(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년10월10일
(11) 등록번호 10-2587233
(24) 등록일자 2023년10월04일

(51) 국제특허분류(Int. Cl.)
H04N 13/156 (2018.01) *G06N 3/08* (2023.01)
H04N 13/15 (2018.01)
(52) CPC특허분류
H04N 13/156 (2018.05)
G06N 3/088 (2023.01)
(21) 출원번호 10-2022-0159773
(22) 출원일자 2022년11월24일
심사청구일자 2022년11월24일
(56) 선행기술조사문헌
KR1020190089683 A*
KR1020220026845 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
박인규
서울특별시 강남구 선릉로 221, 209동 502호(도곡동, 도곡렉슬아파트)
김수지
인천광역시 미추홀구
(74) 대리인
양성보

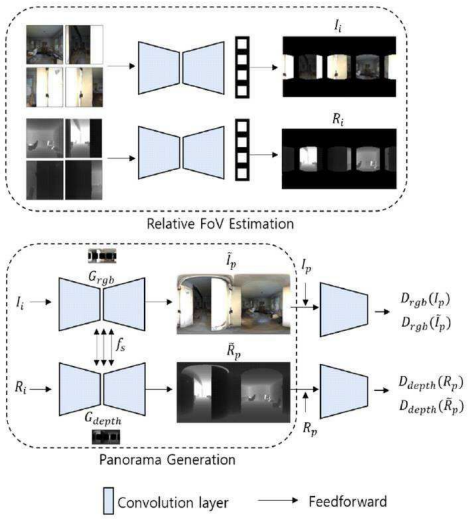
전체 청구항 수 : 총 9 항 심사관 : 옥윤철

(54) 발명의 명칭 **소수의 협소화각 RGBD 영상으로부터 360 RGBD 영상 합성**

(57) 요 약

소수의 협소화각 RGBD 영상으로부터 360 RGBD 영상을 합성하기 위한 방법 및 장치를 개시한다. 일실시예에 따른 영상 합성 방법은 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하는 단계 및 파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하는 단계를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
H04N 13/15 (2018.05)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711174118
과제번호 RS-2022-00155915
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신인재양성(교육운영수익)
연구과제명 [Ezbaro][정부+민간] 인공지능융합혁신인재양성(1차년도)
기 여 율 1/2
과제수행기관명 인하대학교 산학협력단
연구기간 2022.07.01 ~ 2022.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711152759
과제번호 2020-0-01389-003
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 인공지능융합연구센터지원사업(국고)
연구과제명 [Ezbaro][정부] 인공지능융합연구센터지원(3차년도)
기 여 율 1/2
과제수행기관명 인하대학교 산학협력단
연구기간 2022.01.01 ~ 2022.12.31

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

적어도 하나의 프로세서를 포함하는 컴퓨터 장치의 영상 합성 방법에 있어서,

상기 적어도 하나의 프로세서에 의해, 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하는 단계; 및

상기 적어도 하나의 프로세서에 의해, 파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하는 단계

를 포함하고,

상기 파노라마 생성 네트워크의 손실함수는, RGBD(Red Green Blue Depth) 네트워크의 RGB에 대한 제1 손실함수 및 상기 RGBD 네트워크의 깊이(Depth)에 대한 제2 손실함수를 포함하는 것을 특징으로 하는 영상 합성 방법.

청구항 2

제1항에 있어서,

상기 파노라마 생성 네트워크는 유-넷(U-Net) 기반의 적대적 생성 신경망 네트워크를 포함하는 것을 특징으로 하는 영상 합성 방법.

청구항 3

제1항에 있어서,

상기 파노라마 생성 네트워크는 적대적 손실함수로 LSGAN(Least Squares GAN(Generative Adversarial Network))을 사용하여 훈련되는 것을 특징으로 하는 영상 합성 방법.

청구항 4

삭제

청구항 5

제1항에 있어서,

상기 제1 손실함수는 상기 RGBD 네트워크의 RGB에 대한 제1 적대적 손실함수, 상기 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수, 상기 RGBD 네트워크의 RGB에 대한 지각 손실 목적함수 및 상기 생성된 영상과 상기 참값 영상간에 측정된 프레쳇 거리(Frechet distance) 손실함수를 이용하여 결정되는 것을 특징으로 하는 영상 합성 방법.

청구항 6

제1항에 있어서,

상기 제2 손실함수는 상기 RGBD 네트워크의 깊이에 대한 제2 적대적 손실함수, 상기 RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수, 상기 RGBD 네트워크의 깊이에 대한 지각 손실 목적함수 및 상기 생성된 영상과 상기 참값 영상간에 측정된 프레쳇 거리 손실함수를 이용하여 결정되는 것을 특징으로 하는 영상 합성 방법.

청구항 7

적어도 하나의 프로세서를 포함하는 컴퓨터 장치의 영상 합성 방법에 있어서,

상기 적어도 하나의 프로세서에 의해, 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야

(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하는 단계; 및

상기 적어도 하나의 프로세서에 의해, 파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하는 단계

를 포함하고,

상기 파노라마 생성 네트워크는 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상과 상기 RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상에 대하여 입력 영상의 참값 영역을 제외한 나머지 부분에 이진 마스크를 적용한 영상의 특징을 공유하고,

상기 파노라마 생성 네트워크의 마지막 레이어의 출력이 상기 RGBD 네트워크의 마지막 블록에 채널 연결을 수행하여 상기 RGBD 네트워크의 디코더에 전달되는 것을 특징으로 하는 영상 합성 방법.

청구항 8

컴퓨터 장치와 결합되어 제1항 내지 제3항 또는 제5항 내지 제7항 중 어느 한 항의 방법을 상기 컴퓨터 장치에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장되는 컴퓨터 프로그램.

청구항 9

제1항 내지 제3항 또는 제5항 내지 제7항 중 어느 한 항의 방법을 컴퓨터 장치에 실행시키기 위한 컴퓨터 프로그램이 기록되어 있는 컴퓨터 판독 가능한 기록매체.

청구항 10

컴퓨터 장치에 있어서,

상기 컴퓨터 장치에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서

를 포함하고,

상기 적어도 하나의 프로세서에 의해,

시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하고,

파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하고,

상기 파노라마 생성 네트워크의 손실함수는, RGBD(Red Green Blue Depth) 네트워크의 RGB에 대한 제1 손실함수 및 상기 RGBD 네트워크의 깊이(Depth)에 대한 제2 손실함수를 포함하는 것

을 특징으로 하는 컴퓨터 장치.

발명의 설명

기 술 분 야

[0001] 아래의 설명은 소수의 협소화각 RGBD 영상으로부터 360 RGBD 영상 합성하기 위한 영상 합성 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 최근 몰입감을 제공하는 실감형 미디어인 가상현실과 증강현실 기술에 대한 관심이 급상승하고 있다. 3D 장면 이해는 이러한 분야에서 매우 중요한 요소이며 그 중에서 가장 기본이 되는 연구 분야 중 하나는 깊이 추정 연구이다. 깊이 영상은 3차원 공간에서의 깊이 정보를 2차원 평면에 표현한 영상으로서, 공간의 3차원 구조 정보를 포함하므로 뷰 합성, 3D 모델링, 자율주행, 로봇공학 등 다양한 3D 비전 분야에 활용된다.

[0003] 기존의 좁은 FoV 카메라를 이용하여서 장면을 취득하는 경우에 주변 장면의 상당 부분이 소실된 일부분에 해당하는 영상을 얻게 되고, 전 방향 장면을 취득하려면 복수의 카메라를 구축하여 다수의 영상을 처리해야 하기 때문에 많은 비용이 발생하게 된다. 그 대안으로 FoV가 180° 이상인 넓은 FoV를 갖는 소수의 어안렌즈 카메라를 사용하여 전방향 영상을 취득할 수 있다. 어안렌즈는 구형 모델로 기존의 핀홀 카메라 기반의 평면 모델을 그

대로 활용하기 어렵기 때문에 등장방형도법(Equirectangular projection) 영상과 같은 구체 투영 영상을 사용할 수 있다. 그러나 이러한 360° 투영 영상은 경계면과 극 부분에 왜곡 문제가 발생할 수 있고, 최근에 이러한 360° 투영 영상의 특성을 반영한 딥러닝 네트워크도 제안되고 있다.

[0004] 오랜 기간 연구되어온 깊이 추정 연구는 전통적인 방식의 스테레오 영상을 이용한 깊이 추정, 최근 딥러닝 네트워크를 활용한 단안 영상에 대한 깊이 추정방식, 그리고 이 두 기법을 혼합한 방법이 연구되고 있다. 이러한 기법들은 대부분 좁은 FoV를 가지는 단안 영상을 대상으로 한다. 또한 일반적인 영상 데이터셋과 달리 좁은 FoV 영상 기반의 센서를 이용하여 취득된 고품질 360° 데이터셋은 충분하지 않으며, 특히 깊이 영상의 경우 깊이 정보가 취득된 화소에서는 높은 정밀도를 갖지만 딥러닝 네트워크 훈련에 적합하지 않은 희소한 깊이 영상이 취득되는 경우가 많다. 이러한 이유로 희소한 깊이 영상을 조밀한 깊이 영상으로 완성하는 연구가 수행되기도 하였으며, 스테레오 영상을 취득하고자 카메라를 새롭게 배치하여서 영상을 취득하기도 하지만 이 또한 일관성 있는 기준선을 유지하기에 힘들고 작업 비용이 크다는 단점이 있다.

