



카메라 및 LiDAR 센서 기반의 심층 신경망을 이용한 실내 환경에서의 위치 추정에 관한 기본 연구

A Basic Study on Robot Position Estimation in Indoor Environment using Deep Neural Network based on the Camera and LiDAR Sensors

이정우* · 배기덕* · 최영호* · 김영복** · 서진호***†

Jung-Woo Lee*, Gi-Deuk Bae*, Young-Ho Choi*, Young-Bok Kim**
 and Jin-Ho Suh***†

(Received 07 June 2019, Revision received 12 August 2019, Accepted 12 August 2019)

Abstract: It is important to know its own position in order to the robot to move autonomously in the environment. In the outdoor environment, GPS can be used to determine the position, but in the indoor environment, the position is estimated by complicated calculation by LiDAR or image sensors. Currently, SLAM (Simultaneous localization and mapping) method is used to create a map and estimate the position for the robot navigation. This method is disadvantageous in that there is a risk of divergence in complicated algorithm calculations, the processing time is long and uneven using various sensors having a large amount of information relatively, and it is difficult to know the self-position when starting from any location.

In this paper, we propose a deep neural network model that outputs absolute position using Lidar and image sensor data. The position estimation method using this deep neural network can predict the absolute position immediately at every location, and the calculation of the model inference is executed periodically without divergence. In the experiment, training dataset is builded by acquiring LiDAR data and images in the indoor environment, and it is used to learn the deep neural network model for about one-month. As a result, the top-5 accuracy of the inference is about 98%, and the estimated position error is about 14cm on average.

Key Words : Deep Neural Network, Multimodal Sensors, Simultaneous Localization and Mapping (SLAM), Mobile Robot

*** 서진호(ORCID:<https://orcid.org/0000-0002-8673-4528>) : ***† Jin-Ho Suh(ORCID:<https://orcid.org/0000-0002-8673-4528>) : Professor, Department of Mechanical System Engineering, Pukyong National University.
 교수, 부경대학교 기계시스템공학과
 E-mail : suhgang@pknu.ac.kr, Tel : 051-629-6189
 *이정우(ORCID:<https://orcid.org/0000-0001-9091-1729>) : *Jung-Woo Lee(ORCID:<https://orcid.org/0000-0001-9091-1729>) : Senior Researcher, Field Robot R&D Division, Korea Institute of Robot and Convergence(KIRO).
 선임연구원, 한국로봇융합연구원 필드로봇연구본부
 *배기덕(ORCID:<https://orcid.org/0000-0002-8640-4123>) : *Gi-Deuk Bae(ORCID:<https://orcid.org/0000-0002-8640-4123>) : Assistant Researcher, Field Robot R&D Division, KIRO.
 주임연구원, 한국로봇융합연구원 필드로봇연구본부
 *최영호(ORCID:<https://orcid.org/0000-0003-1576-8277>) : *Young-Ho Choi(ORCID:<https://orcid.org/0000-0003-1576-8277>) : Principal Researcher, Field Robot R&D Division, KIRO.
 책임연구원, 한국로봇융합연구원 필드로봇연구본부
 **김영복(ORCID:<https://orcid.org/0000-0001-6035-6744>) : **Young-Bok Kim(ORCID:<https://orcid.org/0000-0001-6035-6744>) : Professor, Department of Mechanical System Engineering, Pukyong National University.
 교수, 부경대학교 기계시스템공학과

1. 서론

로봇은 현재의 위치에서 목적지로 이동하기 위해 자신의 위치를 추정하는 것이 필요하며, 위치 추정의 높은 정확도가 요구된다. 실외 환경에서는 GPS(global positioning system) 수신을 통한 위치 추정이 가능하나, 실내 환경에서는 GPS를 사용하기 어렵기 때문에 SLAM(Simultaneous localization and mapping)에 의한 방법으로 카메라와 LiDAR(light detection and ranging), 초음파, 적외선 센서와 같은 여러 종류의 센서로부터 획득된 데이터로 위치를 추정할 수 있다.

SLAM의 방법은 여러 센서의 데이터를 정보 교합을 통해 칼만 필터 등의 필터링에 의해 위치를 추정하며, 정보의 부정확한 교합이 발생하거나 필터 실행 시 발산하게 되면 위치 추정의 정확도가 떨어지게 된다. 또한 센서 데이터를 시간 축으로 누적하여 계산하게 되므로 이전 위치를 기준으로 한 상대적 위치를 추정하게 되고, 임의 위치에서 시작할 때 위치 추정이 어려우며, 위치 추정 시에도 이전 위치에서 발생한 오차가 전파되면서 이를 보정하는데 일정 시간이 소요된다.

