

# 인공 신경망 기반의 위치 추정과 이미지 생성을 통한 다중 로봇의 지도 병합\*

김재인<sup>01</sup> 이현도<sup>2</sup> 이강훈<sup>3</sup> 김윤성<sup>2</sup> 한동식<sup>2</sup> 장병탁<sup>2</sup>

<sup>1</sup>서울대학교 협동과정 뇌과학전공

<sup>2</sup>서울대학교 컴퓨터공학부

<sup>3</sup>서울대학교 협동과정 인지과학전공

[jkim@bi.snu.ac.kr](mailto:jkim@bi.snu.ac.kr) [hdlee@bi.snu.ac.kr](mailto:hdlee@bi.snu.ac.kr) [khlee@bi.snu.ac.kr](mailto:khlee@bi.snu.ac.kr) [yskim@bi.snu.ac.kr](mailto:yskim@bi.snu.ac.kr)

[dshan@bi.snu.ac.kr](mailto:dshan@bi.snu.ac.kr) [btzhang@bi.snu.ac.kr](mailto:btzhang@bi.snu.ac.kr)

## Multi-robot Map Merging with Pose Estimation and Image Generation based on Neural Network

Jaein Kim<sup>01</sup> Hyundo Lee<sup>2</sup> Ganghun Lee<sup>3</sup> Yoonsung Kim<sup>2</sup> Dong-Sig Han<sup>2</sup> Byoung-Tak Zhang<sup>2</sup>

<sup>1</sup>Interdisciplinary Program in Neuroscience, Seoul National University

<sup>2</sup>School of Computer Science and Engineering, Seoul National University

<sup>3</sup>Interdisciplinary Program in Cognitive Science, Seoul National University

### 요 약

지도 병합 문제는 오랫동안 주요한 다중 로봇 시스템의 문제 중 하나이지만, 기존의 연구들은 병합하는 지도 간의 공통 정보가 충분히 많아야 사용 가능 하거나 특수한 하드웨어를 요한다는 한계를 갖는다. 본 논문에서는 다양한 인공 신경망 모델들을 사용해 상기한 한계들을 극복해 로봇 간의 상대 위치를 추정하고 병합된 지도의 품질을 높이는 시스템을 제안한다. 제안하는 시스템은 1) RGB 이미지에서 로봇의 상대 위치를 추정하고, 2) 변환된 지도 이미지에서 상대 위치의 오차를 회귀하고, 3) 병합된 지도 이미지의 노이즈를 제거하는 3개의 인공 신경망 모델로 구성된다. 모델들에 대한 양적 평가, 시뮬레이션 환경에서의 실험을 통해 제안된 시스템이 성공적으로 지도를 병합할 수 있음을 확인했다.

### 1. 서론

다중 로봇 SLAM에서의 지도 병합은 동일한 환경에서 독립적으로 행동하는 복수의 로봇들의 부분 지도를 하나의 지도로 병합하는 문제이다. 로봇의 부분 지도의 종류나 추가 정보에 따라 지도 병합 문제는 다양한 접근법이 존재하나[1], 일반적으로 로봇들의 시작 위치가 불확실함을 전제한다[2].

기존의 연구들 중에는, 서로 다른 시작 위치를 갖는 각 로봇들의 점유 격자 지도(Occupancy Grid)에서 비전 알고리즘을 사용해 각 개체의 부분 지도들 사이의 최적 변환을 찾는 연구들이 존재한다[3][4]. 하지만 지도 정보만 이용해 변환을 찾는 기존 연구들은 지도들이 충분한 겹치는 영역을 가져야 한다는 한계가 있다[2]. 지도 정보 외에 복수의 로봇 간 상대 위치를 측정해 이용하는 연구의 경우 로봇들이 특수한 전방향 센서와 마커가 장비되어 있음을 가정했으나[5], 이러한 가정은 하드웨어의 제약으로 인한 한계를 갖는다.

본 연구는 기존 지도 병합 연구의 한계를 극복하기 위해,

모바일 로봇이 범용적인 RGB 카메라로 얻은 이미지에서 인공 신경망 모델을 사용해 다른 로봇의 상대 위치를 추정하고 지도 병합에 이용하는 시스템을 설계했다. 해당 시스템은 병합 지도의 품질 향상을 목표로, 위치 추정 외에도 로봇 지도 사이의 방향 오차를 회귀하고 병합 지도의 노이즈를 제거하는 인공 신경망 모델들을 사용한다. 개별 모델들에 대한 양적 평가와 가상 환경에서의 시나리오 수행 결과, 제안한 시스템이 향상된 성능의 지도 병합을 수행할 수 있음을 확인했다.

