit, CPU), 마이크로프로세서, 컨트롤러 및 마이크로컨트롤러로 지칭된다.
- [0013] 컴퓨팅 디바이스(100)는 일반적으로 컴퓨팅 디바이스(100)에 의해 액세스 가능한 임의의 이용 가능한 매체일 수 있는 복수의 컴퓨터 저장 매체를 포함하며, 이는 휘발성 및 비휘발성 매체, 및 이동식 및 비이동식 매체를 포함하지만 이에 제한되지 않는다. 메모리(120)는 휘발성 메모리(예를 들어, 레지스터, 캐시, RAM(Random Access Memory)), 비휘발성 메모리(예를 들어, ROM(Read-Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), 플래시 메모리) 또는 이들의 임의의 조합일 수 있다. 메모리(120)는 본 명세서에 설명된 다양한 구현의 기능을 수행하도록 구성된 프로그램 모듈을 갖는 이미지 처리 모듈(122)을 포함한다. 이미지 처리 모듈(122)은 처리 유닛(110)에 의해 액세스되고 동작되어, 각각의 기능을 구현할 수 있다.
- [0014] 저장 디바이스(130)는 임의의 이동식 또는 비이동식 매체일 수 있으며, 정보 및/또는 데이터를 저장하는데 사용되고 컴퓨팅 디바이스(100)에서 액세스될 수 있는 머신 판독 가능 매체를 포함할 수 있다. 컴퓨팅 디바이스(100)는 추가적인 이동식/비이동식, 휘발성/비휘발성 메모리 매체를 추가로 포함할 수 있다. 도 1에는 도시되지 않았지만, 이동식 및 비휘발성 디스크를 읽고 쓰기 위한 디스크 드라이브가 제공되고, 이동식 비휘발성 디스크를 읽고 쓰기 위한 디스크 드라이브가 제공된다. 이 경우 각 드라이브는 하나 이상의 데이터 매체 인터페이스를 통해 버스(도시되지 않음)에 연결된다.
- [0015] 통신 유닛(140)은 통신 매체를 통해 추가적인 컴퓨팅 디바이스와 통신한다. 추가로, 컴퓨팅 디바이스(100)의 컴포넌트의 기능은 통신을 위해 통신 가능하게 연결된 단일 컴퓨팅 클러스터 또는 다중 컴퓨팅 머신에 의해 구현될 수 있다. 따라서, 컴퓨팅 디바이스(100)는 하나 이상의 다른 서버, 네트워크 개인용 컴퓨터(PC) 또는 다른 일반 네트워크 노드와의 논리 링크를 사용하여 네트워킹 환경에서 동작될 수 있다.
- [0016] 입력 디바이스(150)는 마우스, 키보드, 트래킹 볼, 음성 입력 디바이스 등과 같은 하나 이상의 입력 디바이스를 포함할 수 있다. 출력 디바이스(160)는 디스플레이, 라우드 스피커, 프린터 등과 같은 하나 이상의 출력 디바이스를 포함할 수 있다. 필요에 따라, 컴퓨팅 디바이스(100)는 또한 통신 유닛(140)을 통해 저장 디바이스, 디스플레이 디바이스 등과 같은 하나 이상의 외부 디바이스(미도시), 사용자가 컴퓨팅 디바이스(100)와 상호 작용할 수 있게 하는 하나 이상의 디바이스, 또는 컴퓨팅 디바이스(100)가 하나 이상의 다른 컴퓨팅 디바이스(예를 들어, 네트워크 카드, 모뎀 등)와 통신할 수 있게 하는 임의의 디바이스와 통신할 수 있다. 이러한 통신은 입력/출력(I/O) 인터페이스(미도시)를 통해 수행된다.
- [0017] 컴퓨팅 디바이스(100)는 본 명세서에 설명된 주제의 복수의 구현에서 조명 추정을 구현하는 데 사용될 수 있다. 따라서, 이하에서는 컴퓨팅 디바이스(100)를 "이미지 처리 디바이스(imaging processing device)(100)"라고도 한다. 조명 추정을 수행할 때, 이미지 처리 디바이스(100)는 입력 디바이스(150)를 통해 실제 객체(예를 들어, 커피 잔, 매직 큐브 등)에 관한 입력 이미지(171) 및 입력 이미지(171)에 대응하는 깊이 맵(172)을 수신할 수 있다. 예를 들어, 깊이 맵(172)은 깊이 카메라에 의해 캡처될 수 있다. 이후, 이미지 처리 디바이스(100)는 입력 이미지(171) 및 깊이 맵(172)에 기반하여 트레이닝된 복수의 신경망을 이용하여 입력 이미지(171)의 장면의 조명 조건을 결정할 수 있다. 예를 들어, 조명 조건은 장면의 광원을 나타내는 특정 모양의 이미지(미도시)로 표현될 수 있다. 이미지는 광원의 복수 부분에 대응하는 복수의 픽셀을 포함할 수 있으며, 복수의 픽셀의 각 픽셀 값은 광원의 복수 부분의 각각의 방사 조도(irradiance) 세기를 나타낸다. 또한, 이미지 처리 디바이스(100)

0)는 또한 장면의 결정된 조명 조건에 기반하여 입력 이미지(171)의 장면 내에 가상 객체(예를 들어, 헬로 키티 인형)를 합성하여, 출력 이미지(173)을 생성할 수 있다. 출력 이미지(173)의 가상 객체는 실제 객체와 일치하는 사실적인 효과를 나타낼 수 있다. 이미지 처리 디바이스(100)는 출력 디바이스(160)를 통해 이미지(173)를 더 출력할 수 있다.

[0018] 일부 레거시 솔루션은 또한 실제 장면에 관한 이미지로부터 장면의 조명 조건을 복구하려고 시도한다. 그러나 이러한 레거시 솔루션은 일반적으로 일부 또는 모든 환경 정보를 포함하는 원격 보기(remote-view) 이미지에 적용할 수 있지만, 복수의 애플리케이션에서 꽤 일반적인 근접 보기(close-view) 이미지는 적합하지 않다(즉, 실제 객체만 이미지에 표시되는 반면, 객체의 조명 효과로 이어지는 환경 정보는 거의 보이지 않는다). 근접 보기 이미지의 경우, 일부 다른 레거시 솔루션은 실제 객체의 음영 정보로부터 실제 장면의 조명 조건을 복구하는 것을 제안한다. 그러나 이러한 솔루션은 일반적으로 객체의 재질 또는 객체 표면의 반사를 알고 있어야 한다. 이러한 솔루션은 무작위 재질이나 표면 반사를 가진 객체에 관한 이미지를 기반으로 실제 장면의 조명 조건을 복구할 수 없다.

[0019] 기존의 조명 추정 솔루션의 몇 가지 문제점이 위에서 논의되었다. 본 명세서에 설명된 주제의 구현에 따르면, 상기 문제 및 하나 이상의 다른 잠재적 문제를 해결하기 위한 조명 추정의 솔루션이 제공된다. 이 솔루션은 입력 이미지의 음영 및 그림자 정보를 기반으로 다수 레벨의 신경망을 사용하여 장면의 조명 조건을 점진적으로 추정할 수 있다. 따라서, 솔루션은 장면에서 실제 객체의 물성에 대한 어떠한 가정도 하지 않고, 원격 보기 이미지와 근접 보기 이미지 모두에 적용할 수 있다. 표준 광 베이스 세트의 선형 조합을 사용하여 장면의 광원을 모델링하고 표준 광 베이스 세트로부터 렌더링된 방사 조도 맵 세트를 신경망의 입력으로서 활용함으로써, 솔루션은 조명 추정의 복잡성을 크게 단순화하고, 결과의 정확성을 향상시킬 수 있다. 또한 다중 트리 기반의 점진적 추정을 활용함으로써 솔루션은 각 레벨의 신경망의 트레이닝 및 계산 비용을 저감하고 그 견고성을 향상시킬 수 있다. 솔루션에 따른 조명 추정 결과를 사용하여 렌더링된 가상 객체는 실제 객체와 일치하는 음영 및 그림자 효과를 나타낼 수 있다.

[0020] 솔루션의 다양한 예시적인 구현이 도면을 참조하여 아래에서 더 상세히 설명될 것이다.

[0021] 시스템 구조

[0022] 도 2는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 시스템(200)의 블록도를 예시한다. 일부 구현에서, 시스템(200)은 도 1의 컴퓨팅 디바이스(100)의 이미지 처리 모듈(122)의 적어도 일부로서 구현될 수 있으며, 즉 컴퓨터 프로그램 모듈로서 구현될 수 있다. 대안적으로, 다른 구현에서, 시스템(200)은 하드웨어 디바이스에 의해 완전히 또는 부분적으로 구현될 수 있다. 도 2에 도시된 바와 같이, 시스템(200)은 일반적으로 장면 결정 서버 시스템(210) 및 조명 결정 서버 시스템(220)을 포함할 수 있다. 시스템(200)의 구조 및 기능은 본 명세서에 설명된 주제의 범위에 대한 어떠한 제한도 암시하지 않고 예시 목적을 위해서만 본 명세서에서 설명된다는 것을 이해해야 한다. 본 명세서에 설명된 주제의 구현은 상이한 구조 및/또는 기능으로 구현될 수 있다.