본 연구에서는, 실내 환경에서 기존 SLAM을 통해 획득되는 위치 정보의 수준을 심층 신경망 모델로 대체하는 방법을 제안한다. 위치 추정을 위한 심층 신경망(Deep neural network) 모델을 사용하여 실내 공간의 임의의 위치에서 각 축별 0.5 m의 해상도를 갖는 2차원 상의 절대 위치 좌표를 출력할 수 있도록 설계하였다.

2. 심층 신경망(Deep Neural Network) 기반 위치 추정 방법의 연구 현황

실내 환경에서의 객체 추적과 자세 추정을 위해 깊은 신경망 모델을 사용한 로봇의 위치 추정 방법에 대한 연구¹⁾에서는, 환경 내 고정된 영상 카메라에서 취득된 영상 데이터를 합성곱 신경망(Convolutional neural network) 모델에 적용하여 로봇의 이동을 추적하고 위치를 추정하고 있다. 그러나 카메라 화각과 렌즈의 굴곡에 의한 경계면

에서의 위치 추정 오차가 커지고, 실내 여러 공간에서의 연속된 위치 추정을 위해서는 많은 시스템 자원이 필요하다는 단점이 있다. 따라서 이동하는 로봇에서 취득된 데이터를 사용하여 위치 추정을 하는 것이 필요하다.

다른 연구²⁾에서는 자율주행차량에 장착된 카메라의 영상 데이터와 시각적 주행거리 측정³⁾(Visual odometry)에 의해 구해진 자세 정보를 이용하여 심층 신경망 모델인 VGG-16을 변형한 합성곱 신경망(CNN) 모델로 학습하여 위치를 추정하고 있으며, 참값이 되는 RTK(real time kinematics) GPS 측정 대비 위치 오차가 약 2.3 m 이고, 시각적 SLAM에 의한 추정 대비 위치 오차가 약 1.2 m이다. 본 연구에서는 이와 유사한 심층 신경망 기반의 위치 추정 방법을 제안하며, 실내 환경에서 더 정밀한 위치값을 제공하는 LiDAR 기반의 SLAM 알고리즘으로 생성된 목표값을 사용하여 심층 신경망 모델의 지도 학습(Supervised learning)을 통해 로봇의 위치를 추정한다.

3. 심층 신경망(Deep Neural Network)을 이용한 실내 환경 위치 추정

3.1 실내 환경의 데이터 취득 구조

실내 환경의 데이터 취득을 위한 모바일 로봇 플랫폼은 다양한 형식(Color, IR(infrared), Night Vision, Thermal, Depth)의 영상 취득이 가능한 카메라들과 LiDAR 센서, GPS 센서 등이 장착된 범용 플랫폼⁴⁾이며, 본 연구에서는 일부 형식의 영상(Color, IR, Depth)과 LiDAR 센서 데이터를 취득하여 사용하고 있다. 데이터 취득 환경은 Fig. 1과 같이 한국로봇융합연구원 내 2층 공간의 복도 환경을 대상으로 하고 있다.

위치 추정을 위한 심층 신경망 모델의 학습을 위해 취득되는 데이터로는 카메라로부터의 컬러, 적외선 및 깊이 정보 등 3종에 대한 영상과 LiDAR 센서로부터 획득된 전방향(Omnidirectional) 거리 정보(3차원 좌표 및 레이저의 강도)이다.

심층 신경망 모델의 지도 학습(Supervised

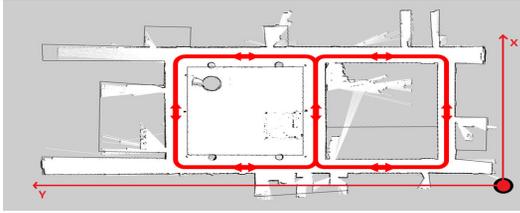


Fig. 1 The starting point of the map(red point) and the trajectory of a mobile robot(red line) in an indoor corridor environment



Fig. 2 An image of the training dataset in which pre-processed data are arranged in a grid ((a) top-left: color image, (b) top-right: LiDAR visualization image, (c) bottom-left: depth image, (d) bottom-right: IR image)

learning)을 위한 목표값이 되는 평면상의 2차원 위치보도 동시에 취득되며, 기존 로봇 시스템에서 사용되고 있는 주행 드라이버의 주행거리(Odometry)와 LiDAR 기반의 SLAM 알고리즘^{6,7)}을 사용하여 로봇의 추정 위치를 구하였고, 이렇게 SLAM으로부터 구해진 목표값은 Fig. 1과 같은 좌표계를 갖는 실내 복도 환경 내에서 실측 대비 평균 약 ± 4 cm의 추정 위치 오차를 갖는 것으로 알려져 있다.⁸⁾