### 2. 연구 방법

본 연구의 시스템은 로봇  $R_0$ 와  $R_1$ 에 대해, 각 로봇의 지도 이미지  $M_0, M_1$ , 각 로봇의 자신의 지도에서의 위치  $\theta_0, \theta_1$ ,  $R_0$ 가  $R_1$ 을 포착한 RGB 이미지  $I$ 를 입력 받아,  $M_0$ 을 기준으로 병합된 지도  $\hat{M}$ 을 출력한다. 시스템의 모든 위치 변환은 아핀 변환이며, 제안하는 시스템의 의사 코드는 알고리즘 1과 같다.

\* 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(2015-0-00310-SW.StarLab, 2017-0-01772-VTT, 2018-0-00622-RMI, 2019-0-01367-BabyMind), 한국산업기술평가원(P0006720-GENKO)의 지원을 받았다.

## 알고리즘 1: 상대 위치 추정 지도 병합 시스템

입력:  $M_0, M_1, \theta_0, \theta_1, I$

출력:  $\hat{M}$

- 1:  $\theta_{relative} \leftarrow PoseEstimation(I)$  // 상대 위치 추정 모델
- 2:  $T \leftarrow \theta_0 \theta_{relative} (\theta_1)^{-1}$  //  $M_1$ 에서  $M_0$ 로의 변환
- 3:  $\tilde{M}_1 \leftarrow T(M_1)$
- 4:  $R \leftarrow OrientationError(M_0, \tilde{M}_1)$  // 방향 오차 회귀 모델
- 5:  $\hat{M}_1 \leftarrow R^{-1}(\tilde{M}_1)$
- 6:  $M_0$ 와  $\hat{M}_1$ 을  $\tilde{M}$ 로 병합
- 7:  $\hat{M} \leftarrow DenoisingAutoencoder(\tilde{M})$  // 노이즈 제거 모델
- 8: return  $\hat{M}$

### 2.1 RGB 이미지에서 로봇의 상대 위치 추정

RGB 이미지  $I$ 에 투영된 로봇의 자세를 추정하는데 DOPE[6] 모델이 사용되었다. 해당 모델은 본래 로봇 암의 pick-and-place 문제를 해결하기 위해 카메라로부터 물체까지의 6DOF 자세( $x, y, z, roll, pitch, yaw$ )를 추정하는 모델이다. 본 연구에서는 2차원 평면에서의 지도 병합에 필요한  $x, y$ 축에 대한 평행이동과  $z$ 축에 대한 회전 값을 사용한다. [6]의 데이터 생성 방법을 참조해 본 연구에서도 가상 환경에서 로봇이 포착된 이미지를 생성해 훈련에 이용했다.



그림 1. DOPE 훈련 이미지의 예시. 왼쪽은 임의의 배경과 장애물에 임의의 자세를 갖는 로봇의 이미지이고, 오른쪽은 Falling Things[7] dataset을 참조해 현실적인 물리를 적용하여 낙하하는 로봇을 포착한 photorealistic 이미지이다

### 2.2 분리된 지도 이미지 데이터 생성

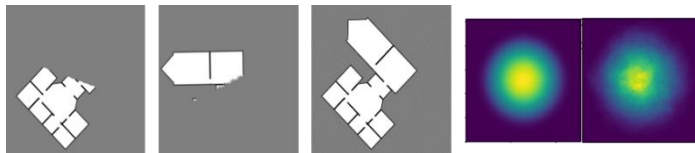


그림 2. 왼쪽부터 순서대로  $M_0, M_1, M$ 의 예시이며, 맨 오른쪽이 두 지도의 픽셀 누적 분포의 히스토그램이다. 왼쪽이 생성 지도, 오른쪽이 SLAM으로 얻은 지도의 히스토그램이다.

지도 이미지를 병합하는 인공 신경망 모델을 훈련하려면, 동일 환경의 다른 시작점에서 두 로봇이 SLAM을 수행해 각각 얻은,  $M_0, M_1, \hat{M}$ 의 label인  $M$ 으로 구성된 set이 필요하다. 본 연구에서는 HouseExpo[8] 데이터의 실내환경 지도를 임의로 나누어 총 103573개 set을 생성했다. 그림 2는 생성된

set의 예시와, 생성 데이터와 가상 환경에서 SLAM으로 얻은 지도 325장의 비교를 위한 픽셀 누적 분포의 히스토그램이다. 회색 픽셀을 제외한 두 히스토그램의 상관관계는 0.9898이다.

### 2.3 두 지도의 방향 오차 회귀

$\theta_{relative}$ 가 지도 변환에 사용될 때, 평행 이동의 오차보다 회전의 오차가 지도 병합의 품질에 미치는 영향이 크다. 따라서  $M_0$ 와  $\hat{M}_1$ 을 입력 받아 두 지도 사이의 변환의 회전 오차를 추정하는 모델이 필요하다.

