

Deep Convolutional Auto-encoder를 이용한 환경 변화에 강인한 장소 인식

Condition-invariant Place Recognition Using Deep Convolutional Auto-encoder

오정현[†], 이범희[†]
Junghyun Oh[†], Beomhee Lee[†]

Abstract: Visual place recognition is widely researched area in robotics, as it is one of the elemental requirements for autonomous navigation, simultaneous localization and mapping for mobile robots. However, place recognition in changing environment is a challenging problem since a same place look different according to the time, weather, and seasons. This paper presents a feature extraction method using a deep convolutional auto-encoder to recognize places under severe appearance changes. Given database and query image sequences from different environments, the convolutional auto-encoder is trained to predict the images of the desired environment. The training process is performed by minimizing the loss function between the predicted image and the desired image. After finishing the training process, the encoding part of the structure transforms an input image to a low dimensional latent representation, and it can be used as a condition-invariant feature for recognizing places in changing environment. Experiments were conducted to prove the effective of the proposed method, and the results showed that our method outperformed than existing methods.

Keywords: Auto-encoder, Feature Extraction, Place Recognition, Deep Learning

1. 서 론

장소 인식(place recognition)은 로봇이 서로 다른 시점에 얻은 데이터로부터 같은 장소를 찾아 위치를 인식하는 기술이다. 이 기술은 로봇의 위치 정보를 제공할 수 있기 때문에 SLAM이나 네비게이션 등 로봇의 다양한 분야와 결합하여 활발히 연구가 수행되고 있다. 최근에는 카메라를 이용하여 얻은 이미지로부터 장소를 인식하는 비전 기반의 장소 인식(visual place recognition)이 주목받고 있다¹⁻³. 카메라는 작은 크기와 경제적인 가격임에도 주변 환경에 대하여 풍부한 정보를 얻을 수 있다는 장점에 로봇에 장착하여 널리 쓰이고 있는 센서이기 때문이다.

그러나 카메라를 통해 얻은 이미지는 외부 환경의 조건에 따라 현격히 다르게 보일 수 있다는 단점이 있다. [Fig. 1]은 같은 장소라고 하더라도 계절이 변화하면 얼마나 다르게 보일 수 있는지를 보여주는 예시 이미지이다.

계절 외에도 시간이나 날씨 등이 변화하면 보이는 풍경도 다르게 보이기 때문에 환경 변화가 존재하는 상황에서의 장소 인식은 가장 어려운 문제 중 하나이다.

본 논문에서는 이러한 환경 변화에서도 강인한 장소 인식을 위하여 딥러닝(deep learning) 구조 중 하나인 Convolutional auto-encoder (CAE)를 이용하여 이미지 descriptor를 추출하는 방법을 제안한다. 이 구조는 원래 입력과 출력 이미지를 같게 주고 학습을 수행하는 대표적인 비지도 학습(unsupervised learning) 구조이다. 하지만 네트워크의 입력과 관계없이 출력을 원하는 조건의 이미지로 주고 학습을 수행하면 입력과 다른 조건의 이미지를 예측할 수 있는 네트워크를 만들 수 있다. 이렇게 학습된 네트워크는 변화하는 환경에서 강인한 특징을 추출할 수 있기 때문에 장소 인식에서 높은 성능을 발휘한다.

Received : Dec. 7. 2018; Revised : Dec. 25. 2018; Accepted : Jan. 4. 2019

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2017R1A2B2002608).

† Researcher, Corresponding author: Electrical and Computer Engineering, Seoul National University, Seoul, Korea (bulley85@snu.ac.kr)

1. Professor, Electrical and Computer Engineering, Seoul National University, Seoul, Korea (bilee@snu.ac.kr)



[Fig. 1] The example images of the place under seasonal changes. A same place looks different according to the time of a day, weather, and seasons, and this makes visual place recognition a challenging problem

제한한 방법을 검증하기 위하여 실험을 수행하였다. 원하는 방향으로 학습이 수행되었는지 이미지 생성 실험을 통해 판단하였다. 그리고 이 네트워크에 학습된 이미지의 비율에 따라 장소 인식 성능이 어떻게 달라지는지를 살펴보았다. 또한 다른 알고리즘과 비교하여 장소 인식 성능이 어느 정도의 수준인지를 분석하였다. 그 결과 제안하는 방법이 기존의 방법들보다 변화하는 환경에서 장소 인식 성능이 향상되었음을 증명할 수 있었다.