3.2 심층 신경망 모델의 학습을 위한 데이터 셋 구축

심층 신경망 모델의 지도 학습을 위한 데이터

셋은 아래 Table 1과 같이 로봇 플랫폼의 주행을 통하여 실내 환경의 기준점으로부터 X축과 Y축에 대해 0.5 m 단위로 취득된 데이터를 모아 공간 축으로 정렬하고, 각 데이터의 타입별로 전처리 과정을 수행하여 최종 데이터셋을 생성한다. 분류(Classification)를 위한 레이블의 개수는 563개이고 각 레이블은 위치 좌표값을 조합하여 표현하고 있다.

전처리 단계에서는, 16-bit의 깊이 정보(Depth) 영상에 대해 유효 거리를 조정하여 8-bit 이미지로 변환하고, LiDAR 센서의 거리 정보는 0.5도 단위의 전방향에 대해 장애물에 대한 최소 거리를 계산하여 흑백 이미지화 한다.

3.3 위치 추정을 위한 심층 신경망 모델의 구성

위치 추정을 위한 심층 신경망 모델은 여러 타입의 카메라 영상 및 LiDAR 센서의 거리 정보를 전처리한 가시화 영상을 입력으로 하여 일정 거리 단위(0.5 m)의 격자로 구분된 2차원 상의 절대 위치 좌표 지점에 대한 분류 결과를 출력하도록 구성하며, 이 심층 신경망 모델은 기 학습되어 이미 알려진 ResNet⁹⁾이나 DenseNet¹⁰⁾, Inception¹¹⁾, VGG¹²⁾, MobileNet¹³⁾ 등 여러 합성곱 신경망 모델 중에 하나를 선정하여 전이 학습(Transfer learning)하여 사용한다.

최적 성능을 내는 기 학습된 여러 합성곱 신경망 모델들 간의 비교 및 평가 방법은, 각 모델에 대해 데이터셋으로부터 하나의 레이블 또는 여러 레이블 간의 임의의 영상들을 입력시켜 모델로부터 추론되는 특징 벡터들(Feature vectors) 간의 교차 상관도(Cross-Correlation)을 계산하고¹⁴⁾, 한 레이블에 속한 영상들에서 추출된 특징 벡터들의 교차 상관도 값이 높을수록, 다른 레이블 간에 영상들에서 추출된 특징 벡터들의 교차 상관도 값이 낮을수록 우리의 데이터셋에 적합한 모델로 판단할 수 있다. 아래 식 (1)에서 교차상관도의 계산을 보이고 있으며, fv_A 와 fv_B 는 교차 상관되는 특징 벡터이다.

$$(fv_A \star fv_B)[n] = \sum_{m=-\infty}^{\infty} \overline{fv_A[m]} \cdot fv_B[m+n] \quad (1)$$

Table 2 Cross-correlation for one-class/multi-class of the random samples in the training dataset

Model Name	Cross-correlation for one-class of the training dataset			
	Mean	Stdev	Min	Max
ResNet50	0.896	0.100	0.553	0.972
VGG16	0.897	0.134	0.448	0.985
Inception ResnetV2	0.865	0.129	0.367	0.982
MobileNet	0.859	0.060	0.682	0.980
DenseNet201	0.877	0.125	0.455	1.000

Model Name	Cross-correlation for multi-class of the training dataset			
	Mean	Stdev	Min	Max
ResNet50	0.702	0.091	0.432	0.869
VGG16	0.661	0.104	0.333	0.849
Inception ResnetV2	0.613	0.151	0.143	0.885
MobileNet	0.569	0.111	0.294	0.786
DenseNet201	0.627	0.091	0.388	0.821

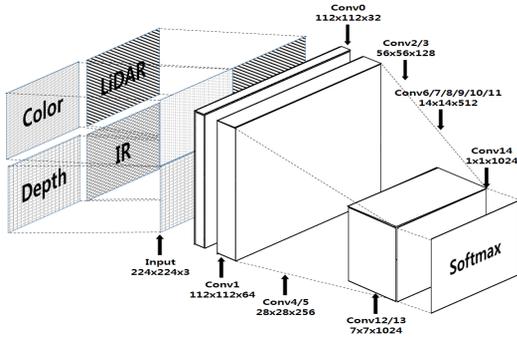


Fig. 3 Structural diagram of proposed DNN model for position estimation

Table 2에 기 학습된 여러 합성곱 신경망 모델들에서 각 모델별로 추론된 특징 벡터들 간의 교차 상관도를 구하여 나타내고 있으며, 시험된 합성곱 신경망 모델들 중에서 교차 상관도와 파라미터 크기를 고려하였을 때 MobileNet 모델이 우리의 데이터셋에 가장 적합한 것으로 판단되어 이를 전이 학습을 위한 모델로 선정하였다.

