

NeRFBot

로봇을 활용한 3D Reconstruction 데이터 취득

1 배경 및 목적 Introduction



NeRF1(Neural Radiance Fields)란 2D 이미지를 학습해 어떤 각도에서든 사실적인 장면을 생성하는 3D Reconstruction 기법 모델이며 해당 모델의 데이터 수집의 과정에는 많은 인적자원과 시간을 필요로 하며 이를 해결하기 위해 데이터 수집 자동화의 필요성을 느낀

추가로 사람이 촬영할 경우 이동중에 카메라 흔들림과 불규칙한 속도 등으로 모션 블러가 발생해 3D 재구성 품질을 저하시킬 수 있어 로봇을 활용한다면 데이터 수집 중에 생기는 부정적인 영향을 줄일 수 있을 것으로 생각

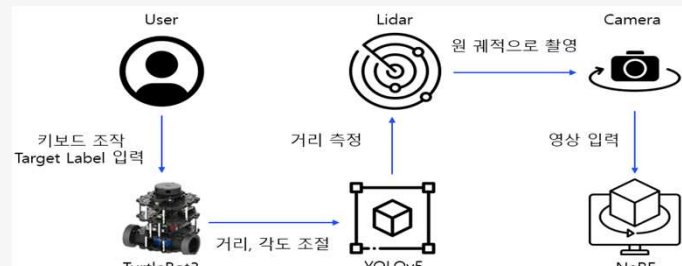
이에 로봇과 사람이 각각 **NeRF 입력 영상 제작에 최적의 카메라 궤적**인 아크 궤적을 포함한 원 형태의 궤적으로 촬영하여 데이터를 수집했을 때 성능에 어떤 영향을 미쳤는지에 대해 알아보려고 함

[그림1] (a) 일관적이지 않은 입력 영상의 결과물 (b) 비교적 일관된 입력 영상의 개선된 결과물



[그림1]

2 기본 동작 원리 | Mechanism

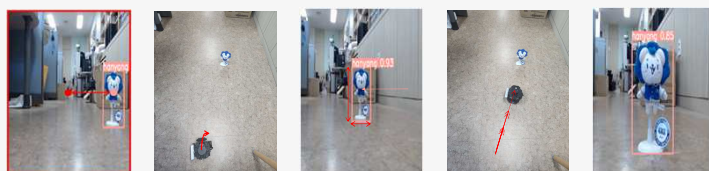


[그림2] 프로젝트의 기본 동작 Mechanism

3 상세 내용 | Details



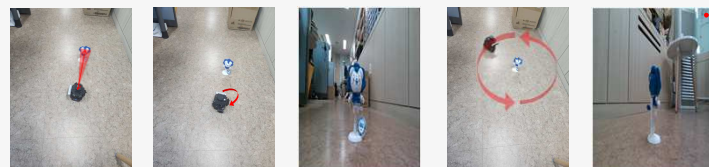
거리, 각도 조절



[그림3] YOLOv5를 이용한 TurtleBot과 Target Object 간의 거리 및 각도를 조절하는 과정

- 키보드로 원격조종[3] 및 촬영할 Object의 Target Label 입력
- Target Object가 카메라 프레임의 중심에 위치하도록 카메라 프레임의 중심 x좌표와 Bounding Box의 중심 x좌표 간의 차이를 기준으로 로봇을 회전
- Bounding Box를 기준으로 높이나 가로가 설정 범위 내의 길이가 되도록 로봇을 앞, 뒤로 이동

데이터 수집



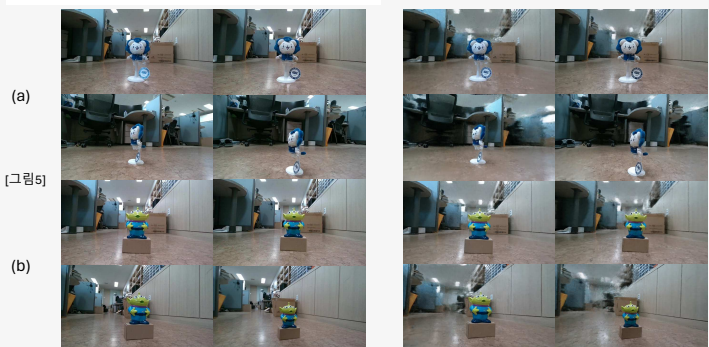
[그림4] Lidar 센서로 Target Object와의 거리를 측정 후, 측정된 거리를 기준으로 원을 그리며 영상을 촬영하는 과정

- Lidar 센서를 이용하여 현재 Target Object와의 거리를 측정
- 로봇을 회전시킨 뒤 피사체를 중심으로 원을 그리며 측면 카메라로 영상을 녹화

모델 결과

- NeRF 모델 중 Instant-NGP[4] 모델을 사용
- COLMAP을 통해 NeRF에 필요한 Camera Parameter와 image 데이터 셋 생성
- 전처리된 데이터를 가지고 Instant-NGP 실행

- 사용 모델 : Instant-NGP
- 카메라 : RealSense D455
- 해상도 : 1280 x 720
- GPU : RTX2060



[그림5] 촬영한 영상을 Input으로 넣어 NeRF 모델에 학습한 결과물 (a) 학년이 (b) 알린

4 결과 및 분석 | Results

평가방법

- 로봇과 사람이 같은 원형 궤적으로 촬영한 각 입력 영상에서 이미지 50, 100, 200장을 추출하여 각 이미지를 NeRF 모델로 학습해 렌더링 된 이미지를 생성
- 로봇이 취득한 데이터와 사람이 취득한 데이터가 Reconstruction 한 결과에 미치는 영향을 확인하기 위해 학습한 이미지와 원본 이미지와의 평가를 진행
- 평가지표로는 이미지 간의 유사도를 평가하는 지표인 *PSNR, *SSIM, *LPIPS, *R/E를 사용

* PSNR (Peak Signal-to-noise ratio) : 신호가 가질 수 있는 최대 전력에 대한 잡음의 전력, 영상 화질 손실 정도를 평가할 때 사용.