[0005] [선행문헌번호]

[0006] 한국공개특허 제10-2020-0095112호

발명의 내용

해결하려는 과제

[0007] 소수의 협소화각 RGBD 영상으로부터 360 RGBD 영상 합성하기 위한 영상 합성 방법 및 장치를 제공한다.

과제의 해결 수단

[0008] 적어도 하나의 프로세서를 포함하는 컴퓨터 장치의 영상 합성 방법에 있어서, 상기 적어도 하나의 프로세서에 의해, 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하는 단계; 및 상기 적어도 하나의 프로세서에 의해, 파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하는 단계를 포함하는 영상 합성 방법을 제공한다.

[0009] 일측에 따르면, 상기 파노라마 생성 네트워크는 유-넷(U-Net) 기반의 적대적 생성 신경망 네트워크를 포함하는 것을 특징으로 할 수 있다.

[0010] 다른 측면에 따르면, 상기 파노라마 생성 네트워크는 적대적 손실함수로 LSGAN(Least Squares GAN(Generative Adversarial Network))을 사용하여 훈련되는 것을 특징으로 할 수 있다.

[0011] 또 다른 측면에 따르면, 상기 파노라마 생성 네트워크의 손실함수는, RGBD(Red Green Blue Depth) 네트워크의 RGB에 대한 제1 손실함수 및 상기 RGBD 네트워크의 깊이(Depth)에 대한 제2 손실함수를 포함하는 것을 특징으로 할 수 있다.

[0012] 또 다른 측면에 따르면, 상기 제1 손실함수는 상기 RGBD 네트워크의 RGB에 대한 제1 적대적 손실함수, 상기 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수, 상기 RGBD 네트워크의 RGB에 대한 지각 손실 목적함수 및 상기 생성된 영상과 상기 참값 영상간에 측정된 프레쳇 거리(Frechet distance) 손실함수를 이용하여 결정되는 것을 특징으로 할 수 있다.

[0013] 또 다른 측면에 따르면, 상기 제2 손실함수는 상기 RGBD 네트워크의 깊이에 대한 제2 적대적 손실함수, 상기 RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수, 상기 RGBD 네트워크의 깊이에 대한 지각 손실 목적함수 및 상기 생성된 영상과 상기 참값 영상간에 측정된 프레쳇 거리 손실함수를 이용하여 결정되는 것을 특징으로 할 수 있다.

[0014] 또 다른 측면에 따르면, 상기 파노라마 생성 네트워크는 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상과 상기 RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상에 대하여 입력 영상의 참값 영역을 제외한 나머지 부분에 이진 마스크를 적용한 영상의 특징을 공유하고, 상기 파노라마 생성 네트워크의 마지막 레이어의 출력이 상기 RGBD 네트워크의 마지막 블록에 채널 연결을 수행하여 상기 RGBD 네트워크의 디코더에 전달되는 것을 특징으로 할 수 있다.

[0015] 컴퓨터 장치와 결합되어 상기 방법을 컴퓨터 장치에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램을 제공한다.

[0016] 상기 방법을 컴퓨터 장치에 실행시키기 위한 프로그램이 기록되어 있는 컴퓨터 판독 가능한 기록매체를 제공한다.

[0017] 컴퓨터 장치에 있어서, 상기 컴퓨터 장치에서 판독 가능한 명령을 실행하도록 구현되는 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서에 의해, 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성하고, 파노라마 생성 네트워크를 이용하여 상기 생성된 시야 영상으로부터 파노라마 영상을 생성하는 것을 특징으로 하는 컴퓨터 장치를 제공한다.

발명의 효과

[0018] 소수의 협소화각 RGBD 영상으로부터 360 RGBD 영상 합성하기 위한 영상 합성 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 일실시예에 따른 RGBD 생성 네트워크의 전체적인 구조의 예를 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 RGBD 생성 네트워크의 특징 공유 모듈의 예를 도시한 도면이다.

도 3은 정제된 데이터셋의 샘플(3D60 데이터셋 샘플)의 예를 도시한 도면이다.

도 4는 생성된 RGBD 영상의 정성평가 결과의 예를 도시한 도면이다.

도 5는 생성된 RGBD 영상의 3D 포인트 클라우드 결과의 예를 도시한 도면이다.

도 6은 프레쳇 거리 손실함수의 사용 전후 결과를 퓨전 모듈 사용 전후 네트워크에 적용하여 비교한 결과의 예를 도시한 도면이다.

도 7은 프리쳇 거리의 검증 결과의 예를 도시한 도면이다.

도 8은 RGB 영상 결과에 대한 기존의 기법과의 정성적 비교 결과의 예를 도시한 도면이다.

도 9는 깊이 영상 결과에 대한 기존의 기법과의 정성적 비교 결과의 예를 도시한 도면이다.

도 10은 본 발명의 일실시예에 따른 컴퓨터 장치의 예를 도시한 블록도이다.

도 11은 본 발명의 일실시예에 따른 영상 합성 방법의 예를 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다. 하기 설명 및 첨부 도면에서 본 발명의 요지를 불필요하게 흐릴 수 있는 공지 기능 및 구성에 대한 상세한 설명은 생략한다.

[0021] 깊이 영상은 3차원 공간상의 거리 정보를 2차원 평면에 나타낸 영상이며 다양한 3D 비전 연구에서 유용하게 사용된다. 기존의 많은 깊이 추정 연구는 주로 좁은 FoV(Field of View) 영상을 사용하여 전체 장면 중 상당 부분이 소실된 영상에 대한 깊이 정보를 추정한다. 본 발명의 실시예들에서는 소수의 좁은 FoV 영상으로부터 360° 전 방향 RGBD 영상을 동시에 생성하는 방법 및 장치를 제공한다. 일례로, 오버랩 되지 않는 4장의 소수 영상으로부터 전체 파노라마 영상에 대해서 상대적인 FoV를 추정하고 360° 컬러 영상과 깊이 영상을 동시에 생성하는 적대적 생성 신경망 기반의 영상 생성 모델을 제공할 수 있으며, 두 모달리티의 특징을 공유하여 상호 보완된 결과를 확인한다. 또한 360° 영상의 구면 특성을 반영한 네트워크를 구성하여 개선된 성능을 보인다.

[0022] 보다 구체적으로, 본 발명의 실시예들에서는 기존의 문제점들을 고려하여 RGBD 영상을 생성하는 컨볼루션 신경망(CNN) 기반의 딥러닝 네트워크를 제공할 수 있으며, 겹치지 않는 n장(일례로, 4장)의 좁은 FoV 영상에 대해서 360° RGBD 파노라마 영상을 생성하는 프레임워크를 제공할 수 있다. 본 발명의 실시예들에서는 3D60 360° 실내 데이터셋에 대하여 일반화된 네트워크임을 보이고, 두 모달리티의 상호작용을 통해서 기존의 단일 네트워크보다 개선된 결과를 보인다. 또한 360° 영상 특성을 고려한 특징을 추출하여 프레쳇 거리(Frechet distance)를 계산하고, 이에 대한 손실함수를 이용하여 복원 성능을 개선하였다.

[0023] 본 발명의 실시예들의 주요 기여점을 요약하면 다음과 같다.

[0024] · 3D60 실내 데이터셋을 이용하여 기존 실외영상 기반 네트워크를 일반화한다.

[0025] · U-Net 기반의 적대적 생성 신경망을 이용하여 컬러(RGB) 영상과 깊이(D) 영상을 동시에 생성한다.

[0026] ·RGBD 두 모달리티의 특징을 공유하도록 하여 상호 보완된 성능을 보인다.

[0027] ·360° 영상 특성을 반영한 특징을 이용한 프레쳇 거리 손실 함수를 통해서 개선된 결과를 보인다.