최종의 위치 추정을 위한 심층 신경망 모델은

MobileNet 모델을 기본으로 하여, 기존 Softmax 레이어를 제거하고 추가적인 Fully-connected 레이어와 새로운 레이블에 대한 Softmax 레이어를 모델 후단에 추가하여 Fig. 3과 같이 최종 심층 신경망 모델을 구성하였다.

4. 실험 결과

학습 데이터셋으로부터 심층 신경망 모델의 전이 학습을 통해 실내 복도 환경에서 총 563개의 절대 위치를 분류하도록 하였다. 학습된 심층 신경망 모델의 추론 속도는 센서 데이터의 전처리 과정을 제외하고 약 100 ms로 측정되었다.

테스트 데이터셋에 대한 위치 추정의 정확도를 아래 Table 3에 나타내고 있으며, Top-3 정확도는 약 96%, Top-5 정확도는 약 98%의 결과를 얻었다. 이는 임의의 위치에서의 센서 데이터를 입력받아 추론하였을 때 위치 추정의 3번째까지의 후보군 중에서 실제 위치가 있는 경우가 약 96%이고, 5번째까지의 후보군 중에서 실제 위치가 있는 경우가 약 98%이었음을 나타낸다. 테스트 데이터셋에 대한 로봇 주행 시의 기준 경로 대비 심층 신경망 모델의 절대 위치 추정 결과는 Fig. 4와 같으며, 0.5 m 간격으로 절대 위치가 추정되는 것을 볼 수 있다. 실제 위치와 추정 위치 간의 RMS 오차는 아래 Table 4에 나타내고 있으며, 왼쪽은 Top-1 추정 위치 오차이고 오른쪽은 Top-2까지의 추정 위치 오차이다. 심층 신경망 모델이 추정 위치를 추론할 때 입력 데이터 내의 어떤 영역에 가중치를 두고 있는지를 가시화 한 예시를 Fig. 5에 보이고 있다. 좌측은 여러 타입의 영상을 격자로

Table 3 Top-N accuracy of a DNN model for position estimation in an indoor environment

Top-N	Accuracy
1	0.786
2	0.933
3	0.959
4	0.971
5	0.977

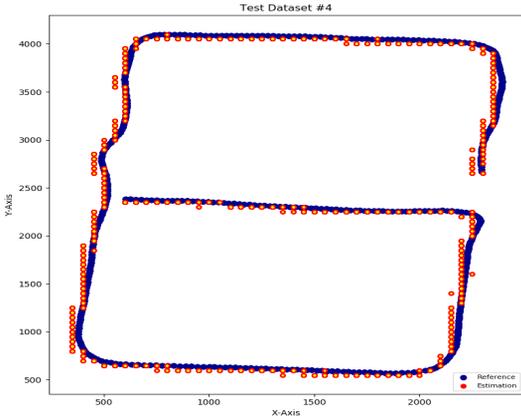


Fig. 4 Absolute position estimation results of a DNN model for test dataset

Table 4 RMSE(Root mean square error) for position estimation inference results of a DNN model

No. of Test Dataset	RMSE [cm]	
	Top-1 estimation	Top-1 & 2 estimation
1	49.78	12.22
2	56.67	14.37
3	54.99	13.86
4	65.35	15.93

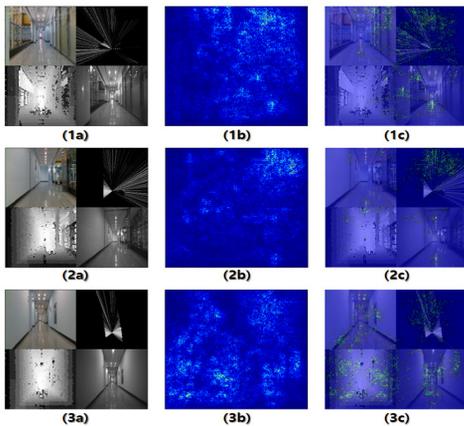


Fig. 5 Example of the feature vector visualization of a DNN model for the input image ((a)left: input original image, (b)center: visualization of the feature vector, (c)right: overlapped image)

배치한 입력 영상이고, 가운데는 위치를 추정하기 위해 심층 신경망 모델에 의해 추론이 수행될 때의 가중치를 가시화하였으며, 우측은 입력 영상에 가중치를 적용한 것으로, LiDAR 거리 정보에서는 주로 전방부 원거리 부분에 가중치를 두고 있고 컬러 및 IR 영상 등에서는 엣지가 검출되는 부분에 가중치를 두도록 학습이 된 것을 볼 수 있다.

