



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0081620
(43) 공개일자 2025년06월05일

- | | |
|---|--|
| <p>(51) 국제특허분류(Int. Cl.)
G06V 10/84 (2022.01) G06V 10/764 (2022.01)
G06V 20/64 (2022.01)</p> <p>(52) CPC특허분류
G06V 10/84 (2023.08)
G06V 10/764 (2023.08)</p> <p>(21) 출원번호 10-2023-0170145
(22) 출원일자 2023년11월29일
심사청구일자 2023년11월29일</p> | <p>(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자
유현우
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(74) 대리인
특허법인더웨이브</p> |
|---|--|

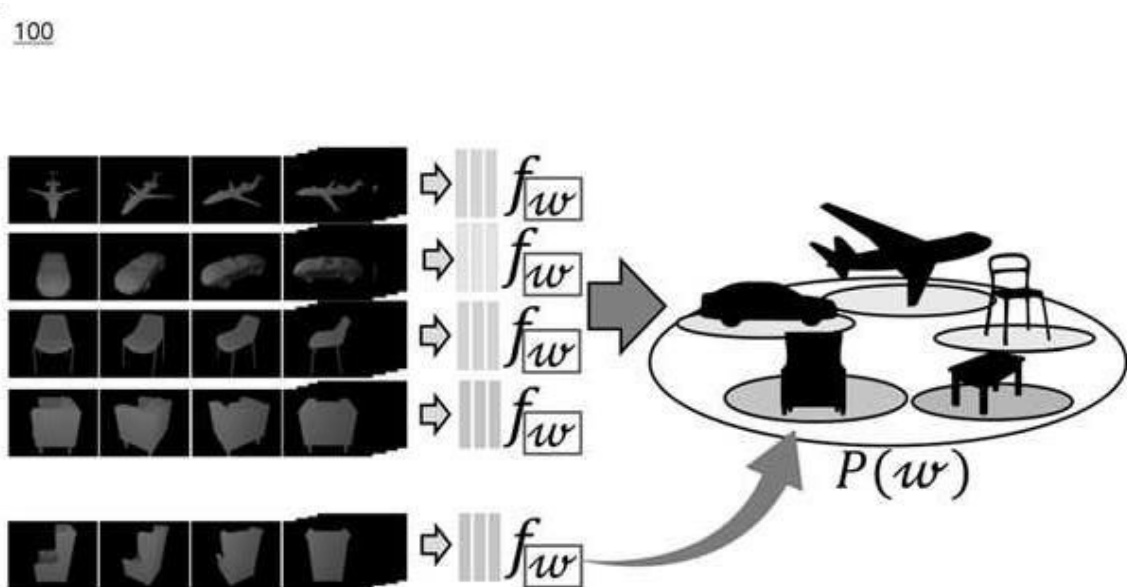
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 3D 오브젝트 인식을 위한 직접적 뉴럴 래디언스 필드 분류

(57) 요약

실시예에 따르면, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치는, 명령어들을 저장하도록 구성되는 메모리; 및 상기 명령어들을 실행함으로써: 3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하고, 상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하고, 상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하고, 상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하도록 구성되는 프로세서; 를 포함한다.

대표도



(52) CPC특허분류

G06V 20/64 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711172456
과제번호	2022R1F1A1066488
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	로봇 주행을 이용한 이중 환경에서의 자기지도학습
기 여 율	1/1
과제수행기관명	울산과학기술원
연구기간	2022.06.01 ~ 2023.05.31

명세서

청구범위

청구항 1

3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치에 있어서,

명령어들을 저장하도록 구성되는 메모리; 및

상기 명령어들을 실행함으로써:

3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하고,

상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하고,

상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하고,

상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하도록 구성되는 프로세서; 를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 2

제1항에 있어서,

상기 프로세서는, 상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하도록 구성되는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 3

제2항에 있어서,

상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 4

제2항에 있어서,

상기 프로세서는, 상기 프라이어 분포($p(W)$)에 따라 상기 네트워크 파라미터들(W)을 제한함과 동시에 상기 NeRF 네트워크들이 상기 관측 시퀀스들을 나타낼 확률을 학습시킴으로써 상기 네트워크 파라미터들(W)을 수렴시키도록 구성되는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 5

제4항에 있어서,

상기 프로세서는, 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통한 초기화에도 불구하고 파괴적 망각(catastrophic forgetting)을 회피하기 위해 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 상기 네트워크 파라미터들(W) 모두에 관하여 상기 NeRF 네트워크들을 트레이닝하도록 구성되는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 6

제1항에 있어서,

상기 프로세서는, 로봇의 온-라인 학습을 위한 경량 분류기를 이용하여 상기 네트워크 파라미터들을 기반으로 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하도록 구성되고,

상기 경량 분류기는 서포트 벡터 머신(SVM) 및 랜덤 포레스트를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치.

청구항 7

메모리에 저장되는 명령어들을 실행하는 프로세서에 의해 수행되는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법에 있어서,

3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하는 단계;

상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하는 단계;

상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하는 단계; 및

상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계; 를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 8

제7항에 있어서,

상기 프라이어 파라미터들로 초기화하는 단계는,

상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하는 단계; 를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 9

제8항에 있어서,

상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 10

제8항에 있어서,

상기 NeRF 네트워크들을 트레이닝하는 단계는,

상기 프라이어 분포($p(W)$)에 따라 상기 네트워크 파라미터들(W)을 제한함과 동시에 상기 NeRF 네트워크들이 상기 관측 시퀀스들을 나타낼 확률을 학습시킴으로써 상기 네트워크 파라미터들(W)을 수렴시키는 단계; 를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 11

제10항에 있어서,

상기 NeRF 네트워크들을 트레이닝하는 단계는,

상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통한 초기화에도 불구하고 파괴적 망각(catastrophic forgetting)을 회피하기 위해 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 상기 네트워크 파라미터들(W) 모두에 관하여 상기 NeRF 네트워크들을 트레이닝하는 단계; 를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 12

제7항에 있어서,

상기 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계는,

로봇의 온-라인 학습을 위한 경량 분류기를 이용하여 상기 네트워크 파라미터들을 기반으로 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계; 를 포함하고,

상기 경량 분류기는 서포트 벡터 머신(SVM) 및 랜덤 포레스트를 포함하는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법.

청구항 13

비일시적 컴퓨터 판독 가능한 기록 매체에 저장되는 컴퓨터 프로그램에 있어서,

상기 컴퓨터 프로그램의 명령어들은, 적어도 하나의 프로세서에 의해 실행될 때, 상기 적어도 하나의 프로세서가:

3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하는 동작;

상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하는 동작;

상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하는 동작; 및

상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 동작; 을 수행하게 하는, 비일시적 컴퓨터 판독 가능한 기록 매체에 저장되는 컴퓨터 프로그램.

청구항 14

제13항에 있어서,

상기 프라이어 파라미터들로 초기화하는 동작은,

상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하는 동작; 을 포함하는, 비일시적 컴퓨터 판독 가능한 기록 매체에 저장되는 컴퓨터 프로그램.

청구항 15

제14항에 있어서,

상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함하는, 비일시적 컴퓨터 판독 가능한 기록 매체에 저장되는 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 뉴럴 래디언스 필드(NeRF) 네트워크의 네트워크 파라미터를 직접적으로 이용하여 3D 오브젝트의 구조 및 형상을 분류하는 기법에 관한 것이다.