* SSIM (Structural Similarity Index Measure) : 주어진 2개의 이미지의 similarity(유사도)를 계산하는 속도

* LPIPS (Learned Perceptual Image Patch Similarity) : 비교할 2개의 이미지를 각각 Network(Alex)에 넣고, 중간 layer의 feature값들을 빼내서 2개의 feature 유사도 측정

* R/E (Reprojection Error) : 3D Point를 2D 이미지에 재투영한 결과 2D 이미지의 특징점 사이의 픽셀 거리가

결과

평균		로봇				사람			
		PSNR↑	SSIM↑	LPIPS↓	R/E↓	PSNR↑	SSIM↑	LPIPS↓	R/E↓
하양이	50장	24.79	0.792	0.25	0.046	25.19	0.802	0.356	0.050
	100장	23.56	0.788	0.29	0.053	20.87	0.739	0.505	0.092
	200장	21.39	0.759	0.32	0.068	22.59	0.773	0.436	0.069
알린	50장	23.40	0.773	0.25	0.053	23.76	0.784	0.433	0.061
	100장	22.13	0.756	0.26	0.063	22.20	0.765	0.455	0.075
	200장	22.98	0.765	0.23	0.057	24.79	0.764	0.475	0.078

[표1] 각 50장, 100장, 200장 별로 NeRF 영상과 원본 영상의 유사도를 평가하는 정량적 척도인 PSNR, SSIM, LPIPS, R/E 측정값의 평균

분산		로봇				사람			
		PSNR	SSIM	LPIPS	R/E	PSNR	SSIM	LPIPS	R/E
하양이	50장	2.11276	0.00101	0.00071	0.00004	8.46065	0.00130	0.00112	0.00013
	100장	3.82942	0.00122	0.00090	0.00008	6.58422	0.00283	0.00134	0.00051
	200장	2.59669	0.00148	0.00069	0.00011	9.54497	0.00180	0.00128	0.00028
알린	50장	2.56842	0.00136	0.00027	0.00007	4.84659	0.00165	0.00187	0.00012
	100장	3.26278	0.00171	0.00048	0.00013	8.14673	0.00222	0.00149	0.00032
	200장	3.06361	0.00157	0.00047	0.00013	7.63344	0.00218	0.00177	0.00031

[표2] 각 50장, 100장, 200장 별로 NeRF 영상과 원본 영상의 유사도를 평가하는 정량적 척도인 PSNR, SSIM, LPIPS, R/E 측정값의 분산

- 평균값에서는 LPIPS와 R/E 값에서는 로봇의 영상이 꾸준히 좋은 결과를 보이지만, PSNR과 SSIM의 경우 사람의 영상이 더 높게 나오기도 하며 차이가 크지 않음
- 그러나 분산값에서는 다소 차이를 보이며, 모든 수치에서 로봇을 활용한 데이터 품질의 편차가 적음을 알 수 있음

5 결론 | Conclusion

- 로봇을 이용한 데이터 수집 자동화를 통해 사람의 노동력을 줄일 수 있음
- 피사체를 중심으로 일정한 속도와 거리를 유지하며 촬영하기에 모든 프레임에서 객체가 잘 중심되게 영상에 담겨 높은 신뢰성을 제공
- 로봇으로 수집한 데이터는 사람이 수집한 데이터에 비해 흔들림이 적어 학습 결과물에도 긍정적인 영향을 미친다는 것을 실험을 통해 검증
- 모든 지표에 대해 분산 값이 적은 것을 통해 일관된 품질의 결과물을 생성했다는 것을 확인할 수 있었으며, 특히 LPIPS와 R/E값 기준으로는 각각 39.28%, 17.66% 개선

개선방안

- 강화학습을 통해 로봇의 데이터 수집 전략을 최적화하여 데이터를 수집
- 자기지도학습으로 변형되거나 다양한 환경에서의 효율적인 Target Label 탐지 능력 개선

참고문헌 | References

- [1] Ben M, Pratul P, S. Matthew T., Jonathan T. B., Ravi R and Ren N. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," ECCV 2020
- [2] Junekoo Kang, Kyeongmin Yu, Bosung Sonn, "NeRF 배경 장면 제작을 위한 최적의 카메라 궤적," 한국인공지능학회 2023
- [3] ROBOTIS. TurtleBot3. GitHub repository. <https://github.com/ROBOTIS-GIT/turtlebot3>
- [4] Thomas M., Alex E, Christoph S. and Alexander K. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," ACM Trans. Graph (SIGGRAPH). 2022