[0029] 1. 관련 연구

[0030] 1-1. 구형 모델에 대한 조밀한 깊이 추정 연구

[0031] 360° 영상을 이용한 깊이 추정연구로는 기존의 좁은 FoV 영상 에서의 깊이 추정 연구와 유사하게 전통적인 방법인 스테레오 영상 기반의 깊이 추정 기법과 딥러닝 기반의 단안 영상에 대한 깊이 추정 연구 기법이 존재한다. 360SD-Net은 딥러닝 기반의 구형 스테레오 영상을 활용한 깊이 추정 네트워크로 Top-Bottom 구조로 카메라를 세팅하여 얻은 스테레오 영상과 구형 파노라마의 왜곡문제를 해결하기 위한 편각(polar angle) 영상을 사용하며 비용 볼륨 구축을 위해서 학습 가능한 시프팅 필터(Shifting filter)를 제안하였다. OmniDepth는 딥러닝 기반의 단안 영상에 대한 깊이 추정 연구로 오토인코더 구조의 360° 영상의 특성을 고려한 네트워크를 제안하였다. 구형 파노라마는 양 극으로 갈수록 왜곡이 심해지는 특징을 가지는데, 이러한 구조에 대해서 전역 컨텍스트를 위한 확장성(Dilated) 컨볼루션과 공간 상화 관계(Spatial correlation)를 위한 1X1 컨볼루션으로 네트워크를 설계하였다. BiFuse 또한 딥러닝 기반의 단안 영상에 대한 깊이 추정 연구이며, 구형 파노라마에서 등장방형 형식과 큐브맵 형식에 의한 경계면 문제를 개선하고자 두 형식을 융합하는 모듈을 제안하여 네트워크를 설계하였다. 또한 큐브맵 형식에서 구형 패딩 기법을 제안하여서 구형 파노라마 영상의 왜곡을 줄이도록 하였다. 조밀한 깊이 추정을 위한 깊이 완성(Depth completion) 연구에서 특징 공유 모듈을 이용하여 깊이 영상을 생성하기 위한 그레이-스케일(Gray-scale) 영상을 생성하여 보조하는 네트워크를 제안하였다. 본 발명의 실시예들에서는 두 모달리티의 융합된 특징을 사용하는 방법을 기반으로 네트워크를 설계하였다.

[0032] 1-2. 마스크 기반의 컨볼루션 신경망 네트워크

[0033] 입력 영상에 마스크를 적용하고 손상된 영역을 채워 넣는 연구는 오래된 연구 주제로 영상 편집과 같은 작업에 사용될 수 있다. 전통적인 방식으로 전파-기반(Diffusion-based) 기법은 참조 영상을 이용하여서 주변 영역으로부터 정보를 가져와서 채워 넣는 방법이며, 패치-기반(Patch-based) 기법은 입력 영상의 손상되지 않은 영역으로부터 화소 정보를 가져와서 비어 있는 영역을 채워 넣는 방식이다. 이러한 방법은 입력 영상에 의존적이며 전체적인 구조에 일관성이 떨어지고 의미 정보를 생성하기 어려운 한계점을 가진다. 최근 딥러닝 네트워크의 발전으로 생성모델은 기존의 방식과는 다르게 입력 영상에 의존하지 않고도 전체 구조에 대해서 의미 정보를 포함한 일관성을 가진 영상 생성이 가능하다는 점에서 마스크 기반의 영상 생성 연구에 많이 활용된다. StructureFlow은 영상에 랜덤하게 이진 마스크를 생성하여 적용하고 누락된 픽셀을 생성하는 딥러닝 기반의 네트워크로 구조 정보를 위해서 edge-preserve smooth 영상을 생성하고 이를 기반으로 세부적인 텍스처 정보를 생성하도록 하여 전체 영상을 복원하는 기법을 제안하였다. MED는 인코더-디코더 기반의 영상 생성 네트워크이며 얕은 레이어의 텍스처 정보와 깊은 레이어 부분의 구조 정보에 대해서 멀티 스케일 커널을 적용하여 융합한 뒤 평활화 처리를 수행하여 디코더에 더하도록 하여서 불규칙한 마스크 영역에 대한 복원을 수행한다.

[0034] 이처럼 360° 영상에 대한 깊이 추정 연구가 많이 수행되고 있지만 기존의 좁은 FoV를 입력으로 활용한 연구와 겹치지 않는 입력 영상에 대하여 전 방향 영상을 생성하는 연구는 드물다. 또한 대부분 영상 생성 연구에서 불규칙하고 영상 전체에 산발적으로 분포된 이진 마스크 영상을 사용하는데, 본 발명의 실시예들에서는 큰 블록 형태의 마스크로 주변 정보가 많지 않은 환경에서의 영상 생성으로 어려운 조건에 대한 네트워크를 보인다.

[0036] 2. 본 발명의 실시예들

[0037] 본 발명의 실시예들에 따른 네트워크는 360° 영상에 대해서 상대적인 FoV를 추정하는 네트워크와 추정된 FoV 영상으로부터 360° 영상을 생성하는 파노라마 생성 네트워크로 구성될 수 있다. 상대적인 FoV를 추정하는 네트워크와 파노라마 생성 네트워크는 별도의 훈련과정을 거치며 FoV 추정 단계 이후에 생성된 결과 영상에 마스크를 적용하고 등장방형 도법 영상 형식으로 변형한 입력에 대하여 파노라마를 생성할 수 있다. 파노라마 생성 결과는 RGBD의 특징을 공유하여 상호 작용된 결과이다.

[0038] 도 1은 본 발명의 일실시예에 따른 RGBD 생성 네트워크의 전체적인 구조의 예를 도시한 도면이고, 도 2는 본 발명의 일실시예에 따른 RGBD 생성 네트워크의 특징 공유 모듈의 예를 도시한 도면이다.

[0039] 2-1. 상대적인 FoV 추정

[0040] 전체 파노라마에 대해서 상대적인 FoV를 추정하는 단계는 파노라마 생성 이전에 수행되며 절대적인 각도가 아니라 전체 파노라마 영상에서 차지하는 크기의 추정 문제로 모델링할 수 있다. 4개의 수평 관측방향 영상을 입력

으로 사용할 수 있으며, 병목 레이어에서는 FoV 각도에 해당하는 256개의 클래스를 출력하고 분류 작업을 수행할 수 있다. 분류작업의 목적함수로 크로스 엔트로피 손실함수를 사용할 수 있으며, 디코더 레이어에서 FoV각도에 대한 패딩이 추가된 마스크 영상을 생성하여 그라운드 트루스(Ground truth) 마스크와의 L1 거리 목적함수를 사용할 수 있다. 기존의 네트워크를 확장하여 RGBD 영상에 대해서 적용할 수 있고, 다음 단계인 파노라마 생성단계에서는 FoV 추정 네트워크가 정확한 FoV를 추정한다는 가정하에 최근 관련 연구에서 많이 사용되는 마스크 비율 중에서 가장 고품질 영상을 생성하기 어려운 비율인 FoV 60° (마스크 비율 약 57%)의 입력에 대해서 실험을 진행하였다. 마스크 기반의 영상 생성 연구에서는 0-60% 사이의 마스크를 생성하여서 비율에 따른 평가 결과를 보이는데, 마스크의 비율이 클수록 생성해야 할 영역이 커지며, 성능이 낮아지는 결과를 확인할 수 있다. 또한 마스크의 형태를 랜덤한 형태, 그리드 형태 그리고 본 발명의 실시예들에서와 유사한 블록 형태로 구성하여 진행된 실험에서 랜덤한 마스크 구성(75% 비율)에서 가장 좋은 결과를 보였고 블록 마스크 구성(50% 비율)은 다른 마스크 형태와 비교하였을 때 가장 낮은 성능을 보였다.

[0041] 2-2. 파노라마 생성

[0042] 파노라마 생성 네트워크는 유-넷(U-Net) 기반의 적대적 생성 신경망 네트워크일 수 있다. 훈련시에 적대적 손실함수로 LSGAN(Least Squares GAN(Generative Adversarial Network))을 사용하며, 각 RGBD 네트워크에 대한 적대적 손실함수 L_{adv1}, L_{adv2}는 아래 수학식 1 및 수학식 2와 같다.

[0043] [수학식 1]

[0044] $$L_{adv1}(G_{rgb}, D_{rgb}) = \frac{1}{2}E_{I_p}[(D_{rgb}(I_i, I_p) - 1)^2] + \frac{1}{2}E_{I_i,\tilde{I}_p}\left[\left(D_{rgb}(I_i, \tilde{I}_p)\right)^2\right]$$

[0045] [수학식 2]

[0046] $$L_{adv2}(G_{depth}, D_{depth}) = \frac{1}{2}E_{R_p}[(D_{depth}(R_i, R_p) - 1)^2] + \frac{1}{2}E_{R_i,\tilde{R}_p}[(D_{depth}(R_i, \tilde{R}_p))^2]$$

[0047] G_{rgb}, D_{rgb}, G_{depth}, D_{depth}는 각각 RGBD 네트워크의 생성기와 식별기일 수 있다. 또한, I_i, $\tilde{I}_p$, R_i, $\tilde{R}_p$는 각각 G_{rgb}, G_{depth}의 입력과 생성된 영상일 수 있고, I_p, R_p는 생성된 영상에 대한 참값 영상일 수 있다.

[0048] 생성된 영상과 참값 영상 사이의 픽셀 손실함수는 L1 손실함수를 사용할 수 있다. 다음 L_{pix1}, L_{pix2}는 RGBD 네트워크의 픽셀 손실함수로 아래 수학식 3 및 수학식 4와 같이 나타낼 수 있다.