[0023] 도 2에 도시된 바와 같이, 시스템(200)의 입력은 평평한 표면에 놓인 실제 객체에 관한 입력 이미지(201)(예를 들어, 입력 이미지(201)는 도 1에 도시된 바와 같은 입력 이미지(171)일 수 있음) 및 입력 이미지(201)에 대응하는 깊이 맵(202)(예를 들어, 깊이 맵(202)은 도 1에 도시된 바와 같은 깊이 맵(172)일 수 있음)일 수 있다. 입력 이미지(201)는 예를 들어 임의의 크기 및/또는 포맷일 수 있다. 일부 구현에서, 입력 이미지(201)는 공통 RGB 3-채널 이미지일 수 있다. 깊이 맵(202)은 깊이 카메라에 의해 캡처될 수 있으며, 각 픽셀의 값은 깊이 카메라와 픽셀에 대응하는 실제 장면의 포인트 사이의 실제 거리를 반영한다. 게다가, 깊이 맵(202)은 입력 이미지(201)와 정렬될 수 있다. 즉, 깊이 맵(202)의 픽셀은 입력 이미지(201)의 픽셀과 일대일 대응이다. 일부 구현에서, 깊이 맵(202)에서 픽셀의 값에 의해 표시되는 깊이는 깊이 맵(202)이 실제 장면의 대략적인 기하학적 구조를 나타낼 수 있는 한, 정확하지 않을 수 있다.

[0024] 장면 결정 서버 시스템(210)은 깊이 맵(202)에 기반하여 입력 이미지(201)에서 실제 장면(220)을 결정할 수 있다.

[0025] 일부 구현에서, 입력 이미지(201)에서 실제 장면(220)을 결정하기 위해, 장면 결정 서버 시스템(210)은 먼저 실제 장면을 설명하는 좌표계를 결정할 수 있다. 예를 들어, 장면 결정 서버 시스템(210)은 실제 장면의 좌표계에서 XY 평면으로서 실제 객체의 배치를 위한 평평한 표면을 정의할 수 있는 반면, XY 평면의 법선은 실제 장면의 좌표계에서 Z 축으로서 작용한다.

- [0026] 일부 구현에서, 입력 이미지(201)에서 실제 장면을 결정하기 위해, 장면 결정 서버 시스템(210)은 먼저 깊이 맵(202)을 좌표계의 포인트 클라우드로 변환할 수 있다. 그 후, 장면 결정 서버 시스템(210)은 포인트 클라우드 내의 실제 객체를 배치하기 위한 평평한 표면을 나타내는 복수의 포인트를 평면 내에 피팅(fit)할 수 있다. 일부 구현에서, 복수의 평면을 포함하는 장면의 경우, 장면 결정 서버 시스템(210)은 실제 객체의 배치를 위한 평면으로서 복수의 평면 중 가장 큰 평면을 사용할 수 있다. 포인트 클라우드 내의 실제 객체의 배치를 위한 평평한 표면의 복수의 포인트를 평면 내에 피팅한 후에, 장면 결정 서버 시스템(210)은 평평한 표면을 나타내는 복수의 포인트를 제거할 수 있다. 실제 객체의 배치를 위한 평평한 표면을 나타내는 복수의 포인트를 평면 내에 피팅한 후, 장면 결정 서버 시스템(210)은 또한 피팅된 평면 위에 위치한 포인트 클라우드의 복수의 포인트를 3차원(3D) 메시(mesh)로 재구성할 수 있다. 예를 들어, 장면 결정 서버 시스템(210)은 피팅된 평면 위에 위치한 포인트 클라우드의 복수의 포인트를 푸아송(Poisson) 표면 재구성을 통해 3차원(3D) 메시로 재구성할 수 있다. 피팅된 평면 및 재구성된 3D 메시는 입력 이미지(201)의 장면(220)을 공동으로 나타낼 수 있다.
- [0027] 장면 결정 서버 시스템(210)에 의해 결정된 장면(220) 및 입력 이미지(201)는 장면의 조명 추정을 위해 조명 조건 결정 서버 시스템(230)에 입력될 수 있다. 일부 구현에서, 장면(220)의 조명 조건을 추정하기 위해, 조명 조건 결정 서버 시스템(230)은 장면(220)에서 환경 광원을 모델링할 수 있다.
- [0028] 일부 구현에서, 조명 조건 결정 서버 시스템(230)은 좌표계의 원점에 중심을 두고 위에서 설명된 좌표계의 축과 정렬된 큐브를 사용하여 장면(220)에서 환경 광원을 모델링할 수 있다. 큐브를 확장함으로써 크로스(cross) 모양의 이미지를 얻을 수 있다. 따라서, 장면(220)에서 환경 광원은 크로스 이미지의 이미지로 표현될 수 있다. 일부 다른 구현들에서, 조명 조건 결정 서버 시스템(230)은 구형 또는 다른 모양과 같은 3차원 모양을 사용하여 장면(220)에서 환경 광원을 모델링할 수 있다. 이에 대응하여, 장면(220)에서 환경 광원은 구형 또는 다른 모양과 같은 3차원 모양에 대응하는 다른 모양의 이미지로 표현될 수 있다. 설명의 편의를 위해, 이러한 맥락에서, (예를 들어, 도 2의 크로스 이미지(240)에 의해 도시된 바와 같이) 크로스 이미지는 장면(200)에서 환경 광원의 표현의 예로서 역할을 한다. 이것은 본 명세서에 설명된 주제의 범위에 대한 어떠한 제한도 제안하지 않고 예시의 목적으로만 제공된다는 것을 이해해야 한다. 본 명세서에 설명된 주제는 다른 표현을 사용하여 장면에서 환경 광원을 모델링할 수 있다.
- [0029] 일부 구현에서, 광원을 나타내는 크로스 이미지(240)는 장면(220)에서 환경 광원의 복수의 부분에 각각 대응하는 복수의 픽셀을 각각 포함할 수 있다. 예를 들어, 복수의 픽셀의 각각의 픽셀 값은 각각 장면(220)에서 환경 광원의 복수의 부분의 각각의 방사 조도 세기를 나타낼 수 있다. 문맥에서, 장면(240)에서 환경 광원을 나타내는 이미지(240)의 각 픽셀은 또한 "표준 광 베이스(canonical light base)"이라고도 지칭되며 장면(220)에서 환경 광원을 나타내는 이미지(240)는 또한 "환경 맵(environment map)"이라고도 지칭된다. 일부 구현에서, 조명 조건 결정 서버 시스템(230)은 장면(220) 및 입력 이미지(201)에 기반하여 환경 맵(240)에서 복수의 픽셀의 각각의 픽셀 값을 결정할 수 있고, 이에 의해 장면(220)에서 조명 조건을 결정할 수 있다.
- [0030] 조명 조건 결정 서버 시스템(230)의 구조 및 작동 원리는 도 3을 참조하여 아래에서 상세히 설명될 것이다.
- [0031] 작동 원리
- [0032] 도 3은 본 명세서에 설명된 주제의 일부 구현에 따른 조명 조건 결정 서버 시스템(230)의 블록도를 예시한다. 도 3에 도시된 바와 같이, 조명 조건 결정 서버 시스템(230)은 일반적으로 방사 조도 맵 생성 모듈(320), 제1 신경망(340), 방사 조도 맵 생성 모듈(360) 및 제2 신경망(350)을 포함할 수 있다. 조명 조건 결정 서버 시스템(230)의 구조 및 기능은 본 명세서에 설명된 주제의 범위에 대한 어떠한 제한도 제안하지 않고 예시의 목적으로만 제공된다는 것을 이해해야 한다. 본 명세서에 설명된 주제의 구현은 상이한 구조 및/또는 기능으로 구현될 수 있다.
- [0033] 진술한 바와 같이, 조명 조건 결정 서버 시스템(230)은 특정 모양의 이미지로서 장면(220)에서 환경 광원을 나타낼 수 있다. 문맥에서 4×4×3 픽셀의 해상도를 가진 크로스 이미지가 환경 광원의 표현의 예로서 사용된다. 그러나, 이것은 본 명세서에 설명된 주제의 범위에 대한 어떠한 제한도 암시하지 않고 단지 예로서 제공된다는 것을 이해해야 한다. 일부 구현들에서, 조명 조건 결정 서버 시스템(230)은 미리 결정된 값으로 장면(220)에서 환경 광원을 나타내는 크로스 이미지(즉, 초기 환경 맵(310))의 4×4×3 픽셀의 값을 초기화할 수 있다. 예를 들어, 초기 환경 맵(310)의 각 픽셀의 값은 정규화된 방사 조도 세기 "1"로 초기화될 수 있다.
- [0034] 방사 조도 맵(IRRADIANCE MAP) 생성
- [0035] 일부 구현에서, 방사 조도 맵 생성 모듈(320)은 장면(220)에 기반하여, 장면(220)의 광원에 의해 야기되는 입력

이미지(201)의 실제 객체에 대한 음영 및 그림자 정보를 결정할 수 있다. 구체적으로, 방사 조도 맵 생성 모듈(320)은 장면(220) 및 초기 환경 맵(310)에 기반하여 초기 환경 맵(310)의 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵(330)을 생성할 수 있다. 여기서 설명된 "방사 조도 맵"은 입력 이미지의 픽셀 각각에 대한 광원의 복사 플럭스(radiant flux)를 기록하고, 이는 광원에 의해 야기되는 입력 이미지의 음영 및 그림자 정보를 반영할 수 있다. 예를 들어, 제1 방사 조도 맵 세트 각각은 입력 이미지(201)와 동일한 크기를 가질 수 있고, 광원의 각 부분에 의해 야기된 입력 이미지(201)의 실제 객체에 대한 음영 및 그림자 정보를 나타낼 수 있다.