배경 기술

[0002] 근래에 포인트 클라우드 또는 3D 복셀 그리드로부터의 피쳐 추출 스킴을 통해 3D 구조 또는 오브젝트를 이해하기 위해 다양한 학습 기반의 기법들이 제안되어 왔다. 컴퓨터 비전의 분야에서 최근의 진전들 중 하나는 뉴럴 래디언스 필드(NeRF)이고, 이는 관측된 환경에 대한 높은 레벨의 디테일로 장면의 합성을 가능하게 한다. 로보틱스에서, NeRF는 3D 장면의 래디언스 필드를 뉴럴 네트워크에 의해 학습되는 연속 함수로 모델링하는 비선형 회귀를 이용한 3D 보간 기법이다. NeRF는 또한 이전의 접근법들과 비교할 때, 프리 트레이닝 데이터의 많은 양에 대한 요구 없이 신규한 3D 구조의 정교한 표현을 허용한다. 더욱이, NeRF는 실시간 트레이닝을 가능하게 하도록 확장되어 왔으며, 동시적 로컬화 및 맵핑(SLAM)과 같은 온라인 애플리케이션들에 대한 기법을 사용하는 것을 가능하게 한다. 예를 들면, 환경의 밀집한 3D 맵을 구성하기 위해 NeRF를 이용하도록 일부 연구들이 제안되었으며, 다른 연구들은 로컬화 태스크를 위한 NeRF 생성 장면들의 사용을 조사했다. NeRF 기반의 SLAM에 더하여, 일부 연구들은 하나의 관측 오브젝트에 대해 하나의 작은 NeRF 모델을 트레이닝함으로써 오브젝트 레벨의 맵핑을 수행한다.

[0003] 그러나, 대체로 3D 피쳐 표현을 요구하는 오브젝트/장소 인식 태스크를 위해 NeRF에 의해 학습되는 3D 공간 필드를 이용하는 것은 도전적이다. 좌표 기반의 뉴럴 표현을 이용함으로써 3D 공간을 이해하기 위해, 2D 렌더링 또는 3D 포인트 샘플링에 의한 다양한 관점들에 대해 네트워크 출력들을 직접적으로 활용하는 것이 필요하다. 네트워크 출력이 획득된다 하더라도, 관측된 오브젝트를 인식하기 위해서는, 포인트 클라우드에 대한 3D 표현, 복셀 그리드에 대한 3D 컨볼루션 뉴럴 네트워크(CNN), 및 옥트리(octree) 기반의 뉴럴 네트워크로부터 계층적 피쳐 추출 단계들이 여전히 필요하며, 이는 복잡한 네트워크 아키텍처 및 대규모의 3D 데이터셋을 필요로 한다.

발명의 내용

해결하려는 과제

[0004] 본 발명의 목적들 중 하나는, 복잡한 네트워크 아키텍처 및 대규모의 3D 데이터셋 없이 3D 오브젝트의 구조 및 형상을 분류하기 위해 NeRF 네트워크의 네트워크 파라미터를 직접적으로 이용하여 3D 오브젝트의 구조 및 형상을 분류하는 기법을 제공하는 것이다.

[0005] 본 발명의 기술적 목적은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또다른 기술적 과제들은 아래의 기재들로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0006] 실시예에 따르면, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치는, 명령어들을 저장하도록 구성되는 메모리; 및 상기 명령어들을 실행함으로써: 3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하고, 상기 프라이어 파라미터들에 의한 정규화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하고, 상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하고, 상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하도록 구성되는 프로세서;를 포함한다.

[0007] 실시예에 따르면, 상기 프로세서는, 상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하도록 구성된다.

[0008] 실시예에 따르면, 상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함한다.

[0009] 실시예에 따르면, 상기 프로세서는, 상기 프라이어 분포($p(W)$)에 따라 상기 네트워크 파라미터들(W)을 제한함과 동시에 상기 NeRF 네트워크들이 상기 관측 시퀀스들을 나타낼 확률을 학습시킴으로써 상기 네트워크 파라미터들

(W)을 수렴시키도록 구성된다.

- [0010] 실시예에 따르면, 상기 프로세서는, 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통한 초기화에도 불구하고 파괴적 망각(catastrophic forgetting)을 회피하기 위해 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 상기 네트워크 파라미터들(W) 모두에 관하여 상기 NeRF 네트워크들을 트레이닝하도록 구성된다.
- [0011] 실시예에 따르면, 상기 프로세서는, 로봇의 온-라인 학습을 위한 경량 분류기를 이용하여 상기 네트워크 파라미터들을 기반으로 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하도록 구성되고, 상기 경량 분류기는 서포트 벡터 머신(SVM) 및 랜덤 포레스트를 포함한다.
- [0012] 실시예에 따르면, 메모리에 저장되는 명령어들을 실행하는 프로세서에 의해 수행되는, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법은, 3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하는 단계; 상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하는 단계; 상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하는 단계; 및 상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계; 를 포함한다.
- [0013] 실시예에 따르면, 상기 프라이어 파라미터들로 초기화하는 단계는, 상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하는 단계; 를 포함한다.
- [0014] 실시예에 따르면, 상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함한다.
- [0015] 실시예에 따르면, 상기 NeRF 네트워크들을 트레이닝하는 단계는, 상기 프라이어 분포($p(W)$)에 따라 상기 네트워크 파라미터들(W)을 제한함과 동시에 상기 NeRF 네트워크들이 상기 관측 시퀀스들을 나타낼 확률을 학습시킴으로써 상기 네트워크 파라미터들(W)을 수렴시키는 단계; 를 포함한다.
- [0016] 실시예에 따르면, 상기 NeRF 네트워크들을 트레이닝하는 단계는, 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통한 초기화에도 불구하고 파괴적 망각(catastrophic forgetting)을 회피하기 위해 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 상기 네트워크 파라미터들(W) 모두에 관하여 상기 NeRF 네트워크들을 트레이닝하는 단계; 를 포함한다.
- [0017] 실시예에 따르면, 상기 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계는, 로봇의 온-라인 학습을 위한 경량 분류기를 이용하여 상기 네트워크 파라미터들을 기반으로 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 단계; 를 포함하고, 상기 경량 분류기는 서포트 벡터 머신(SVM) 및 랜덤 포레스트를 포함한다.
- [0018] 실시예에 따르면, 비일시적 컴퓨터 판독 가능한 기록 매체에 저장되는 컴퓨터 프로그램에 있어서, 상기 컴퓨터 프로그램의 명령어들은, 적어도 하나의 프로세서에 의해 실행될 때, 상기 적어도 하나의 프로세서가: 3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하는 동작; 상기 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 상기 관측 시퀀스들을 이용하여 상기 NeRF 네트워크들을 트레이닝하는 동작; 상기 NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 상기 NeRF 네트워크들의 네트워크 파라미터들을 산출하는 동작; 및 상기 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 상기 네트워크 파라미터들을 직접 이용하여 상기 3D 오브젝트의 카테고리 및 인스턴스를 분류하는 동작; 을 수행하게 한다.
- [0019] 실시예에 따르면, 상기 프라이어 파라미터들로 초기화하는 동작은, 상기 네트워크 파라미터들(W)이 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 상기 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하는 동작; 을 포함한다.
- [0020] 실시예에 따르면, 상기 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 상기 빈 3D 공간은 카메라로부터의 레이

가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함한다.