2. 관련 연구

비전 기반의 장소 인식은 이미지에서 지역적인 특징(local feature)을 이용한 Bag-of-words 방식이 널리 쓰이고 있다^[4,5]. Local feature는 이미지에서 키포인트(keypoint)를 추출한 후 주변 정보로부터 특징을 추출하는데 대표적인 local feature로는 SIFT^[8], SURF^[9], BRISK^[10] 등이 있다. 학습 이미지에서 추출한 feature로부터 Visual dictionary를 만든 후, 장소 인식을 수행하고자 하는 이미지들을 히스토그램으로 만들어 비교하는 방식이다. 이 방식은 속도가 매우 빨라서 실시간으로 로봇 위에서도 동작할 수 있고 또 시점의 변화에 강인하다는 장점 때문에 널리 쓰이고 있다. 그러나 변화하는 환경에서는 local feature의 값이 크게 달라지기 때문에 장소 인식 성능이 매우 떨어진다는 한계가 있다.

이 단점을 극복하기 위해 변화하는 환경에서의 장소 인식은 이미지 전체의 정보를 이용하는 global descriptor를 추출하여 비교하는 방식이 제안되었다. 이미지를 저차원의 이미지로 변환한 후 명암 차를 맞추어 비교하는 sum-of-difference (SAD) 방식^[6]이 제안되었고, 이미지 전체에서 SURF feature를 추출하는 whole-image SURF (WI-SURF) 방식으로 장소 인식을 수행한 연구가 있었다^[11]. BRIEF-gist라는 global descriptor로부터 변화하는 환경 조건에서도 강인한 장소 인식 방법 또한 제안되었다^[12]. 이 방법들은 시점의 변화에는 불리하지만 변화

하는 환경에서는 기존의 local feature 방식보다 우수한 성능을 보였다. 그러나 계절이나 밤낮의 변화와 같이 그 변화가 클 때는 역시 성능이 높지 않다는 한계가 있었다.

최근에는 이러한 한계를 학습(learning)을 이용하여 극복하고자 하는 시도가 계속되고 있다. 지도 학습(supervised learning)과 비지도 학습 방법을 이용하여 장소의 변화를 예측하고 변하는 부분과 변하지 않는 부분을 구분하여 변하지 않는 부분을 이용하여 장소를 인식하는 방법이 제안되었다^[7]. 또한 딥러닝이 머신 러닝(machine learning) 및 컴퓨터 비전(computer vision) 분야에 널리 쓰이기 시작하면서 딥러닝 구조를 적용하고자 하는 시도도 생겨나기 시작하였다. Convolutional neural network (CNN)가 물체 분류(object classification)나 물체 인식(object recognition)에서 높은 성능을 보이고 있다는 점에 착안하여 이미 학습된 네트워크를 장소 인식에 활용하였다^[13]. 이미 학습된 네트워크가 아니라 다양한 장소의 이미지를 CNN으로 직접 학습을 수행하여 네트워크 중간층의 출력 값들을 이미지의 global descriptor로 활용하면 보다 높은 성능을 나타낼 수 있음을 증명하기도 하였다^[14]. 이 방법들은 딥러닝 구조를 장소 인식에 활용하였을 때 높은 활용도를 가질 수 있으며 특히 변화하는 환경에서도 유용하게 쓰일 수 있다.

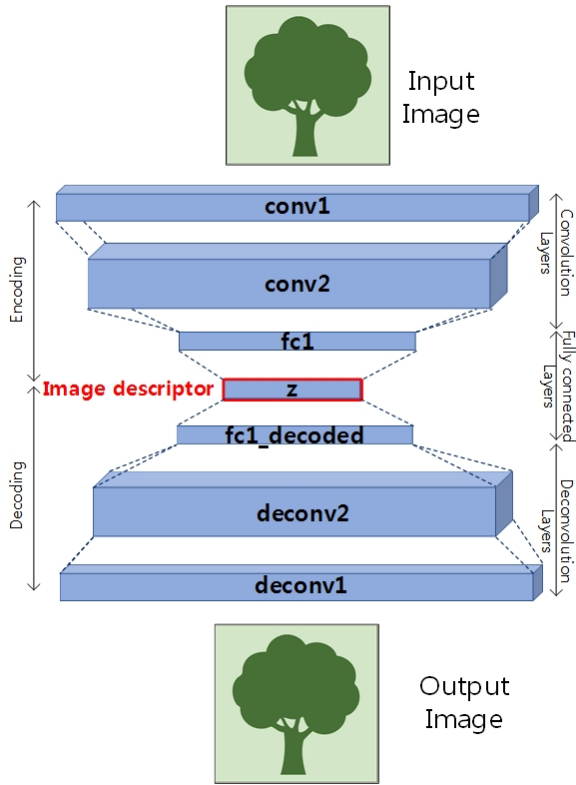
본 논문에서는 이러한 점에 착안하여 CAE^[15]를 이용하여 장소 인식을 수행하였다. 기존의 CNN를 이용한 방식은 중간 layer에서 나오는 결과들이 handcraft feature에 비하여 높은 구분력을 가지는 feature이기 때문에 높은 성능을 가질 수 있었다. 하지만 그 변화가 매우 큰 경우, 즉 낮과 밤의 이미지처럼 밝기에 큰 변화가 생기거나 계절의 변화가 생겨서 나무에 푸른 잎이 생기거나 눈이 와서 하얗게 된 경우에는 그 차이가 너무 크기 때문에 CNN과 같이 구분을 잘하는 feature도 한계가 있다. 이 문제를 극복하려면 근본적으로 새로운 환경의 이미지를 생성(generate)하여 서로 다른 두 조건을 유사하게 한 후 비교하여야 더 높은 성능을 낼 수 있을 것이다.