[0036] 일부 구현에서, 광원 상의 특정 포인트 L(예를 들어, 광원을 나타내는 크로스 이미지의 픽셀)이 주어지면, 이에 대응하는 방사 조도 맵은 입력 이미지(201)에서 각 픽셀에 대해 포인트 L의 복사 플럭스를 기록할 수 있다. 예를 들어, 입력 이미지(201)의 픽셀 x에 대해, 광원 상의 포인트 L에 의해 야기되는 복사 플럭스는 다음과 같이 표현될 수 있다:

수학식 1

[0037]
$$r(x) = V(x, l)L(l) (N(x) \cdot l)$$

[0038] 여기서 N(x)는 픽셀 x에서의 법선 방향을 나타내고, l은 광원 상의 포인트 L에서 픽셀 x까지의 조명 방향을 나타내고, L(l)은 방향 l을 따른 방사 조도 세기를 나타내고, V(x, l)는 l 방향을 따라 광원 상의 포인트 L에 대한 픽셀 x의 가시성을 나타낸다. 유사하게, 주어진 광원 상의 복수의 포인트들의 조합에 대해, 복수의 포인트들의 조합에 의해 야기되는 복사 플럭스는 수학식 1을 적분함으로써 결정될 수 있다. 예를 들어, 입력 이미지(201)의 픽셀 x에 대해 광원 상의 복수의 포인트(예를 들어, "Ω+"로 표시됨)의 조합에 의해 야기되는 복사 플럭스는 다음과 같이 표현될 수 있다:

수학식 2

[0039]
$$R(x) = \int_{\Omega^+} V(x, l)L(l) (N(x) \cdot l) dl$$

[0040] 수학식 1에 기반하여, 방사 조도 맵 생성 모듈(320)은 초기 환경 맵(310)의 픽셀(즉, 장면(220)에서 환경 광원의 포인트)에 대하여, 그에 대응하는 방사 조도 맵을 생성할 수 있다. 이러한 방식으로, 방사 조도 맵 생성 모듈(320)은 초기 환경 맵(310)에서 복수의 픽셀과 일대일 대응을 갖는 제1 세트의 방사 조도 맵(330)을 생성할 수 있다.

[0041] 초기화 네트워크

[0042] 도 3에 도시된 바와 같이, 제1 세트의 방사 조도 맵(330) 및 입력 이미지(201)는 제1 트레이닝된 신경망(340)(이는 본 명세서에서 또한 "초기화 네트워크"라고도 지칭됨)에 입력될 수 있다. 예를 들어, 제1 신경망(340)은 컨볼루션 신경망(convolution neural network, CNN)일 수 있으며, 이는 초기 환경 맵(310)에서 복수의 픽셀의 각각의 픽셀 값을 결정하도록 트레이닝되어 환경 맵(310)에 대응하는 환경 맵(350)을 생성할 수 있다. 조명 추정 은 입력 이미지의 음영 및 그림자 정보를 나타내는 방사 조도 맵을 신경망에 제공함으로써 수행되며, 따라서 본 명세서에 설명된 주제의 구현은 장면에서 실제 객체의 재질 속성에 대한 어떠한 가정도 요구하지 않고 원격 보기 이미지와 근접 보기 이미지 모두에 적용 가능하다. 여기서 CNN은 제1 신경망(340)의 예로서 제공되지만, 본 명세서에 설명된 주제의 구현은 다른 타입의 신경망에 적용될 수 있으며, 본 명세서에 설명된 주제의 범위는 이 측면에서 제한되지 않음을 이해해야 한다.

[0043] 도 4는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 초기화 네트워크(340)의 동작 다이어그램이다. 도 4는 초기 환경 맵(310)에 기반하여 생성된 입력 이미지(201) 및 제1 세트의 방사 조도 맵(330)을 도시한다. 도 4에 도시된 바와 같이, 초기화 네트워크(340)는 초기화 환경(310)에서 복수의 픽셀의 각각의 픽셀 값을 입력 이미지(201) 및 제1 세트의 방사 조도 맵(330)에 기반하여 결정할 수 있고, 따라서 환경 맵(310)에 대응하는 환경 맵(350)을 생성할 수 있다.

[0044] 일부 구현에서, 초기화 네트워크(340) 내부에서, 입력 이미지(201)는 3-계층 인코더에 의해 처리될 수 있고, 각

각의 방사 조도 맵(330)은 다른 2-계층 인코더에 의해 개별적으로 처리될 수 있다. 각각의 방사 조도 맵(330)에 대한 인코더는 동일한 구조 및 가중치를 공유할 수 있다. 인코더의 각 계층은 컨볼루션 계층, 배치(batch) 정규화 계층 및 여기(excitation) 계층(예를 들어, 활성화 함수로서 ReLU(rectified linear unit)을 사용함)를 포함할 수 있으며, 이후에 최대 풀링 계층에 의해 다음 스케일로 다운 샘플링될 수 있다. 인코더를 거친 후, 각각의 방사 조도 맵(330)에서 추출된 특징 맵과 입력 이미지(201)가 연결(concatenate)되고 그 후 다른 3-계층 인코더에 의해 처리된다. 마지막으로, 초기 환경 맵(310)에서 복수의 픽셀의 각각의 픽셀 값은 완전 연결(fully-connected, FC) 계층에 이어 3-계층 인코더로부터 획득될 수 있다. 이러한 픽셀 값은 환경 맵(310)에 대응하는 환경 맵(350)으로 조직화될 수 있다.

[0045] 일부 구현에서, 구성된 트레이닝 데이터 세트는 초기화 네트워크(340)를 트레이닝하는 데 사용될 수 있다. 예를 들어, 3D 객체, 및 3D 객체의 그룹으로부터 무작위로 선택된 환경 맵 및 환경 맵들은 이미지를 렌더링하고 생성하는데 사용될 수 있다. 그런 다음, 렌더링 동안 적용된 렌더링된 이미지 및 환경 맵은 초기화 네트워크(340)의 트레이닝 데이터 세트 내로 수집될 수 있다. 예를 들어, 초기화 네트워크(340)의 트레이닝 절차 동안, L2 손실 함수를 사용하여 트레이닝 데이터와 예측 결과 사이의 차이를 측정할 수 있다.

[0046] 도 3으로 돌아가면, 환경 맵(350)을 생성함으로써, 조명 조건 결정 서버 시스템(230)은 장면(220)의 조명 조건을 미리 결정할 수 있다. 그러나, 장면(220)에 있는 객체의 모양과 레이아웃의 다양성으로 인해 장면의 조명 조건과 입력 이미지(201) 및 깊이 맵(202) 간의 매핑은 비선형일 수 있다. 문제를 해결하기 위해, 조명 조건 결정 서버 시스템(230)은 결정된 환경 맵(350)의 픽셀을 더 리파인하여, 장면(220)의 조명 조건을 보다 정확하게 결정할 수 있다.

[0047] 리파인 네트워크(REFINEMENT NETWORK)

[0048] 일부 구현에서, 조명 조건 결정 서버 시스템(230)은 환경 맵(350)의 각 픽셀을 복수의 서브-픽셀로 리파인하고 복수의 리파인된 서브 픽셀의 각각의 값을 결정함으로써 장면(220)의 조명 조건을 더 정확하게 추정할 수 있다. 예를 들어, 일부 구현에서, 조명 조건 결정 서버 시스템(230)은 쿼드 트리에 기반하여 환경 맵(350)의 픽셀을 리파인할 수 있다. 즉, 조명 조건 결정 서버 시스템(230)은 환경 맵(350)의 각 픽셀을 4 개의 서브 픽셀로 리파인하고 그에 따라 4 개의 서브 픽셀의 값을 결정할 수 있다. 일부 다른 구현에서, 조명 조건 결정 서버 시스템(230)은 또한 다른 방식으로 환경 맵(350)의 픽셀을 분할(divide)할 수 있다. 예를 들어, 각 픽셀은 9 개의 서브 픽셀, 16 개의 서브 픽셀 등으로 분할될 수 있다.