발명의 효과

- [0021] 본 발명의 실시예들에 따르면, 복잡한 네트워크 아키텍처 및 대규모의 3D 데이터셋 없이 3D 오브젝트의 구조 및 형상을 분류하기 위해 NeRF 네트워크의 네트워크 파라미터를 직접적으로 이용하여 3D 오브젝트의 구조 및 형상을 분류하는 기법이 제공될 수 있다.
- [0022] 본 발명의 실시예들에 따른 기술적 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또다른 효과들은 본 문서의 개시에 따라 당업자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0023] 도 1은 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법의 개요를 도시한다.
- 도 2는 일부 실시예에 따른 3D 오브젝트 인식을 위해 직접적 NeRF 분류를 수행하는 시스템을 도시한다.
- 도 3은 일부 실시예에 따른 관측 관점들 및 3D 뉴럴 표현들 간의 관계를 도시한다.
- 도 4는 일부 실시예에 따른 관측에 배경을 추가하는 방식을 도시한다.
- 도 5는 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치를 구성하는 요소들을 도시한다.
- 도 6은 일부 실시예에 따른 직접적 NeRF 분류 방법의 인스턴스 및 카테고리에 대한 분류 결과를 도시한다.
- 도 7은 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법을 구성하는 단계들을 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하에서는 도면을 참조하여 본 발명의 실시예들이 상세하게 설명될 것이다. 이하에서의 설명은 실시예들을 구체화하기 위한 것일 뿐, 본 발명에 따른 권리범위를 제한하거나 한정하기 위한 것은 아니다. 본 발명에 관한 기술 분야에서 통상의 지식을 가진 자가 발명의 상세한 설명 및 실시예들로부터 용이하게 유추할 수 있는 것은 본 발명에 따른 권리범위에 속하는 것으로 해석되어야 한다. 본 발명에 관한 기술 분야에서 통상의 지식을 가진 자에게 널리 알려져 있는 사항들에 대해서는 자세한 설명이 생략된다.
- [0025] 본 발명에서 사용되는 용어는 본 발명에 관한 기술 분야에서 널리 사용되는 일반적인 용어로 기재되었으나, 본 발명에서 사용되는 용어의 의미는 해당 분야에 종사하는 기술자의 의도, 새로운 기술의 출현, 심사기준 또는 판례 등에 따라 달라질 수 있다. 일부 용어는 출원인에 의해 임의로 선정될 수 있고, 이 경우 임의로 선정되는 용어의 의미가 상세하게 설명될 것이다. 본 발명에서 사용되는 용어는 단지 사전적 의미만이 아닌, 명세서의 전반적인 맥락을 반영하는 의미로 해석되어야 한다.
- [0026] 도 1은 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법의 개요를 도시한다.
- [0027] 도 1을 참조하면, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법의 개요(100)가 도시될 수 있다. 시퀀스로 획득되는 각 오브젝트 관측은 파라미터 W 를 갖는 하나의 NeRF로 표현된다. 대규모 데이터셋에 대한 어떠한 프리 트레이닝도 없이 오브젝트 인식 태스크를 위해 W 를 직접 활용하기 위해, 본 발명은 파라미터 공간에 대한 베이 지안 접근법에 기반한 방법을 제공한다(회색 화살표). 제안된 방법은 재방문된 오브젝트를 인식하기 위해 관측 데이터만을 사용하며, 랜덤 포레스트와 같은 간단한 분류기만을 사용한다(청색 화살표). 따라서, 제안된 방법은 온라인 학습에 적용될 수 있다.
- [0028] 본 문서에서, 네트워크들을 직접적으로 비교하고 오브젝트의 관측의 유사성을 찾음으로써 3D 오브젝트 인식을 수행하기 위한 방법이 제안된다. 뉴럴 네트워크를 비교하는 것은 큰 어려움을 가지며, 이는, 네트워크의 출력을 직접적 체크 없이, 다루기 어려운 함수들이 무엇을 나타내는지가 모호하기 때문이다. 이러한 어려움을 해소하기 위해, 본 발명은 네트워크 파라미터를 활용하기 위한 베이 지안 방법론을 도입한다. 네트워크 출력들을 하나씩 계산하는 대신, 본 발명은 네트워크 파라미터들을 분류를 위한 의미 있는 피쳐들로 만들 수 있고, 그에 따라 네트워크가 학습해온 것을 드러낼 수 있다. 제안된 방법의 개요(100)는 도 1에 도시된다.
- [0029] 로봇에 있어서, 문제되는 것은 타겟 환경 내에 존재하는 오브젝트의 인식이며, 세계의 넓은 지역에서 인식 가능한 모든 오브젝트를 인식하는 것이 아니다. 제안된 방법은 트레이닝 데이터를 요구하지 않으며, 오로지 타겟 환경에서의 관측 데이터를 요구할 뿐이다. 또한, 본 발명은 SVM 또는 랜덤 포레스트와 같은 전통적 분류기들을 이

용하여 극도로 빠른 분류 학습을 달성할 수 있으며, 그럼에도 데이터 중심의 접근법들에 비해 경쟁력 있는 결과를 달성할 수 있다. 이는 타겟 위치에 존재하는 오브젝트에 대한 온라인 학습을 가능하게 할 수 있으며, 이러한 점은 모바일 로봇 및 그 적응적 인식에 대해 필수적이다. 실험 결과들은 오브젝트 인식에 대한 벤치마크 데이터셋들에 대한 본 발명의 효과를 설명하며, 유망한 결과들을 보여주면서, 이러한 방향으로 미래 연구를 위한 길을 포장한다.

[0030] 도 2는 일부 실시예에 따른 3D 오브젝트 인식을 위해 직접적 NeRF 분류를 수행하는 시스템을 도시한다.

[0031] 도 2를 참조하면, 3D 오브젝트 인식을 위해 직접적 NeRF 분류를 수행하는 시스템(200)은 3D 오브젝트(210)를 관측하여 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 도출할 수 있다.

[0032] 시스템(200)은 모바일 로봇 시스템을 포함할 수 있다. 직접적 NeRF 분류 장치(220)는 모바일 로봇 시스템에 구비될 수 있다. 모바일 로봇은 3D 오브젝트(210)의 주위를 탐사하며 3D 오브젝트(210)를 관측할 수 있다. 예를 들면, 직접적 NeRF 분류 장치(220)는 3D 오브젝트(210)에 대한 관측 시퀀스를 생성할 수 있다.

[0033] 직접적 NeRF 분류 장치(220)는 3D 오브젝트(210)에 대해 생성되는 관측 시퀀스에 기초하여 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 도출할 수 있다. 3D 오브젝트(210)의 카테고리는 비행기, 자동차, 의자, 책상, 소파, 화장실, 컵, 피아노, 테이블 등을 포함할 수 있다. 동일한 카테고리에서 서로 다른 오브젝트는 서로 다른 인스턴스로 구별될 수 있다. 직접적 NeRF 분류 장치(220)는 NeRF 모델을 학습시킨 후, 학습 완료된 모델의 파라미터들을 직접 이용하여 카테고리 및 인스턴스(230)를 도출할 수 있다.

[0034] 도 3은 일부 실시예에 따른 관측 관점들 및 3D 뉴럴 표현들 간의 관계를 도시한다.

[0035] 도 3을 참조하면, 관측 관점들 및 3D 뉴럴 표현들 간의 관계를 도시하는 이미지(300)가 도시될 수 있다. 이미지(300)에서, 몇몇 관점들(좌측)에 의존하는 관측들을 통해 트레이닝되는, 뉴럴 필드로부터의 3D 구조(우측)가 도시된다. 관점들이 서로 상이하다하더라도, 그들의 3D 표현들은 유사한 3D 기하학적 피쳐들을 가지며, 이는 뉴럴 필드가 포인트 클라우드 또는 멀티 뷰 시퀀스와 같은 전통적 표현들과는 상이한 비선형적 회귀 스킴이기 때문이다.

[0036] A. 문제점 기술

[0037] 로봇이 기하학적 세그멘테이션 방법들을 이용하여 깊이 이미지 $I = \{I_k\}$ 와 같은 2.5D 시퀀스를 하나의 3D 오브젝트의 관측으로 획득한다는 점이 가정된다. 탐사 내지 향해 동안, 관측들을 이용하여 오브젝트 레벨의 이해를 위해 재방문된 인스턴스 또는 카테고리를 인식하는 것이 필요하다. 포인트 클라우드, TSDF, 또는 복셀화된 그리드와 같은 좌표 기반의 표현들은 3D 장면들 또는 오브젝트들을 표현하기 위해 널리 이용된다. 그러나, 이러한 좌표 기반의 표현들은 이산적이고, 희박하며, 제한적 해상도의 표현들을 가지며, 따라서 이들은 UGV 및 UAV의 관점들과 같은 다양한 관점들을 고려하기에 적합하지 않다. 이러한 난점으로 인해, 일반화된 피쳐들을 획득하기 위해, 막대한 3D 데이터셋에 대해 프리 트레이닝된 복잡한 구조를 갖는 뉴럴 네트워크를 사용하는 것이 불가피하다.