자연 이미지의 회전 각도를 회귀하는 인공 신경망 모델의 성능이 기존의 컴퓨터 비전 알고리즘보다 우수한 것이 알려져 있다[9]. 본 연구도  $M_0$ 와  $\hat{M}_1$ 이 연결된 2차원 채널 이미지를 입력 받아 방향 오차의 scalar 값을 회귀하는 단순한 CNN 모델을 사용했으며, 방향 오차는 매 학습마다  $[-30^\circ, 30^\circ]$ 의 구간에서 균일 추출했다.

### 2.4 Denoising Autoencoder 모델로 병합 지도 처리

$\hat{M}_1$ 과  $M_0$ 을 병합한  $\tilde{M}$ 은  $\theta_{relative}, M_0, M_1$ 로 인한 오차가 존재한다. Denoising Autoencoder는 매 학습마다 확률적으로 손상된 입력에서 원본 입력을 복원[10]하는 자기 지도 학습을 수행한다. 본 연구는  $\theta_{relative}$ 의 label에, 이미지의 변의 길이가  $L$ 일때  $[-0.025L, 0.025L]$ 구간에서 평행이동 오차와  $[-10^\circ, 10^\circ]$ 구간에서 회전 오차를 균일 추출해 적용한다.

또한 VQ-VAE-2[11]의 아키텍처를 사용하여 autoencoder가 이산적이고 계층적인 latent variable을 학습하도록 했다. 그리고 복원된 이미지의 품질 향상을 위해 Structural Similarity Index Measure(SSIM)[12]가 복원 손실 함수로 사용되었다. 이는 인공 신경망에서 이미지 복원 시 SSIM이 L2 Norm보다 이미지의 품질이 높기 때문이다[13]. Encoder, decoder를  $E, D$ , 이산화된 latent variable을  $e_j$ 라 했을 때, 손실 함수  $\mathcal{L}$ 은 아래와 같이 복원 오차를 줄이고 latent variable을 이산 공간으로 사상한다.

$$e_k = \operatorname{argmin}_j \|E(\tilde{M}) - e_j\|, \hat{M} = D(e_k)$$

$$\mathcal{L}(M, \tilde{M}, \hat{M}) =$$

$$1 - SSIM(\hat{M}, M) + \|sg[E(\tilde{M})] - e_k\|_2^2 + \beta \|sg[e_k] - E(\tilde{M})\|_2^2$$

## 3. 실험 결과

### 3.1 상대 위치 추정 모델 평가

DOPE의 pre-trained weight를 생성된 로봇 이미지 80000장에 대해 추가로 훈련한 후 test set 이미지 2000장에 대해  $x, y$ 축 평행이동과  $z$ 축 회전의 오차의 절대값을 평가했다. 그 결과, 모델은  $x, y$ 축 평행이동에 대해 각각 평균 0.38m, 0.15m, 표준편차 16.69, 11.44의 오차를 가지며,  $z$ 축 회전에 대해 평균 16.03°, 표준편차 16.62의 오차를 가졌다.

### 3.2 지도 병합 모델의 비교 평가

방향 오차 회귀 모델과 병합 지도의 노이즈 제거 모델을 연결하여, 생성한 지도 데이터의 test set에 대해 성능을 평가했다.  $\theta_{relative}$ 의 오차는 훈련과 같은 조건이며, 방향 오차 회귀 모델, VQ-VAE-2 구조, SSIM 손실 함수의 사용 여부에 따라 구분해 성능을 평가했다. VQ-VAE-2 구조를 사용하지 않은 모델은 latent variable의 차원이 같은 단순한 autoencoder 구조를 사용했으며,  $\hat{M}$ 과  $M$ 의 픽셀 값을 이산화해 픽셀 단위의 정확도를 측정했다. 이때  $\hat{M}$ 에서 회색 영역의 true positive 픽셀은 평가에서 제외하여 지도 대부분을 차지하는 unknown 영역에 대한 과대평가를 방지했다. 표 1과 같이, 제안한 모델의 정확도가 78.82%로 가장 높았다. 또한 그림 3은 test set의 한 예시에 대한 제안한 모델의 중간 과정과 결과를 보여주고 있다.

표 1. 비교 모델들과 제안하는 모델의 정확도이다.

방향오차보정	VQ-VAE-2	SSIM	정확도(%)
X	X	X	69.51
X	X	O	72.55
X	O	X	72.76
X	O	O	75.83
O	X	X	74.14
O	X	O	77.14
O	O	X	77.06
<b>O</b>	<b>O</b>	<b>O</b>	<b>78.82</b>



그림 3. 왼쪽부터  $M$ ,  $M_0$ ,  $M_1$ ,  $M_0$ 와  $\hat{M}_1$ 의 병합 지도,  $\hat{M}$ ,  $\hat{\hat{M}}$ .