[0049] 도 3에 도시된 바와 같이, 일부 구현에서, 장면(200)의 조명 조건을 보다 정확하게 결정하기 위해, 환경 맵(350) 및 장면(220)이 방사 조도 맵 생성 모듈(360)에 입력될 수 있다. 일부 구현에서 환경 맵(350) 내의 픽셀(예를 들어, 0이 아닌 방사 조도 세기를 갖는 픽셀)에 대해, 방사 조도 맵 생성 모듈(360)은 픽셀과 연관된 제2 세트의 방사 조도 맵(370)을 생성할 수 있다. 제2 세트의 방사 조도 맵(370) 및 입력 이미지(201)는 제2 트레이닝된 신경망(380)(본 명세서에서 "리파인 네트워크"라고도 지칭됨)에 입력될 수 있다. 예를 들어, 제2 신경망(380)은 컨볼루션 신경망(CNN)일 수 있으며, 이는 환경 맵(350)의 픽셀이 제2 세트의 방사 조도 맵(370) 및 입력 이미지(201)에 기반하여 복수의 서브 픽셀로 분할되어야 하는지 여부를 결정하고 복수의 서브 픽셀의 각각의 값을 결정하도록 트레이닝될 수 있다. 이러한 방식으로, 조명 조건 결정 서버 시스템(230)은 환경 맵(350)에 대응하는 리파인된 환경 맵(240)을 생성할 수 있다. CNN이 여기서 제2 신경망(380)의 예로서 제공되지만, 본 명세서에 설명된 주제의 구현은 다른 타입의 신경망에 적용할 수 있으며, 본 명세서에 설명된 주제의 범위는 이 측면에서 제한되지 않는다는 것을 이해해야 한다.

[0050] $4 \times 4 \times 3$ 픽셀의 해상도로 환경 맵(350)을 표현하기 위해 L^2 가 사용되고, 쿼드 트리를 기반으로 환경 맵(350)을 리파인함으로써 획득되는 환경 맵(240)(예를 들어, 해상도가 $8 \times 8 \times 3$ 픽셀임)을 표현하기 위해 L^3 이 사용된다고 가정한다. 환경 맵(350)에서 u 번째 열과 v 번째 라인에 위치한 픽셀이 $L^2(u, v)$ 로 표현될 수 있다고 가정하면, 환경 맵(240)의 픽셀에 해당하는 4 개의 서브 픽셀은 다음과 같이 표현될 수 있다: $L^3(2u, 2v)$, $L^3(2u+1, 2v)$, $L^3(2u, 2v+1)$ 및 $L^3(2u+1, 2v+1)$.

[0051] 일부 구현에서, 장면(220)의 조명 조건을 가능한 한 정확하게 결정하기 위해, 환경 맵(350)의 픽셀 $L^2(u, v)$ 에 대해, 방사 조도 맵 생성 모듈(360)은 리파인 네트워크(380)의 입력으로서 다음의 더 미세한 레벨에서 환경 맵(240)의 모든 픽셀에 대한 각각의 방사 조도 맵을 생성할 수 있다. 그러나, 이러한 방식은 아마도 방사 조도 맵의 수를 상당히 증가시키고 리파인 네트워크(380)와 관련된 계산 및 저장 비용에서 현저한 증가를 초래할 것이

다.

[0052] 문제를 해결하기 위해, 일부 구현에서, 환경 맵(350)의 픽셀 $L^2(u, v)$ 에 대해, 방사 조도 맵 생성 모듈(360)은 리파인 네트워크(380)의 입력으로서 미리 결정된 수의 방사 조도 맵을 생성할 수 있다. 예를 들어, 환경 맵(350) 내의 픽셀 $L^2(u, v)$ 에 대해, 방사 조도 맵 생성 모듈(360)은 픽셀 $L^2(u, v)$ 와 연관된 13 개의 방사 조도 맵을 생성할 수 있다. 예를 들어, 13 개의 방사 조도 맵은 픽셀 $L^2(u, v)$ 로부터 분할된 4 개의 서브 픽셀 $L^3(2u, 2v)$, $L^3(2u+1, 2v)$, $L^3(2u, 2v+1)$ 및 $L^3(2u+1, 2v+1)$ 에 해당하는 4 개의 방사 조도 맵; 픽셀 $L^2(u, v)$ 에 인접한 8 개의 픽셀(즉, $L^2(u-1, v-1)$, $L^2(u-1, v)$, $L^2(u-1, v+1)$, $L^2(u, v-1)$, $L^2(u, v+1)$, $L^2(u+1, v-1)$, $L^2(u+1, v)$ 및 $L^2(u+1, v+1)$)에 대응하는 8 개의 방사 조도 맵; 및 환경 맵(350)에서 픽셀 $L^2(u, v)$ 및 인접하는 8 개의 픽셀을 제외한 나머지 픽셀의 조합에 대응하는 방사 조도 맵을 포함할 수 있다.

[0053] 제2 세트의 방사 조도 맵(370)은 위에서 논의된 바와 같이 수학적 식 1 또는 2에 기반하여 생성될 수 있다. 구체적으로, 방사 조도 맵 생성 모듈(360)은 전술한 수학적 식 1에 기반하여 픽셀 $L^2(u, v)$ 로부터 분할된 4 개의 서브 픽셀 $L^3(2u, 2v)$, $L^3(2u+1, 2v)$, $L^3(2u, 2v+1)$ 및 $L^3(2u+1, 2v+1)$ 에 대응하는 4 개의 방사 조도 맵을 각각 생성할 수 있다. 마찬가지로, 방사 조도 맵 생성 모듈(360)은 위의 수학적 식 1에 기반하여 픽셀 $L^2(u, v)$ 에 인접한 8 개의 픽셀(즉, $L^2(u-1, v-1)$, $L^2(u-1, v)$, $L^2(u-1, v+1)$, $L^2(u, v-1)$, $L^2(u, v+1)$, $L^2(u+1, v-1)$, $L^2(u+1, v)$ 및 $L^2(u+1, v+1)$)에 대응하는 8 개의 방사 조도 맵을 각각 생성할 수 있다. 또한, 방사 조도 맵 생성 모듈(360)은 상기 수학적 식 2에 기반하여 환경 맵(350)에서 픽셀 $L^2(u, v)$ 및 인접한 8 개의 픽셀을 제외한 나머지 픽셀의 조합에 대응하는 방사 조도 맵을 생성할 수 있다. 위에서 언급된 13 개의 방사 조도 맵은 리파인 네트워크(380)의 입력으로서 작용하여, 리파인된 환경 맵(240)을 생성하는데 사용될 수 있다.

[0054] 도 5는 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 리파인 네트워크(380)의 동작 다이어그램을 예시한다. 도 5는 입력 이미지(201) 및 환경 맵(350)의 픽셀(540)(본 명세서에서 또한 "제1 픽셀"이라고도 지칭됨)으로부터 생성된 제2 세트의 방사 조도 맵(370)을 도시한다. 도 5에 도시된 바와 같이, 일부 구현에서, 도 3에 도시된 바와 같이 방사 조도 맵 생성 모듈(360)에 의해 생성된 제2 세트의 방사 조도 맵(370)은 픽셀(540)로부터 분할된 4 개의 서브-픽셀에 대응하는 4 개의 방사 조도 맵(510-1... 510-4), 픽셀(540)에 인접한 8 개의 픽셀에 대응하는 8 개의 방사 조도 맵(520-1... 520-8), 및 환경(350)에서 픽셀(540) 및 8 개의 인접한 픽셀 이외의 나머지 픽셀의 조합에 대응하는 방사 조도 맵(530)을 포함할 수 있다. 리파인 네트워크(380)는 제2 세트의 방사 조도 맵(370) 및 입력 이미지(201)에 기반하여, 픽셀(540)이 분할되어야 하는지 여부를 결정할 수 있다. 리파인 네트워크(380)가 픽셀(540)이 분할되어야 한다고 결정하면, 리파인 네트워크(380)는 픽셀(540)로부터 분할된 4 개의 서브 픽셀의 각각의 값을 더 결정할 수 있다. 이러한 방식으로, 환경 맵(350)에 대응하는 리파인된 환경 맵(240)이 생성될 수 있다.

[0055] 일부 구현에서, 리파인 네트워크(380)의 내부 구조는 초기화 네트워크(340)의 내부 구조와 유사할 수 있다. 구체적으로, 입력 이미지(201) 및 각 방사 조도 맵(370)은 상이한 인코더에 의해 개별적으로 처리될 수 있다. 제2 세트의 방사 조도 맵(370) 및 입력 이미지(201)가 개별 인코더에 의해 처리된 후, 추출된 특징이 연결되어 또 다른 3-계층 인코더에 공급될 수 있다. 마지막으로, 출력 결과(즉, 픽셀이 분할되어야 하는지 여부 및 픽셀로부터 분할된 서브 픽셀의 각 값)는 완전 연결(FC) 계층 및 이어지는 3-계층 인코더로부터 획득될 수 있다. 특히, 리파인 네트워크(380)에서, 4 개의 방사 조도 맵(510-1... 510-4)에 대한 4 개의 인코더가 가중치를 공유할 수 있다. 마찬가지로, 8 개의 방사 조도 맵(520-1... 520-8)에 대한 8 개의 인코더가 가중치를 공유할 수 있다. 리파인 네트워크(380)는 픽셀이 분할되어야 하는지 여부에 대한 결정과 관련하여, 분할되어야 할 확률과 분할되지 않아야 할 확률을 출력할 수 있다. 그런 다음, 소프트-맥스 계층이 이들을 최종 결정으로 변환할 수 있다. 4 개의 서브 픽셀의 값의 경우 방사 조도 세기 간에 콘트라스트가 더 강할 수 있다. 일부 구현에서, 출력 값의 범위는 복수의 간격으로 균일하게 분산될 수 있고, 각 서브 픽셀의 값은 이산화된 다항 분류로 출력될 수 있다.