[0038] NeRF와 같은 좌표 기반의 뉴럴 표현들로, 3D 공간 내에서의 비선형적 피팅을 통해 고품질의, 물샐틈없는 3D 재구성을 달성하는 것이 용이해진다. 특정한 3D 장면의 뉴럴 표현은 비선형 보간법에 기초하는 3D 장면의 높은 수준의 3D 표현으로 취급될 수 있다. 그러므로, 이러한 뉴럴 표현은 도 3에 도시된 바와 같은 다양한 관측된 뷰들과 무관하게 밀도 레벨 3D 볼륨들을 지원한다. 이러한 이점들을 활용하기 위해, 본 발명은 네트워크 파라미터들(또는 가중치들)을 3D 오브젝트의 암시적 피쳐로 취급하고, 이하에서 오브젝트 인식을 위해 어떻게 이러한 피쳐들을 직접적으로 활용하는지를 보여줄 것이다.

[0039] B. 뉴럴 네트워크들의 가중치들

[0040] 뉴럴 네트워크를 드러내기 위해, 몇몇 네트워크들의 출력들을 직접 체크하는 것이 명시적인 방식으로 적용될 수 있다. 그러나, 입력 도메인 X 에서의 거의 모든 포인트들을 명시적으로 탐사하는 것이 불가피하며, 이는 비효율적이고 높은 계산적 부담을 요구한다. 직관적으로, 네트워크들의 다양한 정보는 그들의 구조 또는 파라미터와 강하게 관련된다. 이러한 경우, 네트워크들 및 트레이닝 데이터셋으로부터의 그들의 지식 사이의 연관들을 찾기 위해 네트워크 파라미터들을 확률적 랜덤 변수들로 취급함으로써 네트워크 파라미터들을 활용하는 것이 가질 수 있다. 뉴럴 네트워크가 네트워크 구조와 같은 다양한 피쳐들에 의해 결정되기 때문에, 네트워크 파라미터, 하이퍼 파라미터, 트레이닝의 시작에서의 초기 상태 등에 관하여, 이하에서 네트워크들 간의 유일한 차이는 네트워크 파라미터들의 값인 것으로 가정된다. 네트워크 구조 또는 액티베이션 함수와 같은 나머지의 다

른 팩터들 또한 동일하다.

[0041] 이러한 제한에도 불구하고, 네트워크가 동일한 트레이닝 데이터셋으로 트레이닝되더라도 네트워크 파라미터가 동일한 값으로 수렴할 것이라는 점이 여전히 보장되지 않는다. 이러한 현상의 주된 이유는 네트워크 파라미터들의 초기 상태들이 랜덤하게 할당된다는 것이다. 또한 트레이닝 동안 데이터셋의 오더가 랜덤하게 셔플되기 때문에 확률적 경사 하강법(stochastic gradient descent)을 위한 미니배치가 변동한다. 구체적으로, 동일한 학습 공간(데이터셋)을 이용하여 이미지들을 분류하기 위해 딥러닝 네트워크를 트레이닝하는 예시가 고려된다. 네트워크 파라미터들은 전형적으로 랜덤하게 초기화되며 최적화 알고리즘을 이용하여 트레이닝된다. 최적화 알고리즘이 수렴하면서, 동일한 학습 공간 내에서도 파라미터들의 상이한 세트의 결과가 될 수 있다. 따라서, 동일한 학습 공간이 사용되더라도, 랜덤 초기화 및 최적화 알고리즘 사용으로 인해 상이한 파라미터 값들이 획득될 수 있다. 다른 예시는, 상이한 하이퍼 파라미터 값들이 동일한 학습 공간 내에서 상이한 네트워크 파라미터 값들로 이어질 수 있다는 점이다. 하이퍼 파라미터들은 네트워크의 구조, 최적화 알고리즘 등에 영향을 주는 값들이다. 그러므로, 동일한 학습 공간을 가지고도, 상이한 하이퍼 파라미터 값들은 상이한 네트워크 파라미터 값들의 결과가 될 수 있다.

[0042] 간단한 예시로, 뉴럴 네트워크들에 의해 표현되는 함수 $f_{\mathbf{w}}$ 가 다음과 같이 서로 다른 K개의 함수들로 구성된다는 점이 가정된다.

[0043] [수식 1]

$$f_{\mathbf{w}} = f_{\mathbf{w}_{K-1}} \circ \dots \circ f_{\mathbf{w}_k} \circ \dots \circ f_{\mathbf{w}_0}$$

[0044] $\mathbf{w} = \{\mathbf{w}_k\}$ 는 네트워크의 파라미터이다. 여기서, $f_{\mathbf{w}_k}$ 는 비선형 액티베이션 함수 ψ 를 갖는 텐스 레이어(dense layer)임이 가정된다. 함수 $f_{\mathbf{w}_k}$ 는 다음과 같이 표현된다.

[0045] [수식 2]

$$f_{\mathbf{w}_k}(\mathbf{z}_k) = \psi(\mathbf{w}_k^T \mathbf{z}_k)$$

[0046] 여기서 $\mathbf{z}_k$ 는 k번째 히든 레이어의 히든 입력 벡터이다. 네트워크의 비일관성을 보여주기 위해, k번째 파라미터 $\mathbf{w}_k$ 에 임의의 논제로 값 α 가 곱해진다. ReLU 또는 누설 ReLU와 같은 액티베이션 함수들은 입력 벡터의 스케일을 보존하기 때문에, 즉, $\psi(\alpha y) = \alpha \psi(y)$ 이기 때문에, $f_{\mathbf{w}_k}(\alpha \mathbf{z}_k) = \alpha f_{\mathbf{w}_k}(\mathbf{z}_k)$ 이다. 따라서, (k+1)번째 파라미터 $\mathbf{w}_{k+1}$ 에 $1/\alpha$ 가 곱해지면 다음과 같다.

[0047] [수식 3]

$$\begin{aligned} f_{\frac{1}{\alpha}\mathbf{w}_{k+1}} \circ f_{\alpha\mathbf{w}_k}(\mathbf{z}_k) &= \psi\left(\frac{1}{\alpha}\mathbf{w}_{k+1}^T \psi\left(\alpha\mathbf{w}_k^T \mathbf{z}_k\right)\right) \\ &= \frac{1}{\alpha}\alpha\psi\left(\mathbf{w}_{k+1}^T \psi\left(\mathbf{w}_k^T \mathbf{z}_k\right)\right) \\ &= f_{\mathbf{w}_{k+1}} \circ f_{\mathbf{w}_k}(\mathbf{z}_k) \end{aligned}$$

[0048] 이는 스칼라 값들을 곱하기 전과 동일한 결과로 이어진다. 유사하게, $\mathbf{w}_k$ 의 j번째 행이 α 만큼 스케일되면, $\mathbf{w}_{k+1}$ 의 j번째 열을 $1/\alpha$ 만큼 스케일하는 것이 원본 출력을 획득하는 것의 결과가 될 것이다.

[0049] C. 암시적 분류 및 최대 사후 확률(MAP)

[0050] 본 문서의 주요 목적은 트레이닝 동안 네트워크가 학습한 것을 분석함으로써 3D 오브젝트를 분류하는 것이다. 그러나, 앞서 논의된 바와 같이, 동일한 입력에 대해 유사한 출력을 갖는 네트워크들을 고유한 것으로 고려하는 것은 가능하지 않다. 이는 파라미터들의 프라이어 분포들(prior distributions)은 일반적으로 고려되지 않기 때

문이다. 현존하는 접근법들에서의 대부분의 트레이닝 스킴들은 파라미터들에 대한 최대 우도 추정(MLE)에 기반한다. 그 결과, 프라이어 분포 $p(\mathbb{W})$ 와 같은, 파라미터 $\mathbb{W}$ 에 대한 제한이 없으며, 이는 그들의 도메인과 같은 파라미터들의 조건들을 제한하기 어렵게 만든다. 그러므로, 파라미터를 제한하기 위해, 본 발명은 프라이어 분포를 고려할 수 있고, 다음의 결합 확률을 이용하여 최대 사후 확률(MAP)을 수식화할 수 있다.