[0049] [수학식 3]

[0050] $$L_{pix1} = E_{I_p,\tilde{I}_p}[\| I_p - \tilde{I}_p \|_1]$$

[0051] [수학식 4]

[0052] $$L_{pix2} = E_{R,\tilde{R}_p}[\| R_p - \tilde{R}_p \|_1]$$

[0053] 또한 사실적인 영상 생성을 위해서 지각 손실 목적함수로 사전 훈련된 VGG(Visual Geometry Group) 네트워크를 사용할 수 있다. 다음 L_{vgg1}, L_{vgg2}는 RGBD 네트워크의 지각 손실 목적함수로 아래 수학식 5 및 수학식 6과 같이 나타낼 수 있다. $VGG(\cdot)$는 VGG 네트워크로부터 추출된 i번째 특징벡터일 수 있다.

[0054] [수학식 5]

[0055] $$L_{vgg1} = \sum_i [\| VGG_i(I_p) - VGG_i(\tilde{I}_p) \|_1]$$

[0056] [수학식 6]

$$L_{vgg2} = \sum_{i}[\parallel VGG_i(R_p) - VGG_i(\tilde{R}_p) \parallel_1]$$

[0057]

[0058] 360° 영상 특징을 반영한 특징을 추출하기 위해서 본 발명의 실시예들에 따른 모델을 훈련하고 사전 훈련된 모델을 이용하여 RGBD 특징이 공유된 모듈 잠재공간의 마지막 레이어의 특징 f_{latent}를 추출할 수 있다. 그런 다음 종방향 불변성(Longitudinal invariant) 특징 f'_{latent}와 횡방향 등변성(latitudinal equivariant) 특징 f''_{latent}을 추출할 수 있다. 종방향 불변성 특징과 횡방향 등변성 특징은 각각 수학식 7 및 수학식 8과 같이 나타낼 수 있다. c, w, h는 각 채널, 너비, 높이를 의미할 수 있다. 단일 영상의 특징맵 거리를 계산하여 유사도를 향상시키는 VGG 손실 목적함수와 달리 생성된 영상 데이터셋과 그라운드 트루스 영상 데이터셋 사이의 분포 거리를 측정하게 되며 퓨전(Fusion) 모듈의 마지막 레이어를 사용하여 융합된 RGBD 특징벡터 분포의 유사도가 반영될 수 있다.

[0059] [수학식 7]

$$f'_{latent}(c,h) = \frac{1}{W}\sum_{W} f_{latent}(c,h,w)$$

[0060]

[0061] [수학식 8]

$$f''_{latent}(c,h) = cos\phi \cdot f'_{latent}(c,h)$$

[0062]

[0063] 그리고 프레쳇 거리를 측정하여 손실함수로 사용할 수 있다. 프레쳇 거리 $d(\because)$는 그라운드 트루스 영상의 Mean (m, C) 가우시안과 생성된 영상의 Mean (m', C') 가우시안 사이의 프리쳇 거리로서 아래 수학식 9와 같이 나타낼 수 있다. m, C, m', C' 는 각각 그라운드 트루스 영상과 생성된 영상의 평균(Mean)과 공분산(Covariance)일 수 있다. 프레쳇 거리 손실함수는 훈련 중 0번째 반복에서 측정된 값 d_0^2을 나누어서 다음 수학식 10과 같이 정규화할 수 있다.

[0064] [수학식 9]

$$d^2\big((m,C),(m',C')\big) = \parallel m - m' \parallel_2^2 + Tr(C + C' - 2(CC')^{\frac{1}{2}})$$

[0065]

[0066] [수학식 10]

$$L_d = \frac{d^2}{d_0^2}$$

[0067]

[0068] 따라서 파노라마 생성 네트워크의 전체 손실함수 L_{rgb}, L_{depth}는 아래 수학식 11 및 수학식 12와 같으며, 이때 λ_1, λ_2, λ_3는 상수인자일 수 있다.

[0069] [수학식 11]

$$L_{rgb} = \underset{G_{rgb}}{arg\ min}\ \underset{D_{rgb}}{max}\ L_{adv1} + \lambda_1 L_{pix1} + \lambda_2 L_{vgg1} + \lambda_3 L_{d1}$$

[0070]

[0071] [수학식 12]

$$L_{depth} = \arg \min_{G_{depth}} \max_{D_{depth}} L_{adv2} + \lambda_1 L_{pix2} + \lambda_2 L_{vgg2} + \lambda_3 L_{d2}$$

[0072]

[0073] 각 RGBD 네트워크에서 생성된 영상에 대하여 입력 영상의 참값 영역을 제외한 나머지 부분에 해당하는 이진 마스크를 적용한 영상의 특징을 공유하며 각 특징은 다음 수학식 13 및 수학식 14와 같이 나타낼 수 있다.

[0074] [수학식 13]

$$f_{rgb} = F(M\tilde{I}_p)$$

[0075]

[0076] [수학식 14]

$$f_{depth} = F(M\tilde{R}_p)$$

[0077]

[0078] $F()$는 인코더일 수 있고, M은 가장 큰 FoV 클래스에 해당하는 이진 마스크일 수 있다. 각 RGBD 네트워크의 레이어는 픽셀 합계를 수행하며 아래 수학식 15 및 수학식 16과 같이 나타낼 수 있다. 마지막 레이어 f_s는 각 RGBD 네트워크의 마지막 블록에 채널 연결을 수행하여 각 디코더에 전달될 수 있다. $concat(\cdot)$은 채널 기반의 연결 함수일 수 있다. 제안하는 특징 공유 모듈은 도 2에 도시하였다.

[0079] [수학식 15]

$$concat(G_i(f_{rgb}), f_s)$$

[0080]

[0081] [수학식 16]

$$concat(G_r(f_{depth}), f_s)$$

[0082]

[0083] 2-3. 실험 결과

[0084] 총 21,600개의 실내 데이터셋 Matterport3D, Stanford-3D, SunCG에서 깊이 영상을 기준으로 손상된 영역이 일정 값 이상에 해당하는 훈련에 부적절한 데이터셋 3,446개를 제거하는 과정을 거치며, 총 18,154개의 데이터셋에 대하여 훈련 데이터셋(80%), 테스트 데이터셋(20%)으로 나누어서 적용한다.

[0085] 도 3은 정제된 데이터셋의 샘플(3D60 데이터셋 샘플)의 예를 도시한 도면이다. 도 3에서 1행 및 2행은 RGBD 영상 샘플이며, 3행은 깊이 영상을 기준으로 손상된 영역을 시각화한 맵이다.

[0086] 네트워크 훈련에서는 ADAM(Adaptive momenteum) 옵티마이저를 사용하며, 학습률(Learning rate) α = 0.0002, β_1 = 0.5, β_2 = 0.99, 배치 사이즈(batch size)는 2로 설정하고, NVIDIA RTX A6000 GPU를 사용하여 훈련하였다. Matterport3D, Stanford3D, SunCG 데이터셋에 대해서 큐브맵 형식으로 변환하여 4개의 면을 훈련에 사용할 수 있도록 데이터셋을 구축하였으며, 세 개의 데이터셋을 각각 따로 훈련하였다.

[0087] 정량평가에서 생성 모델의 유사도를 위한 PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index Map)과 깊이 영상의 유사도 평가를 위한 절대 차이(Abs Diff), 절대 상대 오차(Abs Rel), 상대 오차 제곱(Sq Rel), 평균 제곱근 오차(RMS), 평균 제곱근 로그 오차(RMS log), 상대 정확도(δ)를 측정하였다. 평가에 사용된 마스크 비율은 최근 관련 연구에서 많이 사용하는 가장 넓은 영역의 마스크 비율인 FoV 60° 입력에 대하여 평가를 진행하였다.

[0088] 정성평가를 위해서 제안하는 퓨전 블록을 사용한 네트워크와 퓨전 블록을 사용하지 않고 RGBD를 각각 따로 구성한 네트워크와 비교하였으며, 프레쳇 거리 손실함수를 사용하기 전과 후의 결과를 비교하였다. 또한 생성된 RGBD 영상에 대한 포인트 클라우드를 생성하여서 비교하였다.

[0089] 도 4는 생성된 RGBD 영상의 정성평가 결과의 예를 도시한 도면이다. 도 4에서는 본 발명의 일실시예에 따른 네트워크와 특징 공유 모듈을 사용하지 않은 네트워크의 결과를 정성적으로 비교하고 있다. 도 4에서 1행 및 3행은 컬러 영상을, 2행 및 4행은 깊이 영상 샘플을 나타낸다. 본 발명의 실시예들에 따른 네트워크의 결과에서 전반적으로 전체적인 레이아웃과 디테일한 정보를 잘 생성하는 것을 확인할 수 있다.

[0090] 도 5는 생성된 RGBD 영상의 3D 포인트 클라우드 결과의 예를 도시한 도면이다. 도 5에서는 3차원 포인트 클라우드 결과를 비교하여 평가하였으며, 이를 통해 생성된 깊이 영상에서 정성적으로 비교하기 어려운 3차원 레이아웃과 잡음의 영향을 확인할 수 있다.