[0056] 일부 구현에서, 리파인 네트워크(380)는 구성된 데이터 세트를 사용하여 트레이닝될 수 있다. 예를 들어, 3D 객체, 및 3D 객체와 조명 조건 세트로부터 무작위로 선택된 조명 조건을 사용하여 이미지가 렌더링될 수 있다. 무작위로 선택된 각 조명 조건에 대해, 사전 정의된 정책을 사용하여 최적의 크로스 이미지 표현을 결정할 수 있다. 그 후, 리파인 네트워크(380)는 입력 이미지로부터 최적의 크로스 이미지 표현으로의 매핑을 근사하도록 트

레이닝될 수 있다. 일부 구현들에서, 예를 들어, 교차 엔트로피 손실 함수(cross-entropy loss function)는 분할 결정 및 서브-픽셀 값 모두에 대해 리파인 네트워크(380)를 트레이닝하는 데 사용될 수 있다.

[0057] 도 3은 2 개의 레벨의 신경망(즉, 초기화 네트워크(340) 및 리파인 네트워크(380))을 갖는 조명 조건 결정 서브 시스템(230)의 예만을 도시하지만, 본 명세서에서 설명된 주제의 구현은 더 적거나 더 많은 레벨의 신경망을 사용하는 조명 조건의 추정에 적용 가능하다는 것을 이해해야 한다. 예를 들어, 일부 구현에서, 계산 비용을 줄이기 위해 오직 한 레벨의 신경망(예를 들어, 초기화 네트워크(340))이 사용될 수 있다. 대안적으로, 일부 다른 구현에서, 더 정확한 결과를 얻기 위해 더 많은 레벨의 리파인 네트워크(예를 들어, 2 개 이상의 레벨의 리파인 네트워크)가 사용될 수 있다.

[0058] 도 6은 본 명세서에 설명된 주제의 일부 구현에 따라 다중 레벨의 신경망(예를 들어, 초기화 네트워크 및 2 개의 레벨의 리파인 네트워크)을 사용하여 장면의 조명 조건을 점진적으로 결정하는 다이어그램을 도시한다. 도 6은 초기화 네트워크에 의해 결정된 환경 맵(610)을 도시한다. 제1 레벨의 리파인 네트워크는 환경 맵(610)의 픽셀(601, 602... 605) 각각에 대해, 픽셀이 더 분할되어야 하는지 여부를 결정하고, 분할되도록 결정되는 픽셀(601, 603 및 605) 각각에 대해, 픽셀로부터 분할된 4 개의 서브 픽셀의 각각의 값을 결정한다. 제2 레벨의 리파인 네트워크는 제1 레벨의 리파인 네트워크로부터 출력된 환경 맵의 픽셀(606, 607... 611) 각각에 대해, 픽셀이 더 분할되어야 하는지 여부를 결정하고, 분할되도록 결정되는 픽셀(606, 607 및 610) 각각에 대해, 픽셀로부터 분할된 4 개의 서브 픽셀의 각각의 값을 결정한다. 이러한 방식으로 최종 환경 맵(620)(예를 들어, 16×16×3 픽셀의 해상도)이 획득될 수 있다.

[0059] 본 명세서에서 설명된 주제의 구현에 따른 조명 추정 솔루션은 위에서 설명되었다. 이상의 설명으로부터 본 명세서에서 설명하는 주제에 따른 조명 추정 솔루션은 장면에서 실제 객체의 재질 속성에 대한 어떠한 가정을 하지 않고도 입력 이미지 및 입력 이미지에 대한 음영 및 그림자 정보를 기반으로 다중 레벨의 신경망을 이용하여 장면의 조명 조건을 점진적으로 추정할 수 있음을 알 수 있다. 표준 광 베이스 세트의 선형 조합으로 장면의 광원을 모델링하고 표준 광 베이스 세트로부터 렌더링된 방사 조도 맵 세트를 신경망의 입력으로서 활용함으로써, 솔루션은 조명 추정의 복잡성을 크게 단순화하고 결과의 정확도를 향상시킬 수 있다. 또한 다중 트리 기반의 점진적 추정을 활용하여, 이 솔루션은 신경망의 각 레벨의 트레이닝 및 계산 비용을 절감하고 그 견고성을 향상시킬 수 있다. 솔루션에 따라 획득된 조명 조건을 사용하여 렌더링된 가상 객체는 실제 객체와 일치하는 사실적인 효과를 나타낼 수 있다.

[0060] 예시적인 프로세스

[0061] 도 7은 본 명세서에 설명된 주제의 일부 구현에 따른 조명 추정을 위한 프로세스(700)의 흐름도를 도시한다. 프로세스(700)는 예를 들어, 도 1에 도시된 바와 같이 이미지 처리 모듈(122)에 의해 구현될 수 있다. 프로세스(700)는 도시되지 않은 추가 블록을 더 포함하고/하거나 도시된 블록을 생략할 수 있음을 이해해야 한다. 본 명세서에 설명된 주제의 범위는 이 측면에서 제한되지 않는다.

[0062] 블록(710)에서, 이미지 처리 모듈(122)은 실제 객체에 관한 입력 이미지 및 입력 이미지에 대응하는 깊이 맵을 획득한다.

[0063] 블록(720)에서, 이미지 처리 모듈(122)은 깊이 맵에 기반하여 입력 이미지의 장면을 결정한다.

[0064] 블록(730)에서, 이미지 처리 모듈(122)은 장면에 기반하여 장면에서의 광원에 의해 야기되는 실제 객체에 대한 음영 및 그림자 정보를 결정한다.

[0065] 블록(740)에서, 이미지 처리 모듈(122)은 입력 이미지 및 음영과 그림자 정보에 기반하여 광원에 의해 야기되는 장면에서의 조명 조건을 결정한다.

[0066] 일부 구현에서, 프로세스(700)는 장면 및 조명 조건에 기반하여 실제 객체 및 가상 객체를 포함하고 가상 객체를 입력 이미지의 장면 내로 합성하는 출력 이미지를 생성하는 단계를 더 포함한다.

[0067] 일부 구현에서, 장면은 실제 객체의 배치를 위한 표면을 포함한다. 장면을 결정하는 단계는 깊이 맵을 포인트 클라우드로 변환하는 단계; 포인트 클라우드에서 표면을 나타내는 복수의 포인트를 평면 내로 피팅하는 단계; 및 상기 평면 위에 위치한 포인트 클라우드의 복수의 포인트를 3 차원 메시로 재구성하는 단계를 포함하고, 평면 및 3 차원 메시는 입력 이미지의 장면을 공동으로 표현한다.

[0068] 일부 구현에서, 광원은 미리 결정된 모양의 이미지에 의해 표현되고, 이미지는 광원의 복수 부분에 대응하는 복수의 픽셀을 포함하고, 복수의 픽셀의 각각의 픽셀 값은 광원의 복수 부분의 각각의 방사 조도 세기를

나타낸다. 일부 구현들에서, 음영 및 그림자 정보를 결정하는 단계는: 이미지 내의 복수의 픽셀들의 각각의 픽셀 값을 미리 결정된 값으로 초기화하는 단계; 및 장면 및 초기화된 이미지에 기반하여 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵을 생성하는 단계를 포함하며, 여기서 각각의 방사 조도 맵은 광원의 대응 부분에 의해 야기된 실제 객체에 대한 대응하는 음영 및 그림자 정보를 나타낸다. 일부 구현에서, 장면의 조명 조건을 결정하는 단계는, 입력 이미지 및 제1 세트의 방사 조도 맵에 기반하여 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계를 포함한다.