[0054] [수식 4]

$$p(\mathbf{x}, \mathcal{W}; \mathcal{I}) = p(\mathcal{W}) p(\mathbf{x} | \mathcal{W}; \mathcal{I})$$

[0055] L_1 또는 L_2 놈(norm)과 같은 전통적 파라미터 정칙화의 경우에서, 예를 들면, 파라미터의 프라이어 분포는 트레이닝 동안 라플라스 분포 및 다변량 등방성 가우시안 분포로 각각 가정된다. 그러나, 이러한 프라이어들은 파라미터 $\mathbb{W}$ 의 각 요소에 대해 특정한 도메인을 고려하지 않는 제로 평균을 가정한다. 다른 말로, 이러한 프라이어들은 파라미터 행렬 내의 모든 요소들($\mathbb{W} \in \mathbb{W}$ 의 행 또는 열 방식 벡터들)을 가정한다. 예를 들면, k 번째 레이어 $\mathbb{W}_k$ 의 모든 행들이 유사한 초기 포인트들로 시작하고 동일한 프라이어를 갖는 것이 가정된다. 최적의 $\mathbb{W}_k$ 가 특정

$\hat{\mathbf{w}}$
벡터들 $\{\}$ 을 가질 것을 필요로 하는 경우, 특히 f 의 출력이 최종 출력 벡터들의 모든 요소들을 합산함으로써

$\hat{\mathbf{w}}$
획득되는 스칼라인 경우, 어떤 행이 $\{\}$ 가 될 것인지가 랜덤하게 결정된다. 그러므로, 트레이닝이 완료되더라도, 무엇이 학습되었는지를 구별하는 것이 어려울 수 있고, 이는 $\mathbb{W}_k$ 의 각 행 $\mathbf{w}$ 에 대한 도메인이 언제나 일정할 수 없기 때문이다.

[0057] D. 프라이어 분포를 위한 프리 트레이닝된 가중치들

[0058] 네트워크가 학습한 것에 따라 분명하게 구별 가능한 파라미터들을 획득하기 위해, 특정한 프라이어들을 분명하게 정의하고 파라미터 공간에 대한 도메인들을 제한하는 것이 필요하다. 트레이닝 동안 파라미터의 도메인 및 초기 포인트 제한을 구현하기 위해, 파라미터에 대한 몇몇 정칙화 항목들이 도입될 수 있다. 먼저, 파라미터 $\mathbb{W}$ 에 대한 도메인 적응이 다음의 정칙화 손실 L 에서와 같이 도메인 분류기 및 도메인 적대적 트레이닝을 이용함으로써 채택될 수 있다.

[0059] [수식 5]

$$\mathcal{L} = - \sum_m^M \mathcal{D}(\mathcal{W}_m)$$

[0060] 여기서 D 는 f 와 동시에 트레이닝되는 도메인 분류기이다. 그러나, 이러한 접근을 구현하는 것은 실제로는 상당한 난점을 가지며, 이는 모든 M 개의 네트워크들이 트레이닝 동안 수식 5의 계산을 위해 능동적으로 관여해야 하기 때문이다.

[0062] 다른 방법은, 예를 들면, $\mathbb{W}_k$ 의 각 행이 서로 독립적임을 표시하는 $\mathbb{W}_k \sim N(\mathbf{I}, \Sigma)$ 와 같은 분포로부터 가중치들을 초기화함으로써, 네트워크 파라미터들에 프라이어를 직접적으로 할당하는 것을 포함한다. 그러나, 이러한 접근은 네트워크가 실제의 3D 구조와 필수적으로 관련되지 않는 프라이어 분포로부터 학습할 것이 강제되는 결과가 될 수 있다. 이러한 가우시안 프라이어를 이용하는 것은 학습 프로세스를 상당 수준 느리게 만들고 또는 심지어 실제에서의 수렴을 어렵게 만든다는 점이 발견된다. 뉴럴 래디언스 필드를 위한 3D 공간을 반영하는 적절한 프라이어 분포들을 찾기 위해, 본 발명은 파라미터에 대해 다음의 프라이어 분포를 사용할 것을 제안한다.

[0063] [수식 6]

$$p(\mathcal{W}) = \mathcal{N}(\mathcal{W}_{pre}, \mathbf{I})$$

[0064] $\mathbb{W}_{pre}$ 는 다음의 최대 우도 추정(MLE)를 푸는 것으로부터 획득된다.

[0066] [수식 7]

$$\mathcal{W}_{pre} = \underset{\mathcal{W}}{\operatorname{argmax}} \sum_{x \in \mathcal{X}} \log p(x | \mathcal{W}; \mathcal{I}_{pre})$$

[0067]

[0068] $\mathcal{I}_{pre}$ 는 어떠한 3D 구조물도 내부에 없는 빈 방과 같은 임의의 3D 공간을 표현하는 프리 트레이닝(pretraining) 데이터셋을 나타낸다. 간결함을 위해, 이하의 식들에서 x 의 합산은 생략된다. 최대 사후 확률(MAP)에 있어서, 수식 4 및 수식 6을 이용함으로써, 본 발명은 결국 프리 트레이닝된 파라미터들의 고려 하에 최적 가중치 $\mathcal{W}^*$ 를 다음과 같이 획득한다.

[0069] [수식 8]

$$\begin{aligned} \mathcal{W}^* &= \underset{\mathcal{W}}{\operatorname{argmax}} \log p(x, \mathcal{W}; \mathcal{I}) \\ &= \underset{\mathcal{W}}{\operatorname{argmax}} (\log p(\mathcal{W}) + \log p(x | \mathcal{W}; \mathcal{I})) \\ &= \underset{\mathcal{W}}{\operatorname{argmin}} (\|\mathcal{W} - \mathcal{W}_{pre}\|_2^2 - \log p(x | \mathcal{W}; \mathcal{I})) \end{aligned}$$

[0070]

[0071] 수식 8에서, 수식 5와는 달리, 정칙화 항이 $\mathcal{L} = \|\mathcal{W} - \mathcal{W}_{pre}\|_2^2$ 과 같이 정의된다. 수식 8에 의해, 네트워크는 파라미터 $\mathcal{W}$ 를 제한하기 위한 프라이어를 따름과 동시에 트레이닝 데이터 $\mathcal{I}$ 를 나타낼 확률을 학습할 수 있다. 정칙화 항 $\mathcal{L}$ 은 수식 7에서와 같이 프리 트레이닝 데이터셋 $\mathcal{I}_{pre}$ 를 암시적으로 표현하는 프리 트레이닝된 파라미터 $\mathcal{W}_{pre}$ 를 따르도록 네트워크 파라미터들 $\mathcal{W}$ 를 강제할 수 있다. 따라서 수식 8은 다음과 같이 제작성될 수 있다.

[0072] [수식 9]

$$\mathcal{W}^* = \underset{\mathcal{W}}{\operatorname{argmax}} (\log p(x | \mathcal{W}; \mathcal{I}_{pre}) + \log p(x | \mathcal{W}; \mathcal{I}))$$

[0073]

[0074] 여기서, 프리 트레이닝의 최대 우도 추정(MLE)에 대한 항목 $-\log p(x | \mathcal{W}; \mathcal{I}_{pre})$ 은 정칙화 항 $\mathcal{L}$ 을 표시한다. 결과적으로, 파라미터를 제한하는 것은 트레이닝 데이터셋 및 프리 트레이닝 데이터셋 모두를 동시에 사용하여 프라이어 분포를 따름으로써 달성될 수 있다. 실제에서는, 네트워크 파라미터 $\mathcal{W}$ 를 프리 트레이닝된 파라미터 $\mathcal{W}_{pre}$ 로 초기화하는 것이 빠른 수렴을 달성하는 것을 돕는다. 그러나, 정칙화 $\mathcal{L}$ 없이, $\mathcal{W}$ 는 파괴적 망각(catastrophic forgetting)으로 인해 제한될 수 없고, 따라서 본 발명은 $\mathcal{W}$ 가 $\mathcal{W}_{pre}$ 로 초기화되었음에도 수식 9를 사용한다.