[0091] 표 1은 생성된 깊이 영상에 대한 다양한 정량평가 결과이며, 특징공유 모듈을 사용하기 전과 후를 비교하였다.

[0092] [표 1]

[0093]

Method	Dataset	Abs Diff	Abs Rel	Sq Rel	RMS	RMS log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
w/o Fusion	Matterport3D	**69.2760**	**1.5925**	0.4759	5.5254	0.1607	0.9128	0.9709	0.9862
	Stanford3D	**62.4588**	**1.3190**	0.4855	5.3640	0.1654	**0.9160**	0.9687	0.9858
	SunCG	106.6007	3.3904	**0.5770**	**4.2691**	0.1562	**0.9111**	0.9525	0.9709
w/ Fusion	Matterport3D	84.7544	1.8244	**0.4563**	**5.3990**	**0.1490**	**0.9204**	**0.9738**	**0.9881**
	Stanford3D	87.2217	1.6361	**0.4773**	**5.1792**	**0.1617**	0.9135	**0.9707**	**0.9866**
	SunCG	**84.2728**	**3.0317**	0.5868	4.3606	**0.1503**	0.9101	**0.9529**	**0.9718**

[0094] 표 2, 표 3에서는 프레쳇 거리 손실함수를 사용하기 전과 후의 결과이다. 표 2는 프레쳇 거리 손실을 사용하지 않았을 때의 RGBD 영상 평가 결과이며, 표 3은 프레쳇 거리 손실을 적용한 RGBD 영상 평가 결과이다.

[0095] [표 2]

[0096]

Method		Dataset	PSNR	SSIM
w/o Fusion	RGB	Matterport3D	17.2832	0.6535
		Stanford3D	18.0947	0.6716
		SunCG	20.7043	0.7987
	Depth	Matterport3D	25.8226	0.9282
		Stanford3D	21.7707	0.9007
		SunCG	28.0096	**0.9430**
w/ Fusion	RGB	Matterport3D	**17.6667**	**0.6705**
		Stanford3D	**18.2139**	**0.6748**
		SunCG	**21.7096**	**0.8072**
	Depth	Matterport3D	**26.1095**	**0.9294**
		Stanford3D	**25.4359**	**0.9185**
		SunCG	**28.4385**	0.9418

[0097] [표 3]

Method		Dataset	PSNR	SSIM
w/o Fusion	RGB	Matterport3D	17.5454	0.6658
		Stanford3D	18.3560	0.6851
		SunCG	20.9933	0.8011
	Depth	Matterport3D	25.7250	0.9276
		Stanford3D	25.2045	0.9168
		SunCG	28.1536	0.9419
w/ Fusion	RGB	Matterport3D	**18.1325**	**0.6850**
		Stanford3D	**18.4782**	**0.6859**
		SunCG	**21.0893**	**0.8051**
	Depth	Matterport3D	**26.1831**	**0.9301**
		Stanford3D	**25.6019**	**0.9187**
		SunCG	**28.3960**	**0.9435**

[0098]

[0099] 생성된 RGBD 영상 모두 특징공유 모듈을 사용하지 않은 네트워크 결과와 비교하였을 때, 제안하는 네트워크의 생성 영상이 더 좋은 결과를 보이며, 프레쳇 거리 손실함수를 사용하기 전과 비교하여서도 더 좋은 결과를 확인할 수 있다.

[0100] 도 6은 프레쳇 거리 손실함수의 사용 전후 결과를 퓨전 모듈 사용 전후 네트워크에 적용하여 비교한 결과의 예를 도시한 도면이다. 도 6에서 각 1,2행과 3,4행은 RGBD 영상 샘플 결과이다.

[0101] 프레쳇 거리의 검증을 위해서 제안하는 모델의 사전 훈련된 모델로부터 특징을 추출하고, 파노라마 영상에 대해서 가우시안 블러(Gaussian blur), 솔트앤페퍼(Salt and pepper) 노이즈를 4가지 단계로 생성하여 프리쳇 거리를 측정하였다.

[0102] 도 7은 프리쳇 거리의 검증 결과의 예를 도시한 도면이다. 도 7에서는 RGBD 영상 모두 그라운드 트루스 영상의 특징 분포와 생성된 영상의 특징 분포의 거리가 장애(Disturbance) 레벨이 커질수록 증가하는 것을 확인할 수 있으며 제안하는 네트워크를 이용하여 특징을 추출하고 프레쳇 거리를 계산하였을 때 블러와 노이즈를 정상적으로 캡처하는 것을 확인할 수 있다. (1)-(4)는 각 장애 레벨을 의미할 수 있다.

[0103] 또한 제안하는 특징공유 모듈에 대해서 잔차블록(Residual block)을 사용하지 않았을 때, 한 개의 잔차블록만을 사용하였을 때 그리고 직접 연결을 수행하였을 때의 3가지 추가적인 케이스에 대해서 실험하였다. 표 4는 프레쳇 거리를 손실하지 않은 네트워크에 대한 3가지 구성에 대한 정량 평가 결과이다.

[0104] [표 4]

Method		Dataset	PSNR	SSIM
w/o Residual blocks	RGB	Matterport3D	17.0306	0.6434
		Stanford3D	**18.5446**	0.6866
		SunCG	21.4293	0.8036
	Depth	Matterport3D	**26.4858**	**0.9312**
		Stanford3D	**25.7579**	**0.9206**
		SunCG	28.0848	**0.9435**
One residual block	RGB	Matterport3D	**18.0804**	**0.6811**
		Stanford3D	18.5318	**0.6889**
		SunCG	21.2478	0.8039
	Depth	Matterport3D	26.1706	0.9303
		Stanford3D	25.6402	0.9203
		SunCG	**28.5691**	0.9428
Directly concate nation	RGB	Matterport3D	17.8931	0.6808
		Stanford3D	18.5357	0.6880
		SunCG	21.3926	0.8024
	Depth	Matterport3D	26.2732	0.9308
		Stanford3D	25.2013	0.9170
		SunCG	28.0640	0.9421

[0105]

[0106] 특징공유 모듈을 사용하지 않았을 때 보다 3가지 특징공유 모듈을 사용하였을 때 전반적으로 더 좋은 결과를 볼 수 있다. 다른 모델과의 비교를 위해서 RGB 영상 생성에 대한 모델로 인페인팅 모델과 w/o Fusion 네트워크인 기존의 모델에 대해서 비교하였다. 두 모델 모두 같은 조건인 무작위 FoV로 재훈련하였으며, FoV 60° 에 대해서 정량적, 정성적 평가를 수행하였다. 그리고 깊이 영상 비교를 위해서 깊이 추정 모델의 훈련된 모델 그리고 같은 조건으로 다시 훈련한 인페인팅 모델, 기존의 모델을 비교하였다.

[0107] 표 5와 표 6에서 Matterport3D 데이터셋 1,759개에 대한 PSNR, SSIM 평균을 구하여 비교하였다. 표 5는 기존 기법과의 RGB 영상 생성 결과를 비교한 예를 나타내고 있으며, 표 6은 기존 기법과의 깊이 영상 생성 결과를 비교한 예를 나타내고 있다.

[0108] [표 5]

Method	PSNR	SSIM
Ours	**18.1325**	**0.6850**
EdgeConnect [23]	15.1862	0.6039
Sumantri [14]	17.2832	0.6535

[0109]

[0110] [표 6]

Method	Abs Diff	Abs Rel	Sq Rel	RMS	RMS log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Ours	84.7544	**1.8244**	**0.4563**	5.3990	**0.1490**	**0.9204**	**0.9738**	**0.9881**
Bifuse(GT)	116.4334	4.4129	2.5221	9.3509	0.6489	0.3782	0.6052	0.7351
Bifuse(Generated)	95.0981	3.9783	2.1558	9.6091	0.7790	0.2706	0.4764	0.6367
HoHoNet(GT)	**48.8880**	2.9683	2.0377	8.2796	0.3242	0.6760	0.8788	0.9376
HoHoNet(Generated)	85.0457	3.5670	2.0867	8.6423	0.3710	0.5941	0.8304	0.9160
EdgeConnect	124.0315	3.7339	1.4945	7.9192	0.4960	0.5717	0.6956	0.8092
Sumantri	62.4457	2.9313	1.3701	**5.3454**	0.1620	0.9156	0.9723	0.9866

[0111]

[0112] 도 8은 RGB 영상 결과에 대한 기존의 기법과의 정성적 비교 결과의 예를 도시한 도면이고, 도 9는 깊이 영상 결

과에 대한 기존의 기법과의 정성적 비교 결과의 예를 도시한 도면이다. 깊이 영상의 경우에 다른 모델과 비교하기 위해서 Top, Bottom 부분을 제외하고 평가하였다. 또한 참값 RGB 영상을 입력으로 생성된 깊이 영상과 제안하는 모델의 RGB 영상을 입력으로 생성된 깊이 영상을 비교하였다.