### 3.3 가상 환경에서의 로봇 시스템 동작

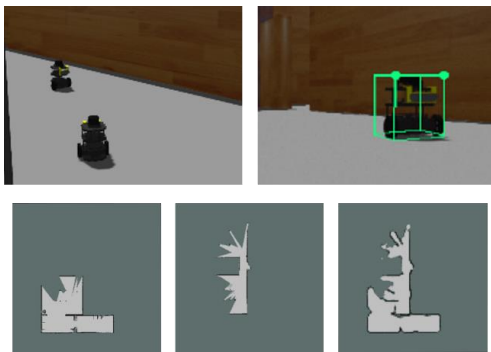


그림 4. 위는 시뮬레이션중인 다중 로봇과 RGB 이미지에 포착된 로봇의 추정된 상대 위치를 나타낸 그림이다. 아래는 위의 조건에서 제안된 시스템의  $M_0$ ,  $\hat{M}_1$ ,  $\hat{\hat{M}}$ 에 해당한다.

ROBOTIS사의 TurtleBot3 Burger 로봇에 Intel사의 RealSense D435 카메라를 부착한 모델을 사용하여 Gazebo 시뮬레이션 환경에서 SLAM을 수행하는 다중 로봇 시스템을 설계하고, 본 연구가 제안한 시스템을 사용해 지도를 병합하는 시나리오를

수행했다. 그림 4와 같이, 제안한 시스템은 가상 환경에서 두 지도 간의 겹치는 영역이 적은 상황에서도 지도를 병합하는 시나리오를 성공적으로 수행했다.

## 4. 결론

본 연구는 각 인공 신경망 모델에 필요한 데이터를 새롭게 생성하여 학습시킨 후, 독립적으로 학습된 모델들을 결합한 시스템으로 가상 환경에서 다중 로봇의 지도 병합의 성능을 향상하는데 기여했다. 또한 본 연구는 기존 연구의 지도 정보에 대한 제약과 하드웨어에 대한 제약을 극복해, 이전보다 범용적인 다중 지도 병합 시스템을 제안했다.

## 참고문헌

- [1] I. Andersone, "Heterogeneous Map Merging: State of the Art," *Robotics*, vol.8, no.3, 74, 2019.
- [2] S. Saedi, M. Trentini, M. Seto, H. Li, "Multiple-Robot Simultaneous Localization and Mapping: A Review," *Journal of Field Robotics*, vol.33, no.1, pp.3-46, 2016.
- [3] S. Carpin, "Fast and accurate map merging for multi-robot systems," *Autonomous Robots*, vol.25, no.3, pp.305-316, 2008.
- [4] V. Ferrao, C. Vinhal, G. da Cruz, "An occupancy grid map merging algorithm invariant to scale, rotation and translation," *2017 Brazilian Conference on Intelligent Systems (BRACIS)*, pp.246-251, 2017.
- [5] X. S. Zhou, S. I. Roumeliotis, "Multi-robot SLAM with Unknown Initial Correspondence: The Robot Rendezvous Case," *2006 IEEE/RSJ international conference on intelligent robots and systems*, pp.1785-1792, 2006.
- [6] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, S. Birchfield, "Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects," *Proceedings of The 2nd Conference on Robot Learning*, pp.306-316, 2018.
- [7] J. Tremblay, T. To, S. Birchfield, "Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.2038-2041, 2018.
- [8] T. Li, D. Ho, C. Li, D. Zhu, C. Wang, M. Q. H. Meng, "HouseExpo: A Large-scale 2D Indoor Layout Dataset for Learning-based Algorithms on Mobile Robots," *arXiv preprint arXiv:1903.09845*, 2019.
- [9] P. Fischer, A. Dosovitskiy, T. Brox, "Image Orientation Estimation Convolutional Networks," *German Conference on Pattern Recognition*, 2015.
- [10] P. Vincent, H. Larochelle, Y. Bengio, P. A. Manzagol, "Extracting and Composing Robust Features with Denoising Autoencoders," *Proceedings of the 25th international conference on Machine learning*, pp.1096-1103, 2008.
- [11] A. Razavi, A. van den Oord, O. Vinyals, "Generating Diverse High-Fidelity Images with VQ-VAE-2," *Advances in Neural Information Processing Systems*, vol.32, pp.14866-14876, 2019.
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol.13, no.4, pp.600-612, 2004.
- [13] H. Zhao, O. Gallo, I. Frosio, J. Kautz, "Loss Functions for Image Restoration with Neural Networks," *arXiv preprint arXiv:1511.08861*, 2015.