[0075] 트레이닝 프로세스를 요약하면, 실제 트레이닝 동안, 본 발명은 프리 트레이닝된 파라미터 $\mathcal{W}_{pre}$ 로 초기화함으로써 네트워크의 프라이어 분포를 정렬한다. 그 다음 본 발명은 3D 장면의 우도를 학습하기 위해 트레이닝 데이터 $\mathcal{I}$ 에 대해 네트워크를 트레이닝하고, 네트워크는 동시에 프리 트레이닝 데이터 $\mathcal{I}_{pre}$ 를 학습함으로써 프라이어 분포를 따른다. 이러한 방식으로, 3D 공간 또는 오브젝트를 학습하는 네트워크의 파라미터 $\mathcal{W}$ 는 각 벡터에 대해 동일한 도메인 내로 제한된다. 결과적으로, 각 차원 또는 벡터를 따라 상이한 3D 공간들 또는 오브젝트들에 대해 트레이닝되는 네트워크들의 파라미터들을 비교하는 것이 의미 있게 되며, 이는 네트워크가 학습한 것을 분류하고 드러낼 수 있게 한다.

[0076] 도 4는 일부 실시예에 따른 관측에 배경을 추가하는 방식을 도시한다.

[0077] 도 4를 참조하면, 관측에 배경을 추가하는 방식을 도시하는 이미지(400)가 도시될 수 있다. 프리 트레이닝을 위해 사용되는 환경으로 관측의 배경을 변경함으로써, 타겟 오브젝트 및 프리 트레이닝 데이터 모두에 대한 레이들(rays)을 샘플링하여 단일의 트레이닝 세션에서의 최대 사후 확률(MAP)을 수행하는 것이 용이해진다.

[0078] 실제에서, 본 발명은 네트워크를 트레이닝하기 위해 수식 9를 활용한다. 수식 7에서 I_{pre} 를 선택하는 것은 트레이닝 수렴성 및 파라미터의 프라이어 모두에 대해 중요하다. 3D 오브젝트를 이해하는 태스크에 있어서, 내부에 어떤 오브젝트도 없는 빈 방이 프리 트레이닝 3D 장면으로 선택된다. 본 발명은 빈 방이 무한 지면 또는 완전한 빈 고악나과 같은 다른 3D 환경과 비교할 때 효과적이라는 점을 발견했다. NeRF의 비용 함수는 다음과 같다.

[0079] [수식 10]

$$\sigma = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))dt,$$

$$T(t) = \exp\left(-\int_{t_n}^t \sigma(r(s))ds\right)$$

[0080]

[0081] 본 발명의 경우에 있어서, RGB 항들은 버려진다. $T(t)$ 는 레이 r 을 따르는 누적된 트랜스미턴스를 표시한다. 즉, 뉴럴 래디언스 필드의 비용 함수를 계산하기 위해, 누적된 트랜스미턴스 $T(t)$ 를 적분함으로써 표현되는 강도(또는 밀도)를 달성하기 위해 카메라로부터의 레이가 3D 공간에서의 3D 구조의 임의의 표면을 터치해야 한다. 레이 가 어떤 표면도 만나지 못하고 그 레이에 따른 픽셀이 정보를 포함하지 않으면, 해당 픽셀은 트레이닝 동안 버려진다. 따라서, 강건한 프리 트레이닝 데이터를 달성하기 위해, 모든 레이들이 의미 있는 정보로 획득되도록 3D 빈 방과 같은 바운디드 3D 환경을 활용하는 것이 필요하다.

[0082] 본 발명의 경우, 3D 오브젝트들이 트레이닝 데이터로 사용된다. 수식 9에서 도식된 바와 같이, 네트워크가 프리 트레이닝된 모델에 의해 초기화되더라도 네트워크를 프리 트레이닝 데이터 및 트레이닝 데이터 모두에 대해 트레이닝하는 것이 필요하며, 이는 파괴적 망각이 네트워크가 프라이어 분포를 유지하는 것을 막기 때문이다. 수식 9에 의해, 네트워크는 주어진 3D 오브젝트 및 빈 방을 동시에 학습해야 한다. 따라서 본 발명은 도 4에 도식된 바와 같이 단순히 3D 오브젝트의 배경을 빈 방으로 변경함으로써 이러한 케이스를 구현하며, 그에 따라 프리 트레이닝 데이터(빈 방)의 레이들 및 트레이닝 데이터(3D 오브젝트)의 레이들이 한번에 수집되고 계산될 수 있다.

[0083] 본 발명이 바운디드 방을 활용함에도 불구하고, 여전히 트레이닝의 수렴성의 문제가 존재한다. 수식 9에서 최대 사후 확률(MAP)을 수행하기 위해 네트워크에 의해 표현되는 우도를 활용할 때, 타겟 분포, 우도, 및 프라이어가 상당한 비유사성을 보이는 경우, 트레이닝 프로세스가 수렴에 실패할 수 있다. 예를 들면, 프라이어 분포로서의 빈 방을 이해하기 위해, 네트워크는 방의 중심에서 제로 밀도를 추정해야 한다. 그러나, 빈 방과 비교할 때 반대의 상황을 나타내는, 방의 중심에 위치하는 3D 오브젝트의 학습을 시도할 때, 로컬 최적 상태(빈 방)을 벗어나는 것 및 바람직한 위치에서 3D 오브젝트를 생성하는 것이 어려울 수 있다. 본 발명은 빈 방으로 초기화된 트레이닝 프로세스가 가끔씩 수렴에 실패하고 3D 오브젝트의 생성에 실패하는 것을 확인했다. 그러므로, 본 발명은 방의 중심에서 신규한 오브젝트 형태의 용이한 생성을 허용하는 프라이어 분포를 규정하기 위해 방의 중심에 큰 구조물을 갖는 장면들을 트레이닝한다. 실제에서 본 발명은 안정적인 트레이닝 및 그 수렴을 위해 큰 구체 장면(large-sphere scene)을 사용한다.

[0084] 도 5는 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치를 구성하는 요소들을 도식한다.

[0085] 도 5를 참조하면, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 장치(220)는 메모리(221) 및 프로세서(222)를 포함할 수 있다. 다만 이에 제한되는 것은 아니고, 일부 구성이 직접적 NeRF 분류 장치(220)에서 생략되거나, 다른 범용적인 구성이 직접적 NeRF 분류 장치(220)에 더 포함될 수 있다.

[0086] 메모리(221)는 직접적 NeRF 분류 장치(220)에서 처리되는 각종 명령어들, 컴퓨터 프로그램, 소프트웨어, 모바일 애플리케이션 또는 데이터를 저장하기 위한 구조를 가질 수 있다. 예를 들면, 메모리(221)는 ROM, PROM, EPROM, EEPROM, 플래시 메모리, PRAM, MRAM, RRAM, FRAM 등과 같은 비휘발성 메모리, 또는 DRAM, SRAM, SDRAM, PRAM, RRAM, FeRAM 등의 휘발성 메모리로 구현될 수 있고, HDD, SSD, SD, Micro-SD 등의 형태, 또는 이들에 관한 조합의 형태로 구현될 수 있다.

[0087] 프로세서(222)는 직접적 NeRF 분류 장치(220)의 동작을 위해 요구되는 처리 과정들을 수행하기 위한 구조를 가

질 수 있다. 프로세서(222)는 각종 연산들을 처리하기 위한 다수의 논리 게이트들의 어레이 또는 범용적인 마이크로 프로세서로 구현될 수 있고, 단일의 프로세서 또는 복수의 프로세서들로 구성될 수 있다. 예를 들면, 프로세서(222)는 마이크로프로세서, CPU, GPU 및 AP 중 적어도 하나의 형태로 구현될 수 있다.