[0113] 본 발명의 실시예들에서는 소수의 영상으로부터 RGBD 영상을 동시에 생성하는 적대적 생성 신경망 기반 네트워크를 제공할 수 있다. 두 모달리티의 특징을 공유한 생성 모델에 대하여 단일 네트워크보다 개선된 성능을 정량적, 정성적으로 확인하였으며, 360° 영상 특징이 반영된 프레쳇 거리 손실함수를 적용하여서 개선된 성능을 보였다. 또한 특징 공유 모듈에 대한 절제연구(Ablation study)를 통해서 기존의 단일 네트워크보다 높은 성능을 확인하였으며, 그 중에서 제안하는 구조에서 좋은 성능을 보이는 것을 볼 수 있었다. 기존의 RGBD 영상 생성 모델과 비교하였을 때에도 정성적, 정량적으로 우수한 성능을 보이는 것을 확인하였다. 기존의 파노라마 영상에 대한 생성 모델과는 달리 높은 비율의 마스크가 적용된 겹치지 않는 소수의 영상으로부터 고품질 RGBD 영상을 동시에 생성한다는 점과 360° 특징이 반영되고 RGBD의 상호보완된 결과를 갖는 점에서 차별점을 갖고, 고품질 RGBD 영상 생성을 통해서 복잡한 3D 장면 재구성에 기여할 수 있다.

[0114] 본 발명의 실시예들에 따른 영상 합성 장치는 적어도 하나의 컴퓨터 장치에 의해 구현될 수 있다. 이때, 컴퓨터 장치에는 본 발명의 일실시예에 따른 컴퓨터 프로그램이 설치 및 구동될 수 있고, 컴퓨터 장치는 구동된 컴퓨터 프로그램의 제어에 따라 본 발명의 실시예들에 따른 영상 합성 방법을 수행할 수 있다. 상술한 컴퓨터 프로그램은 컴퓨터 장치와 결합되어 영상 합성 방법을 컴퓨터에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장될 수 있다.

[0115] 도 10은 본 발명의 일실시예에 따른 컴퓨터 장치의 예를 도시한 블록도이고, 도 11은 본 발명의 일실시예에 따른 영상 합성 방법의 예를 도시한 흐름도이다. 컴퓨터 장치(Computer device, 1000)는 도 10에 도시된 바와 같이, 메모리(Memory, 1010), 프로세서(Processor, 1020), 통신 인터페이스(Communication interface, 1030) 그리고 입출력 인터페이스(I/O interface, 1040)를 포함할 수 있다. 메모리(1010)는 컴퓨터에서 판독 가능한 기록매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 여기서 ROM과 디스크 드라이브와 같은 비소멸성 대용량 기록장치는 메모리(1010)와는 구분되는 별도의 영구 저장 장치로서 컴퓨터 장치(1000)에 포함될 수도 있다. 또한, 메모리(1010)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 메모리(1010)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 메모리(1010)로 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 인터페이스(1030)를 통해 메모리(1010)에 로딩될 수도 있다. 예를 들어, 소프트웨어 구성요소들은 네트워크(Network, 1060)를 통해 수신되는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 컴퓨터 장치(1000)의 메모리(1010)에 로딩될 수 있다.

[0116] 프로세서(1020)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(1010) 또는 통신 인터페이스(1030)에 의해 프로세서(1020)로 제공될 수 있다. 예를 들어 프로세서(1020)는 메모리(1010)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다.

[0117] 통신 인터페이스(1030)는 네트워크(1060)를 통해 컴퓨터 장치(1000)가 다른 장치(일례로, 앞서 설명한 저장 장치들)와 서로 통신하기 위한 기능을 제공할 수 있다. 일례로, 컴퓨터 장치(1000)의 프로세서(1020)가 메모리(1010)와 같은 기록 장치에 저장된 프로그램 코드에 따라 생성한 요청이나 명령, 데이터, 파일 등이 통신 인터페이스(1030)의 제어에 따라 네트워크(1060)를 통해 다른 장치들로 전달될 수 있다. 역으로, 다른 장치로부터의 신호나 명령, 데이터, 파일 등이 네트워크(1060)를 거쳐 컴퓨터 장치(1000)의 통신 인터페이스(1030)를 통해 컴퓨터 장치(1000)로 수신될 수 있다. 통신 인터페이스(1030)를 통해 수신된 신호나 명령, 데이터 등은 프로세서(1020)나 메모리(1010)로 전달될 수 있고, 파일 등은 컴퓨터 장치(1000)가 더 포함할 수 있는 저장 매체(상술한 영구 저장 장치)로 저장될 수 있다.

[0118] 입출력 인터페이스(1040)는 입출력 장치(I/O device, 1050)와의 인터페이스를 위한 수단일 수 있다. 예를 들어, 입력 장치는 마이크, 키보드 또는 마우스 등의 장치를, 그리고 출력 장치는 디스플레이, 스피커와 같은 장치를 포함할 수 있다. 다른 예로 입출력 인터페이스(1040)는 터치스크린과 같이 입력과 출력을 위한 기능이 하나로 통합된 장치와의 인터페이스를 위한 수단일 수도 있다. 입출력 장치(1050) 중 적어도 하나는 컴퓨터 장치(1000)와 하나의 장치로 구성될 수도 있다. 예를 들어, 스마트폰과 같이 터치스크린, 마이크, 스피커 등이

컴퓨터 장치(1000)에 포함된 형태로 구현될 수 있다.

[0119] 또한, 다른 실시예들에서 컴퓨터 장치(1000)는 도 10의 구성요소들보다 더 적은 혹은 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 컴퓨터 장치(1000)는 상술한 입출력 장치(1050) 중 적어도 일부를 포함하도록 구현되거나 또는 트랜시버(transceiver), 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수도 있다.

[0120] 본 실시예에 따른 영상 합성 방법은 영상 합성 장치를 구현하는 컴퓨터 장치(1000)에 의해 수행될 수 있다. 이때, 컴퓨터 장치(1000)의 프로세서(1020)는 메모리(1010)가 포함하는 운영체제의 코드나 적어도 하나의 컴퓨터 프로그램의 코드에 따른 제어 명령(instruction)을 실행하도록 구현될 수 있다. 여기서, 프로세서(1020)는 컴퓨터 장치(1000)에 저장된 코드가 제공하는 제어 명령에 따라 컴퓨터 장치(1000)가 도 11의 방법이 포함하는 단계들(1110 및 1120)을 수행하도록 컴퓨터 장치(1000)를 제어할 수 있다.

[0121] 단계(1110)에서 컴퓨터 장치(1000)는 시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야(Field of View, FoV)를 추정함으로써, 시야 영상을 생성할 수 있다. 이미 설명한 바와 같이, 시야를 추정하는 문제는 파노라마 생성 이전에 수행되며 절대적인 각도가 아니라 전체 파노라마 영상에서 차지하는 크기의 추정 문제로 모델링될 수 있다. n(일례로, n=4) 개의 수평 관측방향 영상을 입력으로 사용할 수 있으며, 시야 추정 네트워크의 병목 레이어에서는 시야 각도에 해당하는 256개의 클래스를 출력하고 분류 작업을 수행할 수 있다. 분류작업의 목적함수로 크로스 엔트로피 손실함수를 사용할 수 있으며, 디코더 레이어에서 시야 각도에 대한 패딩이 추가된 마스크 영상을 생성하여 그라운드 트루스(Ground truth) 마스크와의 L1 거리 목적함수를 사용할 수 있다. 기존의 네트워크를 확장하여 RGBD 영상에 대해서 적용할 수 있다.

[0122] 단계(1120)에서 컴퓨터 장치(1000)는 파노라마 생성 네트워크를 이용하여 생성된 시야 영상으로부터 파노라마 영상을 생성할 수 있다. 일실시예로, 파노라마 생성 네트워크는 유-넷(U-Net) 기반의 적대적 생성 신경망 네트워크를 포함할 수 있으며, 파노라마 생성 네트워크는 적대적 손실함수로 LSGAN(Least Squares GAN(Generative Adversarial Network))을 사용하여 사전 훈련될 수 있다.

[0123] 한편, 파노라마 생성 네트워크의 손실함수는, RGBD(Red Green Blue Depth) 네트워크의 RGB에 대한 제1 손실함수 및 RGBD 네트워크의 깊이(Depth)에 대한 제2 손실함수를 포함할 수 있다. 일례로, 제1 손실함수는 수학식 11의 L_{rgb}에 대응할 수 있고, 제2 손실함수는 수학식 12의 L_{depth}에 대응할 수 있다. 수학식 11에 나타난 바와 같이, 제1 손실함수는 RGBD 네트워크의 RGB에 대한 제1 적대적 손실함수(일례로, 수학식 1의 L_{adv1}), 상기 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수(일례로, 수학식 3의 L_{pix1}), 상기 RGBD 네트워크의 RGB에 대한 지각 손실 목적함수(일례로, 수학식 5의 L_{vgg1}) 및 상기 생성된 영상과 상기 참값 영상간에 측정된 프레쳇 거리(Frechet distance) 손실함수(일례로, 수학식 11의 L_{d1})를 이용하여 결정될 수 있다. 또한, 수학식 12에 나타난 바와 같이 제2 손실함수는 RGBD 네트워크의 깊이에 대한 제2 적대적 손실함수(일례로, 수학식 2의 L_{adv2}), RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상 및 대응하는 참값 영상간의 픽셀 손실함수(일례로, 수학식 4의 L_{pix2}), RGBD 네트워크의 깊이에 대한 지각 손실 목적함수(일례로, 수학식 6의 L_{vgg2}) 및 생성된 영상과 참값 영상간에 측정된 프레쳇 거리 손실함수(일례로, 수학식 12의 L_{d2})를 이용하여 결정될 수 있다.

