

열화상 카메라를 이용 실내 깊이 데이터셋 제작 및 딥러닝 기반 깊이 이미지 추정

이동우¹, 조의현¹, 이현범^{2*}

경북대학교 전자전기공학부¹, 아주대학교 기계공학과²

Deep-learning-based Depth Image Estimation Using Thermal Camera With Dataset Creation

Dongwoo Lee¹, Euihyeon Cho^{2*}, Hyeonbeom Lee^{1,2}

Key Words : Deep-learning(딥러닝), Depth Estimation(깊이 추정), Monocular Camera(단안 카메라) Robot Operating System(로봇구동시스템)

서론

최근 단안 카메라를 활용한 깊이 추정(MDE)기술은 드론과 같은 로보틱스 분야의 연구에 중요한 연구 주제로 부상하고 있다

그러나, 기존의 단안 카메라 깊이 추정 알고리즘에 쓰이는 연구의 데이터셋 대부분은 NYU dataset[1]과 같은 표준 RGB 이미지와 깊이 이미지로만 이루어져 있어, 어둡거나 연무가 낀 재난 현장과 같은 특별한 환경에서의 깊이 추정에 제한이 있을 수 있다.

이러한 문제점에 대한 해결책으로, 본 논문은 열화상 이미지의 활용을 통한 깊이 추정 방안을 제시한다. 새로운 데이터셋인 KNUdataset은 세 가지 다른 카메라 이미지를 Homography를 이용하여 정합시켰다. 이를 깊이 추정에서 많이 쓰이는 기본 구조인 U-Net을 기반으로 하는 DenseDepth[2] 모델을 사용해 학습하고 추론하였다. DenseDepth에서 NYUdataset과 KNUdataset을 각각 학습에 사용했을 때 정확도를 비교하였다.

본론

카메라 보정과 정합

RGB와 depth 카메라는 Microsoft의 Azure kinetic dk(4096x3072) 카메라를 사용하였다. Thermal 카메라는 FLIR boson640(640x512)을 사용하였으며, 이를 kinetic 카메라 상단에 부착하여 이미지 데이터 구축을 위한 시스템을 구축하였다.

먼저, 보정(Callibration)과정에서 체크보드 판을 이용해 각 카메라의 내부 파라미터와 왜곡 계수를 계산한 후 각 카메라의 왜곡현상을 해결하였다.



Fig. 1. Align RGB & Thermal using Homography

열화상 카메라의 경우는 체크보드 판을 햇빛에 장시간 노출시킨후 흰색과 검은색의 온도차이를 이용해 체크보드의 코너점을 찾아 보정작업을 진행했다.

이후 정합(Align)과정에서, RGB와 depth 카메라는 동일한 카메라 모듈에서 ROS를 이용해 정합된 영상을 제공하지만 열화상 카메라의 경우 다른 카메라 모듈이므로, 추가적인 정합이 필요하였다. 이를 위해 열화상 카메라와 나머지 카메라 모듈의 위치를 고정시킨 후, 동일한 물체를 촬영해 SIFT⁽³⁾ 알고리즘을 이용해 특징점을 생성하였다. 각 카메라 이미지에서 특징점을 생성한 후 특징점 벡터의 distance가 작은(유사도가 큰) 좋은 매칭점을 선별하여 RANSAC⁽⁴⁾ 알고리즘을 이용해 두 이미지간의 Homography(변환 행렬)를 구하고 이를 Fig. 1.와 같이 적용해 열화상 카메라와 나머지 카메라간 정합을 완성했다.

데이터셋 제작

Densedepth⁽²⁾의 input 크기는 640x480 이므로, 이를 맞추기 위해 열화상과 RGB image중 FOV가 상대적으로 작은 열화상 카메라를 기준으로 이미지가 나오지 않은 가장자리(Fig. 1.의 검은색 부분)를 잘라낸 후, 640x480 사이즈로 resize시켜 사용하였다.

이후 열화상, RGB, 그리고 Depth 카메라를 ROS를 이용해 실행시킨 후 사진을 실시간으로 Homography를 적용하고 저장할 수 있는 ROS 노드를 만들고 데이터셋을 제작하였다. 학습용으로 800장, 테스트용으로 100장의 사진을 촬영했다. 이때 테스트용 사진은 밝은 환경 50장, 어두운 환경 50장을 촬영하여 결과가 밝은 환경, 어두운 환경 어느 쪽에도 극단적으로 치우치지 않게 고려하였다.