- [0069] 일부 구현에서, 광원을 나타내는 이미지는 크로스 모양의 이미지이다.
- [0070] 일부 구현에서, 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는, 입력 이미지 및 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 신경망을 사용하여 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계를 포함한다.
- [0071] 일부 구현들에서, 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는, 입력 이미지 및 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 복수의 신경망을 사용하여 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 점진적으로 결정하는 단계를 포함한다.
- [0072] 일부 구현에서, 복수의 트레이닝된 신경망은 적어도 제1 신경망 및 제2 신경망을 포함한다. 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는, 입력 이미지 및 제1 세트의 방사 조도 맵에 기반하여, 복수의 픽셀에서 제1 픽셀의 제1 픽셀 값 및 제1 신경망을 사용하여 제1 픽셀에 인접하는 적어도 하나의 픽셀의 픽셀 값을 결정하는 단계; 제1 픽셀, 적어도 하나의 픽셀 및 장면에 기반하여, 제2 세트의 방사 조도 맵- 제2 세트의 방사 조도 맵은 제1 픽셀의 복수의 서브 픽셀, 적어도 하나의 픽셀 및 복수의 픽셀 중 제1 픽셀 및 적어도 하나의 픽셀이 아닌 픽셀들의 조합에 각각 대응함 - 을 생성하는 단계; 상기 입력 이미지 및 제2 세트의 방사 조도 맵에 기반하여, 제1 픽셀이 제2 신경망을 사용하여 분할되어야 하는지 여부를 결정하는 단계; 및 상기 제1 픽셀이 분할되지 않아야 한다는 결정에 응답하여, 제1 픽셀 값을 제1 픽셀의 최종 픽셀 값으로서 사용하는 단계를 포함한다.
- [0073] 일부 구현들에서, 이미지 내의 복수의 픽셀들의 각각의 픽셀 값을 결정하는 단계는, 제1 픽셀이 분할되어야 한다는 결정에 응답하여, 제1 픽셀을 복수의 서브-픽셀들로 분할하는 단계; 및 복수의 서브-픽셀들의 각각의 서브-픽셀 값들을 결정하는 단계를 포함한다.
- [0074] 예시적인 구현
- [0075] 본 명세서에 설명된 주제의 일부 예시적인 구현이 아래에 나열될 것이다.
- [0076] 제1 양태에서, 본 명세서에 설명된 주제는 전자 디바이스를 제공하고, 상기 전자 디바이스는 처리 유닛; 및 상기 처리 유닛에 결합되고 상기 처리 유닛에 의한 실행을 위한 명령어들을 저장하는 메모리를 포함한다. 상기 명령어들은 상기 처리 유닛에 의해 실행될 때 상기 디바이스로 하여금 동작(act)들을 수행하게 하고, 상기 동작들은: 실제 객체에 관한 입력 이미지 및 상기 입력 이미지에 대응하는 깊이 맵을 획득하는 동작; 상기 깊이 맵에 기반하여 상기 입력 이미지의 장면을 결정하는 동작; 상기 장면에 기반하여, 상기 장면의 광원에 의해 야기된 실제 객체의 음영 및 그림자 정보를 결정하는 동작; 및 상기 입력 이미지와 상기 음영 및 그림자 정보에 기반하여, 상기 광원에 의해 야기되는 상기 장면의 조명 조건을 결정하는 동작을 포함한다.
- [0077] 일부 구현에서, 상기 동작들은 상기 장면 및 조명 조건에 기반하여, 실제 객체 및 가상 객체를 포함하는 출력 이미지를 생성하는 동작을 더 포함하고, 상기 출력 이미지는 상기 가상 객체를 상기 입력 이미지의 장면 내로 합성한다.
- [0078] 일부 구현에서, 상기 장면은 상기 실제 객체의 배치를 위한 표면을 포함한다. 상기 장면을 결정하는 동작은: 상기 깊이 맵을 포인트 클라우드로 변환하는 동작; 상기 포인트 클라우드 내의 복수의 포인트 - 상기 복수의 포인트는 표면을 나타냄 - 를 평면 내로 피팅하는 동작; 및 상기 평면 위에 위치한 상기 포인트 클라우드 내의 복수의 포인트를 3 차원 메시로 재구성하는 동작을 포함하고, 상기 평면 및 상기 3차원 메시는 상기 입력 이미지의 장면을 공동으로 나타낸다.
- [0079] 일부 구현에서, 상기 광원은 미리 결정된 모양의 이미지로 표현되고, 상기 이미지는 상기 광원의 복수의 부분에 대응하는 복수의 픽셀을 포함하고, 상기 복수의 픽셀의 각각의 픽셀 값은 상기 광원의 복수의 부분의 각각의 방사 조도 세기를 나타낸다. 일부 구현에서, 상기 음영 및 그림자 정보를 결정하는 동작은: 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 미리 결정된 값으로 초기화하는 동작; 및 상기 장면 및 초기화된 이미지에 기반

하여 상기 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵 - 방사 조도 맵 각각은 광원의 대응 부분에 의해 야기되는 실제 객체에 대한 대응하는 음영 및 그림자 정보를 나타냄 - 을 생성하는 동작을 포함한다. 일부 구현에서, 상기 장면의 조명 조건을 결정하는 동작은: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작을 포함한다.

[0080] 일부 구현에서, 상기 광원을 나타내는 이미지는 크로스 모양의 이미지이다.

[0081] 일부 구현에서, 상기 이미지 내의 상기 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 신경망을 사용하여 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작을 포함한다.

[0082] 일부 구현에서, 상기 이미지 내의 상기 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 복수의 트레이닝된 신경망을 사용하여 점진적으로 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작을 포함한다.

[0083] 일부 구현에서, 상기 복수의 트레이닝된 신경망은 적어도 제1 신경망 및 제2 신경망을 포함한다. 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 복수의 픽셀 내의 제1 픽셀의 제1 픽셀 값 및 상기 제1 신경망을 사용하여 상기 제1 픽셀에 인접한 적어도 하나의 픽셀의 픽셀 값들을 결정하는 동작; 상기 제1 픽셀, 상기 적어도 하나의 픽셀 및 상기 장면에서 기반하여 제2 세트의 방사 조도 맵 - 상기 제2 세트의 방사 조도 맵은 상기 제1 픽셀의 복수의 서브-픽셀, 상기 적어도 하나의 픽셀 및 상기 복수의 픽셀 중 상기 제1 픽셀과 상기 적어도 하나의 픽셀이 아닌 픽셀들의 조합에 각각 대응함 - 을 생성하는 동작; 상기 입력 이미지 및 상기 제2 세트의 방사 조도 맵에 기반하여, 상기 제1 픽셀이 상기 제2 신경망을 사용하여 분할되어야 하는지 여부를 결정하는 동작; 및 상기 제1 픽셀이 분할되지 않아야 한다는 결정에 응답하여, 상기 제1 픽셀 값을 상기 제1 픽셀의 최종 픽셀 값으로서 사용하는 동작을 포함한다.

[0084] 일부 구현에서, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 동작은: 상기 제1 픽셀이 분할되어야 한다고 결정하는 것에 응답하여, 상기 제1 픽셀을 복수의 서브-픽셀로 분할하는 동작; 및 상기 복수의 서브-픽셀의 각각의 서브-픽셀 값을 결정하는 동작을 포함한다.

[0085] 제2 측면에서, 본 명세서에 기술된 주제는 컴퓨터 구현 방법을 제공한다. 상기 방법은, 실제 객체에 관한 입력 이미지 및 상기 입력 이미지에 대응하는 깊이 맵을 획득하는 단계; 상기 깊이 맵에 기반하여 상기 입력 이미지의 장면을 결정하는 단계; 상기 장면에서 기반하여, 상기 장면에서의 광원에 의해 야기되는 실제 객체에 대한 음영 및 그림자 정보를 결정하는 단계; 및 상기 입력 이미지와 상기 음영 및 그림자 정보에 기반하여, 상기 광원에 의해 야기되는 상기 장면의 조명 조건을 결정하는 단계를 포함한다.

[0086] 일부 구현에서, 상기 방법은 상기 장면 및 조명 조건에 기반하여, 실제 객체 및 가상 객체를 포함하는 출력 이미지를 생성하는 단계를 더 포함하고, 상기 출력 이미지는 상기 가상 객체를 상기 입력 이미지의 장면 내로 합성한다.

[0087] 일부 구현에서, 상기 장면은 상기 실제 객체의 배치를 위한 표면을 포함한다. 상기 장면을 결정하는 단계는: 상기 깊이 맵을 포인트 클라우드로 변환하는 단계; 상기 포인트 클라우드 내의 복수의 포인트 - 상기 복수의 포인트는 표면을 나타냄 - 를 평면 내로 피팅하는 단계; 및 상기 평면 위에 위치한 상기 포인트 클라우드 내의 복수의 포인트를 3 차원 메시로 재구성하는 단계를 포함하고, 상기 평면 및 상기 3차원 메시는 상기 입력 이미지의 장면을 공동으로 나타낸다.

[0088] 일부 구현에서, 상기 광원은 미리 결정된 모양의 이미지로 표현되고, 상기 이미지는 상기 광원의 복수의 부분에 대응하는 복수의 픽셀을 포함하고, 상기 복수의 픽셀의 각각의 픽셀 값은 상기 광원의 복수의 부분의 각각의 방사 조도 세기를 나타낸다. 일부 구현에서 상기 음영 및 그림자 정보를 결정하는 단계는: 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 미리 결정된 값으로 초기화하는 단계; 및 상기 장면 및 초기화된 이미지에 기반하여 상기 복수의 픽셀에 대응하는 제1 세트의 방사 조도 맵 - 방사 조도 맵 각각은 광원의 대응 부분에 의해 야기되는 실제 객체에 대한 대응하는 음영 및 그림자 정보를 나타냄 - 을 생성하는 단계를 포함한다. 일부 구현에서 상기 장면의 조명 조건을 결정하는 단계는: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계를 포함한다.

[0089] 일부 구현에서, 상기 광원을 나타내는 이미지는 크로스 모양의 이미지이다.