5. 결 론

실내 복도 환경에서 여러 종류의 영상(컬러, 적외선, 깊이 정보) 및 LiDAR 거리 정보를 획득하여 절대 위치를 추정하는 심층 신경망 모델을 설계 및 학습하였으며, 위치 추정에 대한 추론 정확도는 Top-5 에러가 약 2%로 구해졌다. 테스트 데이터셋에 의한 추론 시 위치 오차는 약 14 cm로, 이는 참값 대비 동일하거나 또는 바로 인접한 격자의 위치를 추정하였음을 나타낸다.

향후에는 추정하고자 하는 위치 간격을 10 cm 수준으로 하여 해상도를 높이고, 실내 주행에 대한 복도 이동 방향의 영상뿐만 아니라 전방향에 대한 영상을 추가 획득하여 데이터셋의 일반성을 높이며, 시간축으로 공간을 이동하는 로봇의 특성을 반영하여 순환 신경망(Recurrent neural network)을 반영하도록 심층 신경망 모델을 보완하여 위치 추정의 강인성을 높이고자 한다.

후 기

이 논문은 부경대학교 자율창의기술연구비 (2018년)에 의하여 연구되었음.

Author contributions

J. W. Lee; Writing-original draft. G. D. Bae; Validation. Y. H. Choi; Project administration. Y. B. Kim; Supervision. J. H. Suh; Writing-review & editing.

References

1. L. Hoyer, C. Steup and S. Mostaghim, 2018, "A Robot Localization Framework Using CNNs for Object Detection and Pose Estimation", arXiv preprint arXiv: 1810.01665.
2. A. Forechi, T. Oliveira-Santos, C. Badue and A. F. Souza, 2018, "Visual Global Localization with a Hybrid WNN-CNN Approach", arXiv preprint arxiv: 1805.03183.
3. D. Nister, O. Naroditsky, and J. Bergen, 2004, "Visual odometry", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 652-659. (DOI:10.1109/CVPR.2004.1315094)
4. J. Y. Lee, S. H. Hong, M. S. Jung, K. R. Han and J. H. Suh, 2018, "Development of the Robot from 250 to 500A Pipes", Journal of Korean Society for Power System Engineering, Vol. 22, No. 5, pp. 13-26. (DOI:10.9726/kspse.2018.22.5.013)
5. Memcached, www.memcached.org
6. M. Cummins and P. Newman, 2007, "Probabilistic Appearance Based Navigation and Loop Closing", International Conference on Robotics and Automation, pp. 2042-2048. (DOI:10.1109/ROBOT.2007.363622)
7. M. Labbe and F. Michaud, 2013, "Appearance-based loop closure detection for online large-scale and long-term operation", IEEE Transactions on Robotics, Vol. 29, No. 3, pp. 734-745. (DOI:10.1109/TRO.2013.2242375)
8. T. Krajník, J. P. Fentanes, M. Hanheide and T. Duckett, 2016, "Persistent Localization and Life-long Mapping in Changing Environments using the Frequency Map Enhancement", International Conference on Intelligent Robots and Systems, pp. 4558-4563.
9. K. He, X. Zhang, S. Ren and J. Sun, 2016, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. (DOI:10.1109/CVPR.2016.90)
10. G. Huang, Z. Liu, L. V. Maaten and K. Q. Weinberger, 2017, "Densely Connected Convolutional Networks", IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261-2269. (DOI:10.1109/CVPR.2017.243)
11. C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, 2017, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 4278-4284.
12. K. Simonyan and A. Zisserman, 2015, "Very Deep Convolutional Networks for Large-Scale Image Recognition", International Conference on Learning Representations. (DOI:10.1109/ACPR.2015.7486599)
13. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, 2017, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv preprint arXiv: 1704.04861.
14. A. E. Usoro, 2015, "Some Basic Properties of cross-correlation functions of n-dimensional vector time series", Journal of Statistical and Econometric Methods, Vol. 4, No. 1, pp. 63-71.