학습설정

50 epoch, batch size는 2, learnig rate는 0.0001로 설정하여 학습을 진행하였다. 사용한 모델은 Densedept의 모델 구조를 그대로 사용하였다.

$$\max\left(\frac{y_p}{\hat{y}_p}, \frac{\hat{y}_p}{y_p}\right) = \delta_i < 1.25^i \quad (1)$$

Table 1. Comparison of Result Accuracy

Train dataset & input	δ_1	δ_2	δ_3
NYUdataset & RGB	0.240	0.488	0.635
NYUdataset & thermal	0.309	0.511	0.637
KNUdataset & thermal	0.520	0.679	0.747

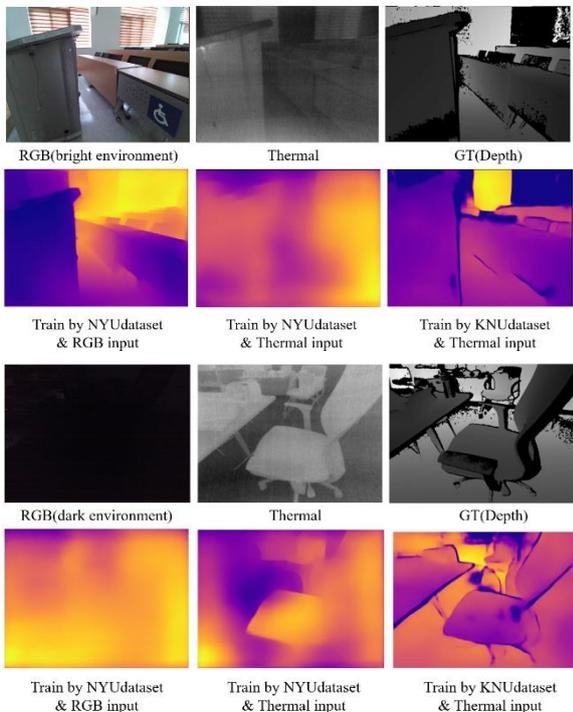


Fig. 2. Comparison of different train dataset & input inference result (bright/dark environment)

학습 결과 분석

Table 1은 기존의 NYUdataset⁽¹⁾ 과 본 논문에서 제작한 KNUdataset을 Densedept⁽²⁾ 네트워크 구조에 각기 훈련시킨 모델에 RGB, 열화상 input을 넣었을 때의 정확도를 비교한 결과이다. 정확도는 수식 (1)의 δ_i 를 이용해 측정하였다. y_p 는 Ground Truth, \hat{y}_p 는 모델에서 추론한 픽셀 값이다. KNUdataset을 학습한 모델에 열화상 input을 적용 시 기존의 NYU dataset을 학습한 모델에 RGB input을 적용하였을 때 보다 116% 높은 정확도가 나왔다. 또한 Fig.2.과 같이 식별이 어려운 어두운 환경에서 열화상 image를 KNUdataset에 학습된 모델로 추론하였을 경우 가장 정확한 결과를 얻는 것을 확인할 수 있다.

결론

본 논문의 결과는 Fig.2에서 볼 수 있듯이 깊이 추정을 위해 열화상 input을 사용하면 어두운 재난 현장에서도 더 나은 깊이 추정이 가능하다는 것을 보여준다. 또한 추후 연구에서 네트워크 구조를 수정하거나 RGB, 열화상 image를 동시에 input으로 사용하면 정확도가 더욱 향상될 것이라고 기대한다.

후기

This work was supported in by the Unmanned Vehicle Advanced Research Center (UVARC) by the Ministry of Science and ICT, Republic of Korea under Grant NRF-2020M3C1C1A01086411

참고문헌

- 1) Silberman, Nathan, et al. "Indoor segmentation and support inference from rgb-d images." *European Conference on Computer Vision(ECCV)*, 2012. pp 746-760.
- 2) I. Alhashim, and P. Wonka, "High Quality Monocular Depth Estimation via Transfer learning" *arXiv preprint*, arXiv:1812.11941.
- 3) David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints" *International Journal of Computer Vision* 2004, pp. 91-110.
- 4) Fischler, Martin A., and Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." *ACM*, (1981), pp.381-395