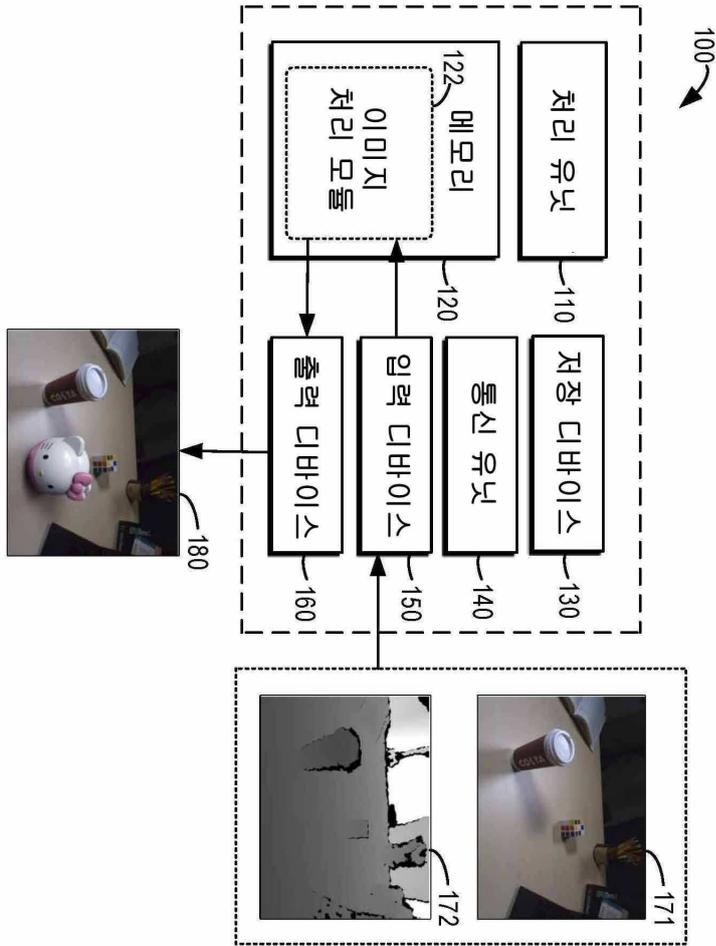
- [0090] 일부 구현에서, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 트레이닝된 신경망을 사용하여 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계를 포함한다.
- [0091] 일부 구현에서, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 복수의 트레이닝된 신경망을 사용하여 점진적으로 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계를 포함한다.
- [0092] 일부 구현에서, 상기 복수의 트레이닝된 신경망은 적어도 제1 신경망 및 제2 신경망을 포함한다. 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는: 상기 입력 이미지 및 상기 제1 세트의 방사 조도 맵에 기반하여, 상기 복수의 픽셀 내의 제1 픽셀의 제1 픽셀 값 및 상기 제1 신경망을 사용하여 상기 제1 픽셀에 인접한 적어도 하나의 픽셀의 픽셀 값들을 결정하는 단계; 상기 제1 픽셀, 상기 적어도 하나의 픽셀 및 상기 장면 에 기반하여 제2 세트의 방사 조도 맵 - 상기 제2 세트의 방사 조도 맵은 상기 제1 픽셀의 복수의 서브-픽셀, 상기 적어도 하나의 픽셀 및 상기 복수의 픽셀 중 상기 제1 픽셀과 상기 적어도 하나의 픽셀이 아닌 픽셀들의 조합에 각각 대응함 - 을 생성하는 단계; 상기 입력 이미지 및 상기 제2 세트의 방사 조도 맵에 기반하여, 상기 제1 픽셀이 상기 제2 신경망을 사용하여 분할되어야 하는지 여부를 결정하는 단계; 및 상기 제1 픽셀이 분할되지 않아야 한다는 결정에 응답하여, 상기 제1 픽셀 값을 상기 제1 픽셀의 최종 픽셀 값으로서 사용하는 단계를 포함한다.
- [0093] 일부 구현에서, 상기 이미지 내의 복수의 픽셀의 각각의 픽셀 값을 결정하는 단계는: 상기 제1 픽셀이 분할되어야 한다고 결정하는 것에 응답하여, 상기 제1 픽셀을 복수의 서브-픽셀로 분할하는 단계; 및 상기 복수의 서브-픽셀의 각각의 서브-픽셀 값을 결정하는 단계를 포함한다.
- [0094] 제3 측면에서, 본 명세서에 설명된 주제는 비일시적 컴퓨터 저장 매체에 유형적으로(tangibly) 저장된 컴퓨터 프로그램 제품을 제공하고, 컴퓨터 프로그램 제품은 디바이스에 의해 실행될 때 디바이스로 하여금 본 명세서에 설명된 주제의 제2 측면의 방법을 구현하게 하는 컴퓨터 실행 가능한 명령어들을 포함한다.
- [0095] 제4 측면에서, 본 명세서에 설명된 주제는 컴퓨터 실행 가능 명령어들이 저장된 컴퓨터 판독 가능 매체를 제공하며, 컴퓨터 실행 가능 명령어들은 디바이스에 의해 실행될 때 디바이스로 하여금 본 명세서에 설명된 주제의 제2 측면의 방법을 구현하게 한다.
- [0096] 본 명세서에서 설명된 기능은 적어도 부분적으로 하나 이상의 하드웨어 로직 컴포넌트에 의해 수행될 수 있다. 예를 들어 제한 없이 사용될 수 있는 하드웨어 로직 컴포넌트의 예시 타입은 FPGA(Field-Programmable Gate Array), ASIC(Application-specific Integrated Circuit), ASSP(Application-specific Standard Product), SOC(System-on-a-chip system), CPLD(Complex Programmable Logic Device) 등을 포함한다.
- [0097] 본 명세서에 설명된 주제의 방법을 수행하기 위한 프로그램 코드는 하나 이상의 프로그래밍 언어의 임의의 조합으로 작성될 수 있다. 이러한 프로그램 코드는 범용 컴퓨터, 특수 목적 컴퓨터 또는 기타 프로그램 가능한 데이터 처리 장치의 프로세서 또는 컨트롤러에 제공될 수 있으며, 따라서 프로그램 코드는 프로세서 또는 컨트롤러에 의해 실행될 때 순서도 및/또는 블록 다이어그램에 지정된 기능/동작이 구현되게 한다. 프로그램 코드는 전체적으로 머신에서, 부분적으로 머신에서, 독립형 소프트웨어 패키지로서, 부분적으로는 머신, 부분적으로는 원격 머신에서 또는 전체적으로 원격 머신 또는 서버에서 실행될 수 있다.
- [0098] 본 개시의 맥락에서, 머신 판독 가능 매체는 명령어 실행 시스템, 장치 또는 디바이스에 의해 또는 이와 관련하여 사용하기 위한 프로그램을 포함하거나 저장할 수 있는 임의의 유형의(tangible) 매체일 수 있다. 머신 판독 가능 매체는 머신 판독 가능 신호 매체 또는 머신 판독 가능 저장 매체일 수 있다. 머신 판독 가능 매체는 전자, 자기, 광학, 전자기, 적외선 또는 반도체 시스템, 장치 또는 디바이스, 또는 이들의 임의의 적절한 조합을 포함할 수 있지만 이에 제한되지는 않는다. 머신 판독 가능 저장 매체의 보다 구체적인 예는 하나 이상의 유선을 가진 전기 연결, 휴대용 컴퓨터 디스켓, 하드 디스크, RAM(Random Access Memory), ROM(Read-Only Memory), 소거 및 프로그램 가능 ROM(EPROM 또는 플래시 메모리), 광 섬유, 휴대용 CD-ROM, 광학 저장 디바이스, 자기 저장 디바이스 또는 이들의 임의의 적절한 조합을 포함할 것이다.
- [0099] 더욱이, 동작이 특정 순서로 묘사되어 있지만, 원하는 결과를 얻기 위하여 그러한 동작이 도시된 특정 순서 또는 순차적 순서로 수행되거나 모든 예시된 동작이 수행될 것을 요구하는 것으로 이해되어서는 안 된다. 특정 상황에서는 멀티 태스킹 및 병렬 처리가 유리할 수 있다. 마찬가지로, 몇 가지 특정 구현 세부 사항이 위의 논의에 포함되어 있지만, 이들은 본 명세서에 설명된 주제의 범위에 대한 제한으로 해석되어서는 안 되며, 특정 구

현에 특유할 수 있는 특징에 대한 설명으로서 해석되어야 한다. 개별 구현의 맥락에서 설명된 특정 특징은 단일 구현에서 조합하여 또한 구현될 수도 있다. 반대로, 단일 구현의 맥락에서 설명된 다양한 특징은 또한 여러 구현에서 개별적으로 또는 임의의 적절한 하위 조합으로 구현될 수 있다.