[0124] 또한, 도 2의 특징 공유 모델 및 수학식 13 내지 수학식 16을 통해 설명한 바와 같이, 파노라마 생성 네트워크는 RGBD 네트워크의 RGB에 대한 생성기에 의해 생성된 영상과 RGBD 네트워크의 깊이에 대한 생성기에 의해 생성된 영상에 대하여 입력 영상의 참값 영역을 제외한 나머지 부분에 이진 마스크를 적용한 영상의 특징을 공유하고, 파노라마 생성 네트워크의 마지막 레이어의 출력이 RGBD 네트워크의 마지막 블록에 채널 연결을 수행하여 RGBD 네트워크의 디코더에 전달될 수 있다.

[0125] 이상에서 설명된 시스템 또는 장치는 하드웨어 구성요소, 또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 어플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를

접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0126] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록매체에 저장될 수 있다.

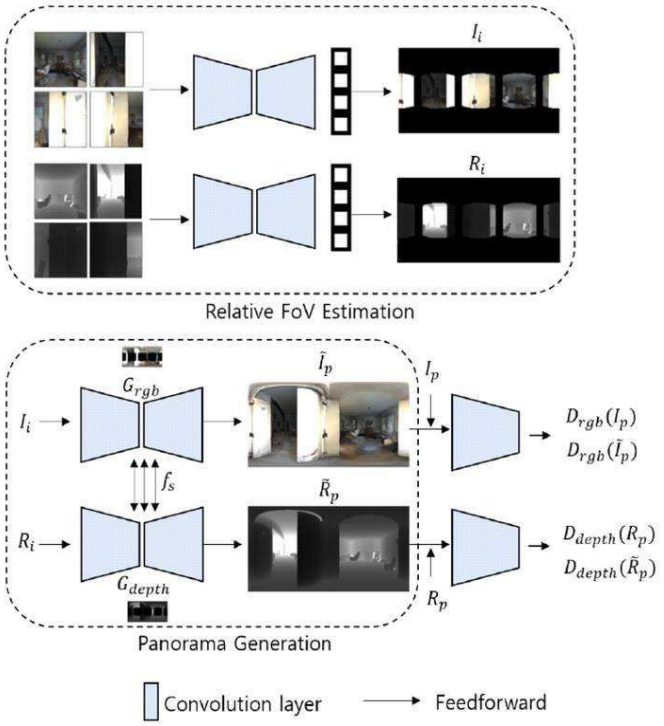
[0127] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수개 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0128] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

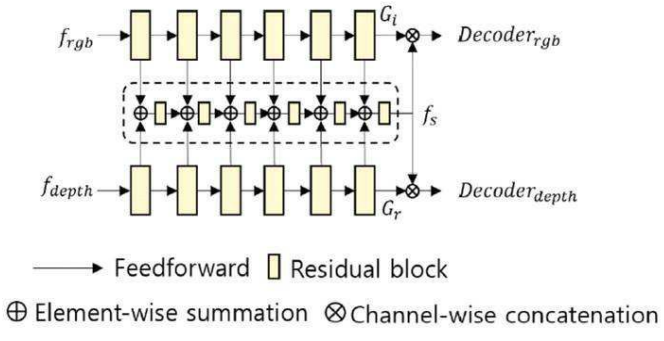
[0129] 그러므로, 다른 구현들, 다른 실시예들 및 청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