- [0088] 프로세서(222)는, 메모리(221)에 저장되는 명령어들을 실행함으로써, 3D 오브젝트(210)에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화하도록 구성될 수 있다. 프리-트레이닝된 프라이어 파라미터들(W_{pre})은 NeRF 네트워크들의 네트워크 파라미터들(W)의 도메인을 제한할 수 있고, 이를 통해 NeRF 네트워크들이 학습한 내용이 3D 오브젝트(210)에 대한 정보를 반영할 수 있다.
- [0089] 프로세서(222)는, 메모리(221)에 저장되는 명령어들을 실행함으로써, 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 관측 시퀀스들을 이용하여 NeRF 네트워크들을 트레이닝하도록 구성될 수 있다. NeRF 네트워크들은 관측 시퀀스들과 같은 실제 트레이닝 데이터에 의해 트레이닝됨과 동시에, 프라이어 파라미터들에 의한 정칙화에 관한 프리 트레이닝 데이터에 의해 트레이닝될 수 있다.
- [0090] 프로세서(222)는, 메모리(221)에 저장되는 명령어들을 실행함으로써, NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 NeRF 네트워크들의 네트워크 파라미터들을 산출하도록 구성될 수 있다. 프라이어 파라미터들에 의한 정칙화를 규정하는 베이지안 방법론 및 관측 시퀀스들에 기초하는 학습을 통해 네트워크 파라미터들(W)이 수렴하면, 네트워크 파라미터들(W)은 3D 오브젝트(210)에 관한 정보를 나타낼 수 있다.
- [0091] 프로세서(222)는, 메모리(221)에 저장되는 명령어들을 실행함으로써, 3D 오브젝트(210)의 구조 및 형상을 암시적으로 표현하는 네트워크 파라미터들을 직접 이용하여 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 분류하도록 구성될 수 있다. 네트워크 파라미터들(W)이 수렴하면, 네트워크 파라미터들(W)은 3D 오브젝트(210)의 구조 및 형상을 암시적으로 표현할 수 있다. 네트워크 파라미터들(W)을 직접 이용한다는 것은, NeRF 네트워크들의 출력값들을 이용하는 대신, 학습 과정을 통해 수렴한 네트워크 파라미터들(W)의 가중치 값들을 이용하는 것을 의미할 수 있다.
- [0092] 실시예에 따르면, 프로세서(222)는, 네트워크 파라미터들(W)이 프리-트레이닝된 프라이어 파라미터들(W_{pre})에 기반하는 프라이어 분포($p(W)$)를 따르도록, 3D 구조물들이 존재하지 않는 빈 3D 공간을 표현하는 프리-트레이닝 데이터셋(I_{pre})에 기초하여 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 획득하도록 구성될 수 있다. 이에 대해 보다 상세하게는 도 4 및 수식 10 부분이 참조될 수 있다. 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 생성하기 위한 프리-트레이닝 데이터셋(I_{pre})을 활용함으로써, 네트워크 파라미터들(W)이 따르게 되는 프라이어 분포($p(W)$)가 의도한 바에 따라 결정될 수 있다.
- [0093] 실시예에 따르면, 프리-트레이닝 데이터셋(I_{pre})에 의해 표현되는 빈 3D 공간은 카메라로부터의 레이가 3D 공간의 임의의 표면을 터치하여 NeRF 강도를 계산할 수 있도록 하는 바운디드 공간을 포함할 수 있다. 이에 대해 보다 상세하게는 도 4 및 수식 10 부분이 참조될 수 있다. 바운디드 공간이 프리-트레이닝 데이터셋(I_{pre})으로 활용되는 경우, 카메라로부터의 레이가 임의의 표면에 접촉하지 않는 경우가 발생하지 않기 때문에, 일부 픽셀이 아무런 정보도 갖지 못하는 경우가 회피될 수 있다. 예를 들면, 빈 3D 공간은 빈 방(empty room)일 수 있다.
- [0094] 실시예에 따르면, 프로세서(222)는, 프라이어 분포($p(W)$)에 따라 네트워크 파라미터들(W)을 제한함과 동시에 NeRF 네트워크들이 관측 시퀀스들을 나타낼 확률을 학습시킴으로써 네트워크 파라미터들(W)을 수렴시키도록 구성될 수 있다. NeRF 네트워크들의 학습을 통해 네트워크 파라미터들(W)이 수렴하는 과정에서, 네트워크 파라미터들(W)은 프라이어 분포($p(W)$)에 의해 제한됨과 동시에, 관측 시퀀스들이 갖는 정보를 반영할 수 있다.
- [0095] 실시예에 따르면, 프로세서(222)는, 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통한 초기화에도 불구하고 파괴적 망각(catastrophic forgetting)을 회피하기 위해 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 네트워크 파라미터들(W) 모두에 관하여 NeRF 네트워크들을 트레이닝하도록 구성될 수 있다. NeRF 네트워크들이 프리-트레이닝된 프라이어 파라미터들(W_{pre})을 통해 초기화되었음을 이유로, 학습 과정에서 네트워크 파라미터들(W)만을 이용한다면, 파괴적 망각으로 인해 NeRF 네트워크들이 프라이어 분포($p(W)$)를 유지하지 못할 수 있다. 이를 방지하기 위해, 프리-트레이닝된 프라이어 파라미터들(W_{pre}) 및 네트워크 파라미터들(W) 모두가 NeRF 네트워크들의 트레이닝에 관여할 수 있다.
- [0096] 실시예에 따르면, 프로세서(222)는, 로봇의 온-라인 학습을 위한 경량 분류기를 이용하여 네트워크 파라미터들

을 기반으로 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 분류하도록 구성될 수 있고, 경량 분류기는 서포트 벡터 머신(SVM) 및 랜덤 포레스트를 포함할 수 있다. 로봇이 탐사와 항해를 하면서 관측 시퀀스들을 수집함과 동시에 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 분류하는 온 라인 학습을 위해서는, 상대적으로 낮은 연산 리소스를 소모하는 경량 분류기가 필요할 수 있다. 이를 통해, 로봇은 탐사/항해와 동시에 3D 오브젝트(210)의 카테고리 및 인스턴스(230)를 분류할 수 있다.