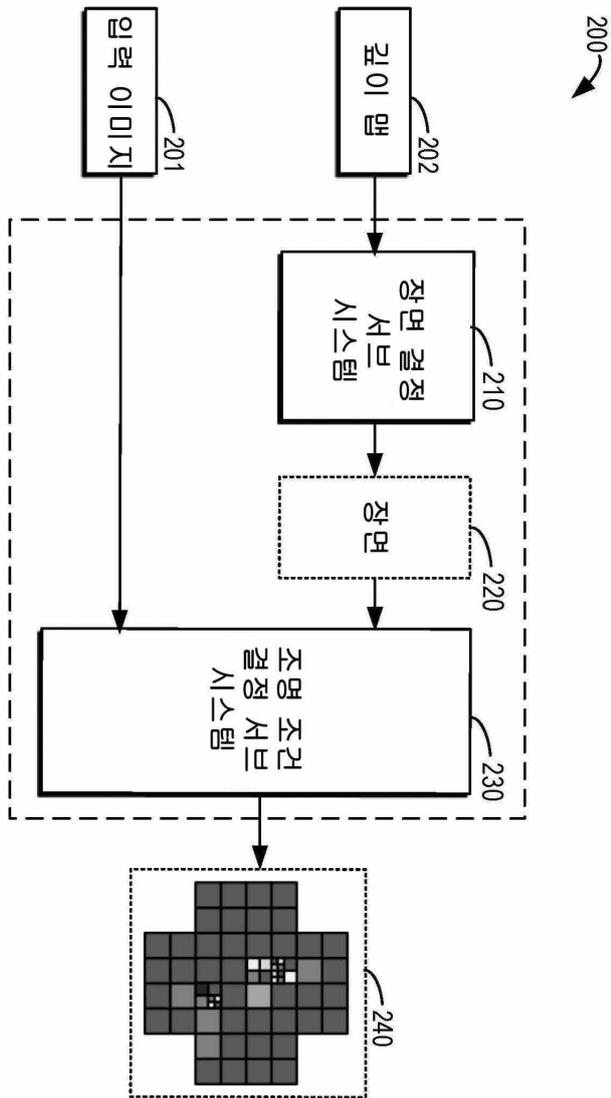
[0100] 주제가 구조적 특징 및/또는 방법론적 행위에 특유한 언어로 설명되었지만, 첨부된 청구범위에 명시된 주제가 반드시 위에서 설명된 특정 특징 또는 동작으로 제한되는 것은 아니라는 것을 이해해야 한다. 오히려, 위에서 설명된 특정 특징 및 동작은 청구범위를 구현하는 예시적인 형태로서 개시된다.

도면

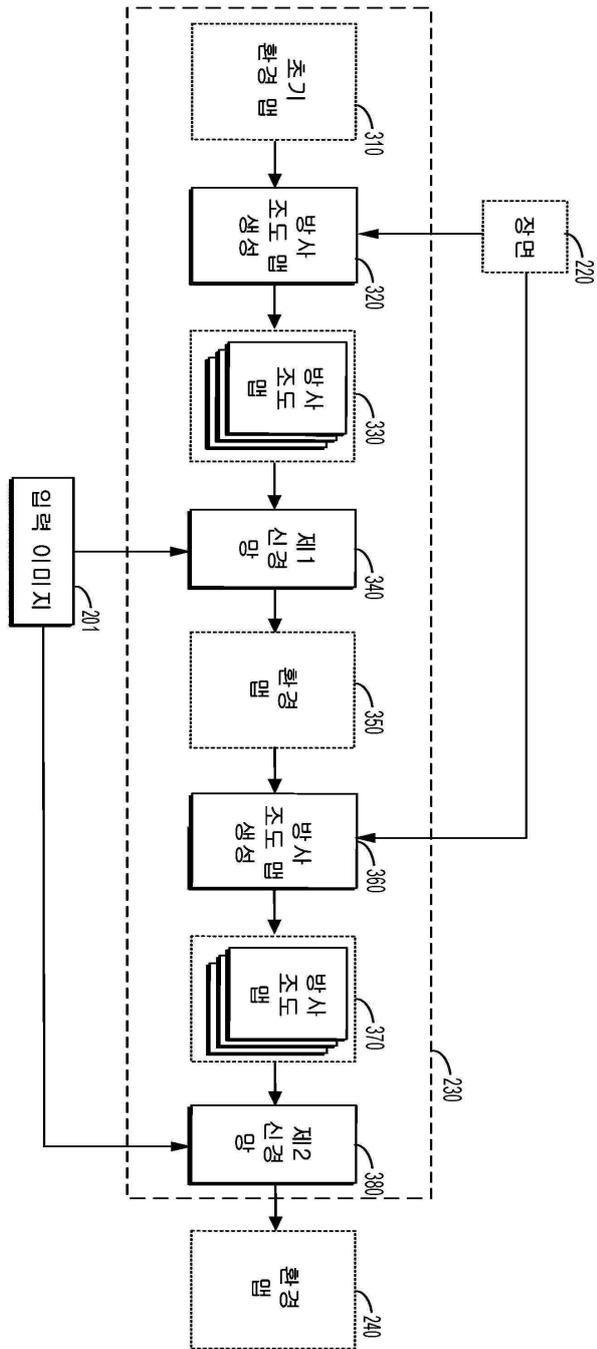
도면1



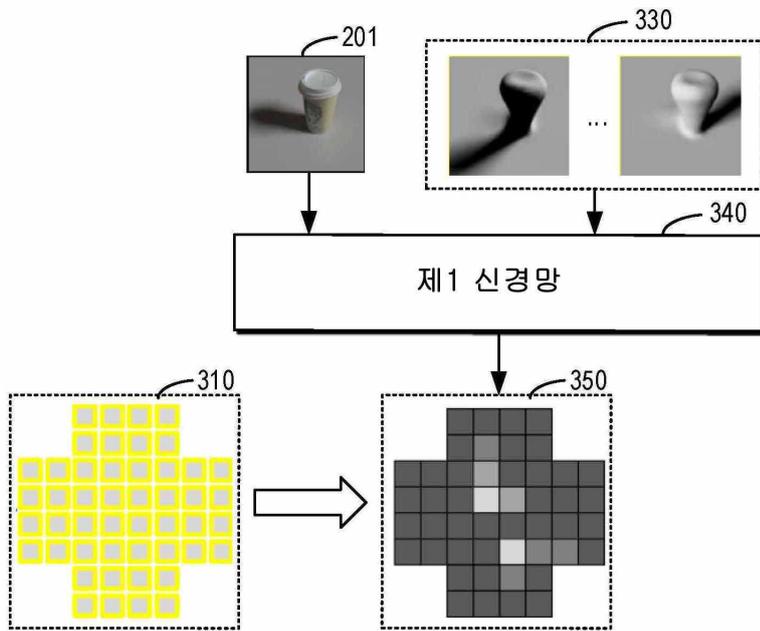
도면2



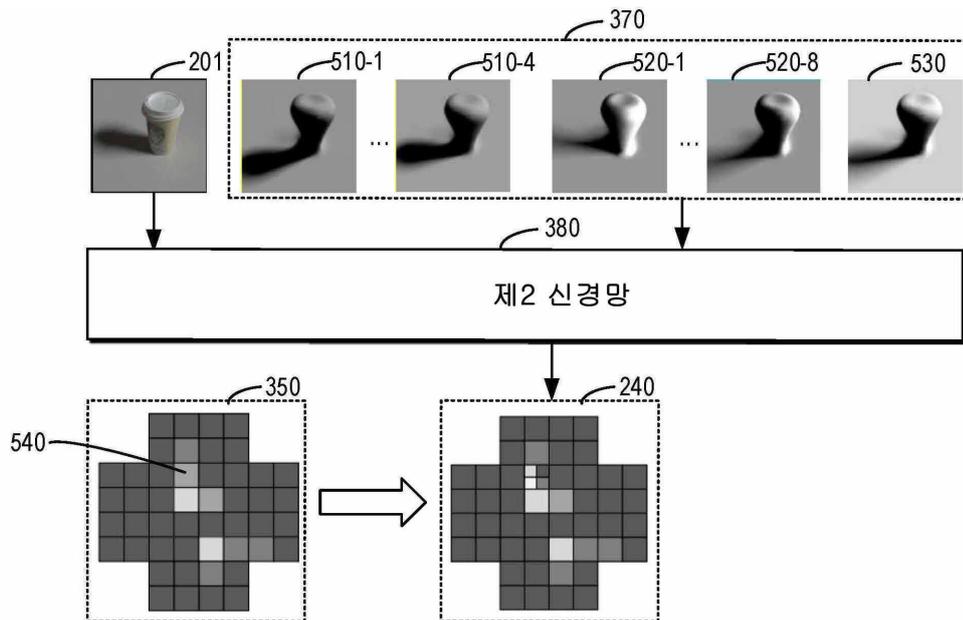
도면3



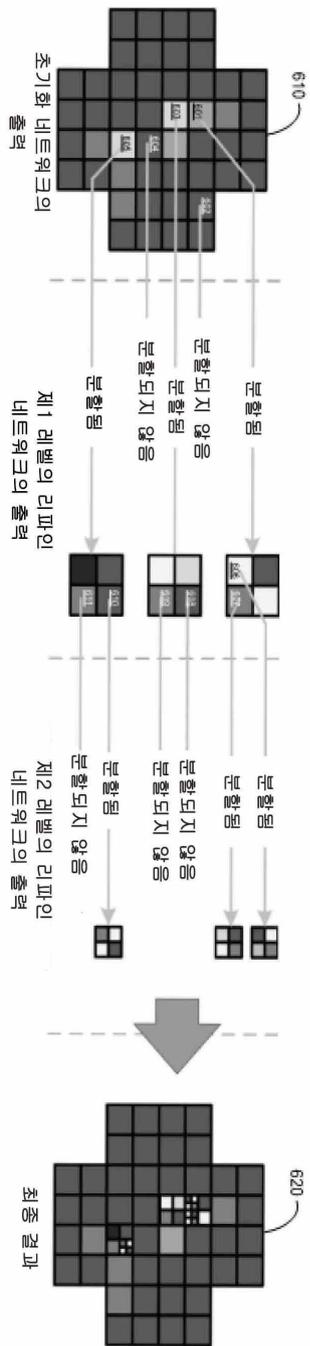
도면4



도면5



도면6



도면7