도면

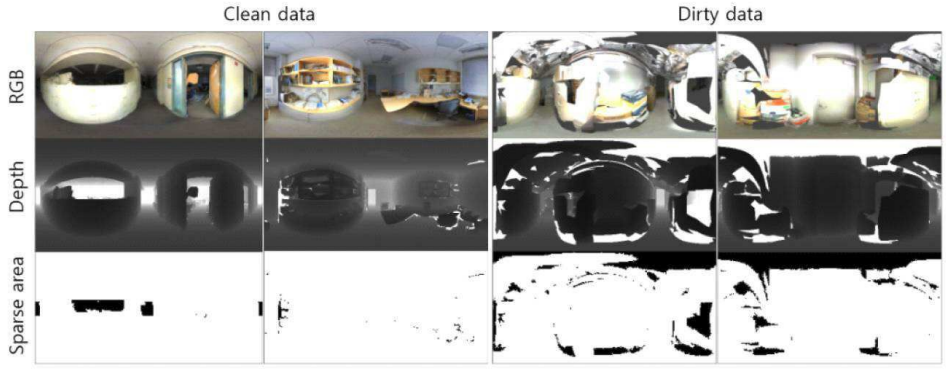
도면1



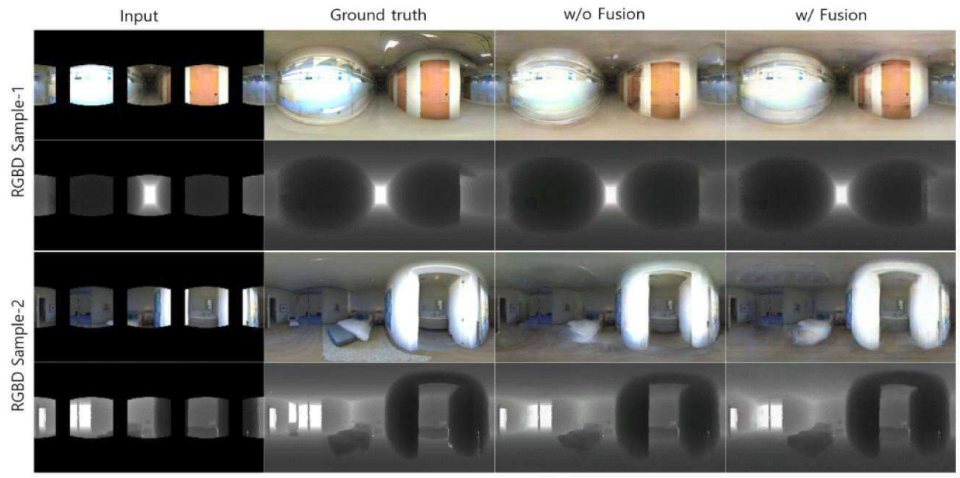
도면2



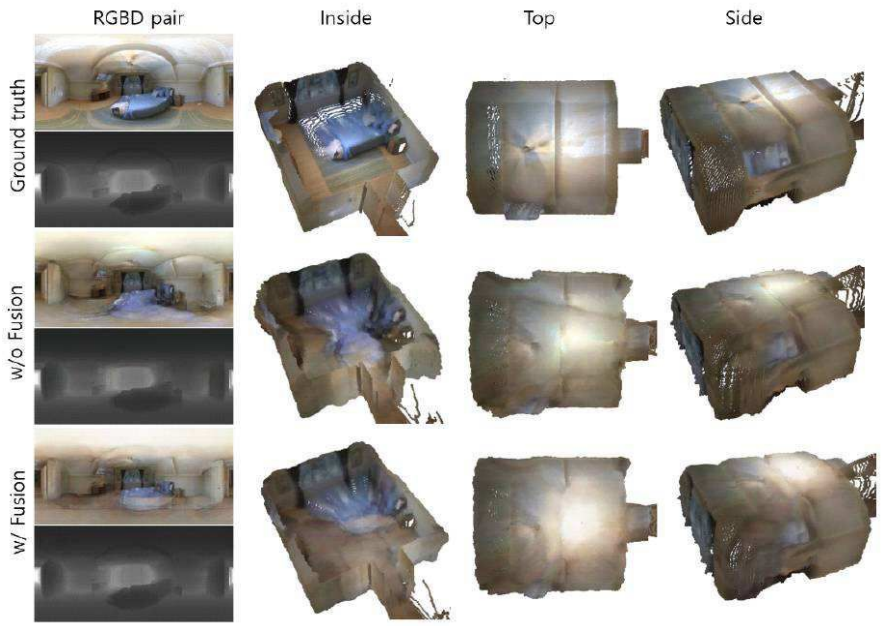
도면3



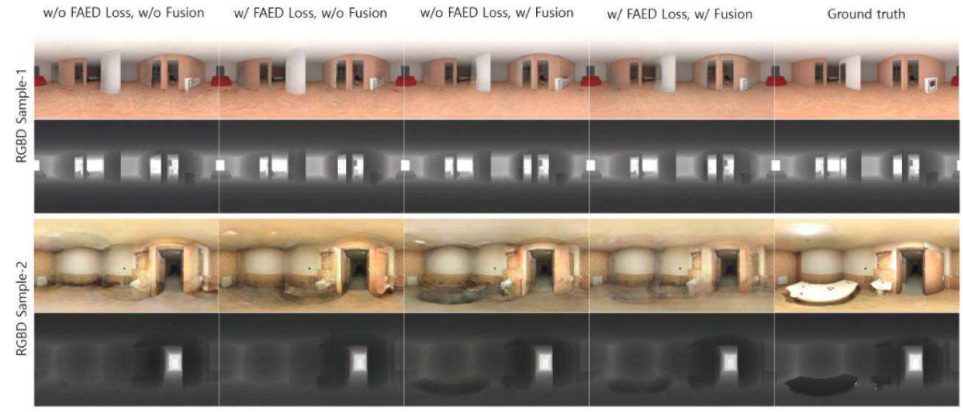
도면4


Input
Ground truth
w/o Fusion
w/ Fusion
RGBD Sample-1
RGBD Sample-2

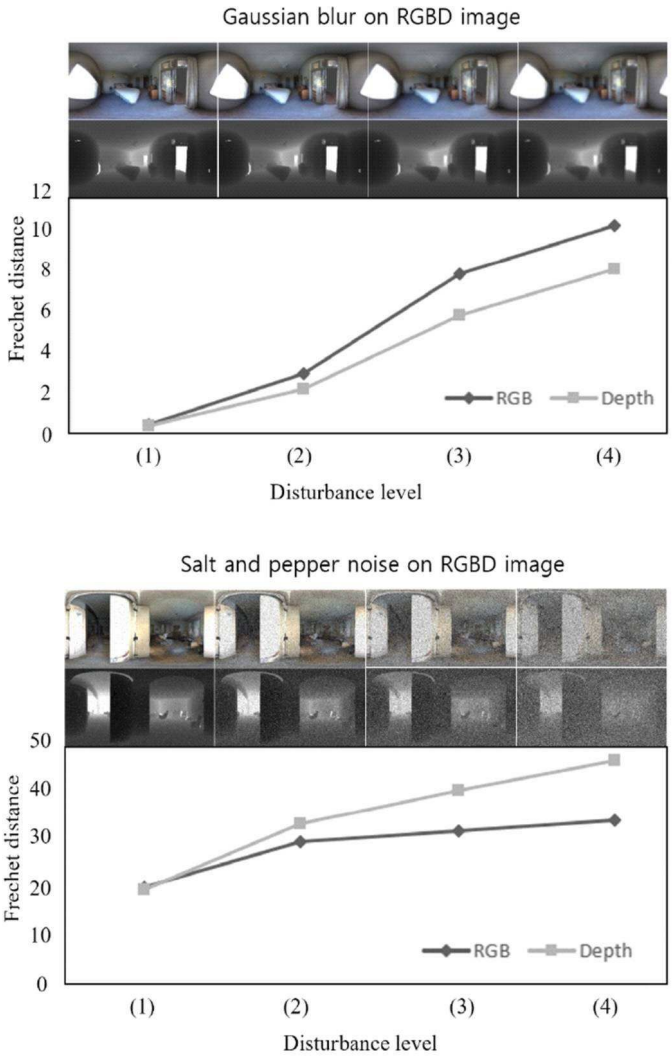
도면5


RGBD pair
Inside
Top
Side
Ground truth
w/o Fusion
w/ Fusion

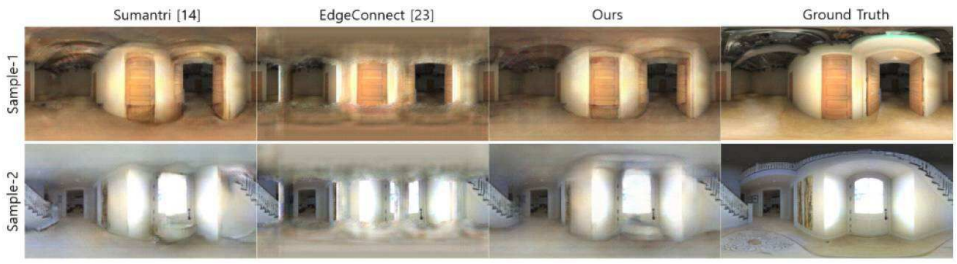
도면6


w/o FAED Loss, w/o Fusion
w/ FAED Loss, w/o Fusion
w/o FAED Loss, w/ Fusion
w/ FAED Loss, w/ Fusion
Ground truth
RGBD Sample-1
RGBD Sample-2

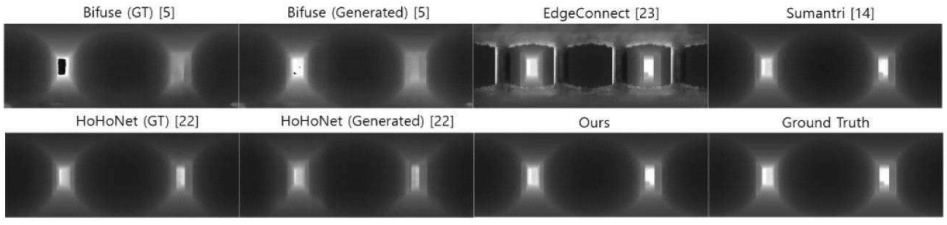
도면7


Gaussian blur on RGBD image
Frechet distance
12
10
8
6
4
2
0
RGB
Depth
(1)
(2)
(3)
(4)
Disturbance level
Salt and pepper noise on RGBD image
Frechet distance
50
40
30
20
10
0
RGB
Depth
(1)
(2)
(3)
(4)
Disturbance level

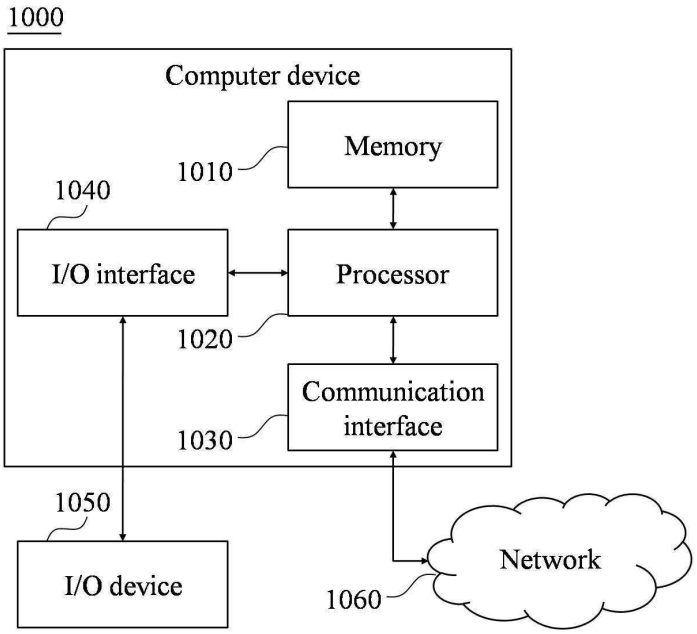
도면8


Sumantri [14]
EdgeConnect [23]
Ours
Ground Truth
Sample-1
Sample-2

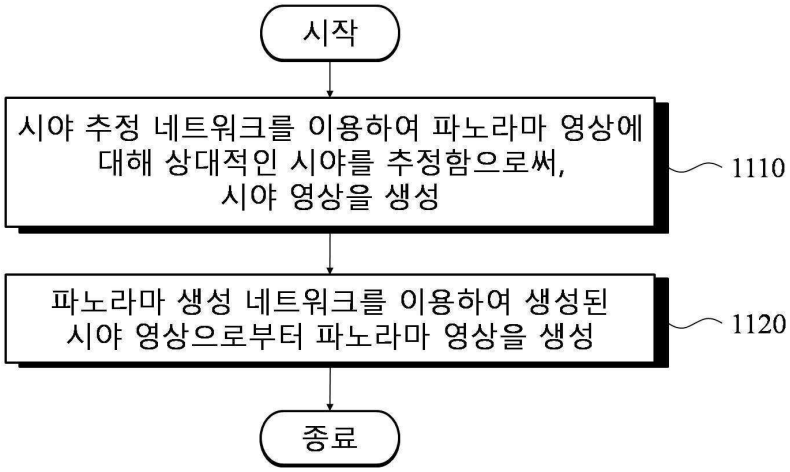
도면9


Bifuse (GT) [5]
Bifuse (Generated) [5]
EdgeConnect [23]
Sumantri [14]
HoHoNet (GT) [22]
HoHoNet (Generated) [22]
Ours
Ground Truth

도면10


1000
Computer device
Memory
1010
1040
I/O interface
Processor
1020
Communication interface
1030
1050
I/O device
Network
1060

도면11


시작
시야 추정 네트워크를 이용하여 파노라마 영상에 대해 상대적인 시야를 추정함으로써, 시야 영상을 생성
1110
파노라마 생성 네트워크를 이용하여 생성된 시야 영상으로부터 파노라마 영상을 생성
1120
종료