- [0097] 도 6은 일부 실시예에 따른 직접적 NeRF 분류 방법의 인스턴스 및 카테고리에 대한 분류 결과를 도시한다.
- [0098] 도 6을 참조하면, 직접적 NeRF 분류 방법의 인스턴스 및 카테고리에 대한 분류 결과를 도시하는 테이블(600)이 도시될 수 있다.
- [0099] 테이블(600)은 인스턴스 분류 및 카테고리 분류 모두에 대한 성능 평가를 나타낸다. 성능 평가에 있어서, 비행기, 자동차, 의자, 책상, 소파, 화장실, 컵, 피아노, 테이블의 9개의 카테고리들 및 각 카테고리마다 2개의 인스턴스들이 선별되며, 총 18개의 인스턴스들이 활용된다. W_k 는 네트워크 f 의 k 번째 레이어의 파라미터를 표시한다. '(b)'는 $W - W_{pre}$ 가 사용되었음을, '(n)'은 정규화 파라미터가 사용되었음을, '(bn)'은 양자 모두가 적용되었음을 나타낸다.
- [0100] 테이블(600)에서, 제1 종래 기술(Pointnet++) 및 제2 종래 기술(PCT)와 비교할 때, 본 발명의 $W_0(bn)$ 및 W_0 가 더 높은 분류 성능을 보인다는 점이 확인될 수 있다. 한편, 통상적인 딥 뉴럴 네트워크의 경우와 유사하게, 네트워크 f 의 첫번째 레이어(W_0)가 후속 레이어들(W_1, W_2)보다 더 높은 성능을 보인다는 점이 확인될 수 있다.
- [0101] 도 7은 일부 실시예에 따른 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법을 구성하는 단계들을 도시한다.
- [0102] 도 7을 참조하면, 3D 오브젝트 인식을 위한 직접적 NeRF 분류 방법(700)은 단계(710) 내지 단계(740)를 포함할 수 있다. 다만 이에 제한되는 것은 아니고, 일부 단계가 생략되거나 다른 범용적인 단계가 추가될 수 있으며, 직접적 NeRF 분류 방법(700)의 단계들은 도시된 순서와 상이한 순서로 실행될 수 있다.
- [0103] 직접적 NeRF 분류 방법(700)은 직접적 NeRF 분류 장치(220)에서 시계열적으로 처리되는 단계들로 구성될 수 있다. 따라서, 이하에서 생략되는 내용이라 할지라도, 이상에서 직접적 NeRF 분류 장치(220)에 대해 설명되는 내용은 직접적 NeRF 분류 방법(700)에 대해서도 동일하게 적용될 수 있다.
- [0104] 직접적 NeRF 분류 방법(700)의 단계(710) 내지 단계(740)는 직접적 NeRF 분류 장치(220)의 메모리(221) 및 프로세서(222)에 의해 수행될 수 있다.
- [0105] 단계(710)에서, 직접적 NeRF 분류 장치(220)는, 3D 오브젝트에 대해 생성되는 관측 시퀀스들에 각각 대응하는 NeRF 네트워크들을 프리-트레이닝된 프라이어 파라미터들로 초기화할 수 있다.
- [0106] 단계(720)에서, 직접적 NeRF 분류 장치(220)는, 프라이어 파라미터들에 의한 정칙화(regularization)를 규정하는 베이지안 방법론에 따라 관측 시퀀스들을 이용하여 NeRF 네트워크들을 트레이닝할 수 있다.
- [0107] 단계(730)에서, 직접적 NeRF 분류 장치(220)는, NeRF 네트워크들에 대한 트레이닝을 통해 수렴하는 NeRF 네트워크들의 네트워크 파라미터들을 산출할 수 있다.
- [0108] 단계(740)에서, 직접적 NeRF 분류 장치(220)는, 3D 오브젝트의 구조 및 형상을 암시적으로 표현하는 네트워크 파라미터들을 직접 이용하여 3D 오브젝트의 카테고리 및 인스턴스를 분류할 수 있다.
- [0109] 실시예에 따르면, 직접적 NeRF 분류 방법(700)은 컴퓨터로 판독 가능한 저장 매체에 저장되는 컴퓨터 프로그램의 형태로 구현될 수 있다. 즉, 컴퓨터 프로그램은 직접적 NeRF 분류 방법(700)을 구현하기 위한 명령어들을 포함할 수 있고, 프로그램의 명령어들은 컴퓨터로 판독 가능한 저장 매체에 저장될 수 있다. 컴퓨터 프로그램은 모바일 애플리케이션을 포함할 수 있다.
- [0110] 실시예에 따르면, 컴퓨터 판독 가능 저장 매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 컴퓨터 프로그램 명령어들을 저장하고 실행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 컴퓨터 프로그램 명령어는 컴파일러에 의해 만들어지는 기계어 코드 및 인터프리터 등을 사용하여 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

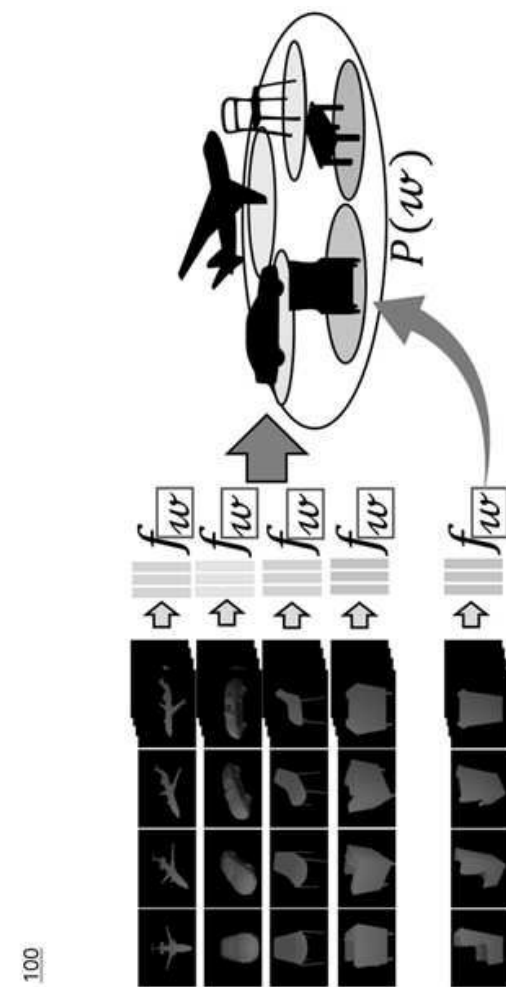
[0111] 이상에서 본 발명의 실시예들이 상세하게 설명되었으나 본 발명에 따른 권리범위가 이에 한정되는 것은 아니고, 다음의 청구범위에 기재되어 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명에 따른 권리범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0112] 100: 개요 210: 3D 오브젝트
220: 직접적 NeRF 분류 장치 220: 카테고리 및 인스턴스
221: 메모리 222: 프로세서
700: 직접적 NeRF 분류 장치

도면

도면1

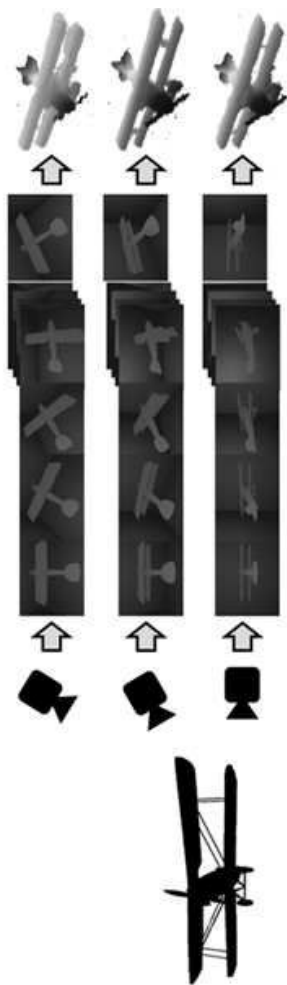


도면2

200

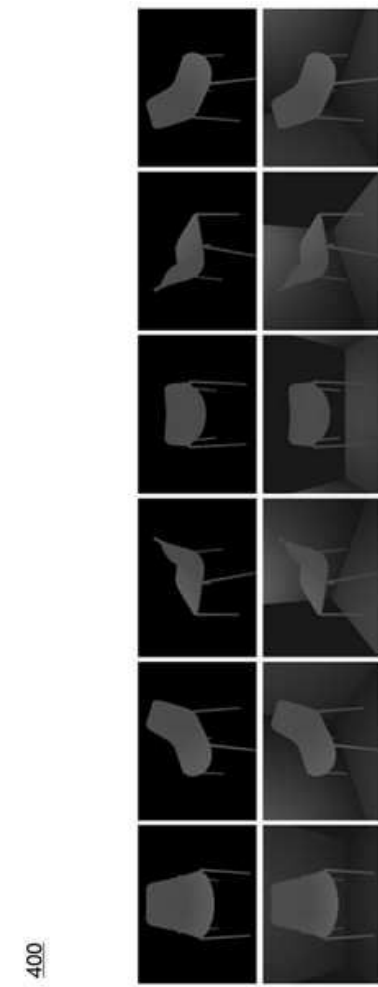


도면3

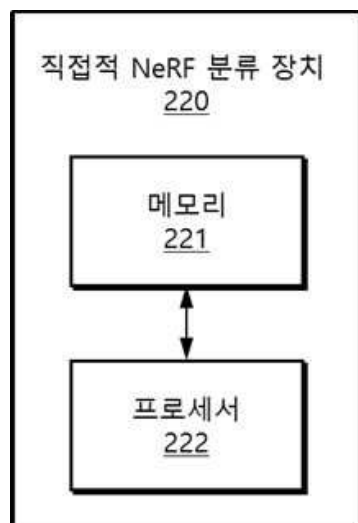


300

도면4



도면5



도면6

600

Deep features	Pointnet++ [2] PCT [3]	instance	category
Proposed	$W_0(\text{bn})$	0.9507	0.9793
	$W_0(n)$	0.9480	0.9741
	$W_0(b)$	0.9446	0.9896
	W_0	0.9431	0.9907
	W_1	0.7329	0.8263
	W_2	0.3282	0.4235

도면7