본 논문에서는 이 목적을 달성하기 위해 CAE를 변형하여 원하는 조건의 이미지를 생성할 수 있는 새로운 학습 방법을 제안하였다. 이 구조는 Variational auto-encoder^[16]와 같이 새로운 데이터를 생성할 수 있는 생성 모델(generative model)이다. 이 점에 착안하여 환경의 변화를 학습시킨다면 이를 통해 새로운 조건의 이미지를 생성할 수 있기 때문에 변화하는 환경에서도 강인한 장소 인식을 달성할 수 있다.

3. 환경 변화에 강인한 장소 인식 방법

3.1 Convolutional Auto-encoder 구조

CAE는 입력과 출력이 모두 자기 자신이기 때문에 따로 라벨(label)이 따로 필요 없는 대표적인 비지도 학습구조 중 하나이다.



[Fig. 2] The structure of the proposed CAE. This structure is divided into the encoding and the decoding parts. The first part transforms an input image to a low dimensional latent representation, and the second part reconstructs it to the input image

[Fig. 2]는 CAE의 구조를 자세히 나타낸 그림이다.

그림에서 살펴볼 수 있듯이 CAE는 크게 입력 이미지를 압축하는 Encoding 단계와 다시 복원하는 Decoding 단계로 나눌 수 있다. Encoding 단계에서는 입력 이미지 $\mathbf{x} \in \mathbb{R}^{m \times n}$ convolution layer와 fully connected layer를 거쳐 저차원의 벡터인 $\mathbf{z} \in \mathbb{R}^d$ 로 압축된다. 그 이후에 Decoding 단계에서는 다시 fully connected layer와 deconvolution layer를 거쳐 입력 이미지 \mathbf{x} 를 복원하는 구조로 이루어져 있다. 이때 Encoding 함수를 $f_\theta : \mathbf{x} \rightarrow \mathbf{z}$, Decoding 함수를 $g_\phi : \mathbf{z} \rightarrow \mathbf{x}$ 라고 정의하고 θ, ϕ 는 각각 함수의 파라미터라고 하면, 두 함수는 네트워크를 거치면서 복잡한 비선형 함수의 형태를 가지게 된다. 이 때 네트워크의 학습 과정은 결국 파라미터 θ, ϕ 의 값을 알아내는 과정이라고 할 수 있다. 이를 수식으로 표현하면 식 (1)과 같다.

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} L(\mathbf{x}, g_\phi f_\theta(\mathbf{x})) \quad (1)$$

여기서 L은 손실 함수(Loss function)이며 목적에 따라 다양한 형태의 함수들이 있다. 본 논문에서는 식 (2)와 같이 크기가

$m \times n$ 인 두 이미지 \mathbf{x}, \mathbf{y} 사이에 정의된 mean squared error 함수 $L(\mathbf{x}, \mathbf{y})$ 를 사용하였다.

$$L(\mathbf{x}, \mathbf{y}) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (\mathbf{x}_{ij} - \mathbf{y}_{ij})^2 \quad (2)$$

학습 데이터가 여러 개인 경우 각각의 손실 함수의 합을 최소화하는 파라미터를 찾도록 학습이 이루어 지게 된다. 학습은 stochastic gradient descent 혹은 Adam과 같이 최적화 툴이 널리 쓰이고 있다. 네트워크의 구조에 따라 적합한 툴을 이용하여 학습을 수행하게 된다. 학습이 끝난 후에 이 네트워크는 자기 자신을 복원할 수 있기 때문에 denoising과 같은 용도로 쓰일 수 있다^[7].

3.2 CAE를 이용한 특징 추출 및 장소 인식 방법

이제 변화하는 환경에서 장소 인식을 수행하기 위해 두 환경에서 얻은 이미지쌍 집합 $\{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \dots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}$ 이 학습 데이터로 주어져 있다고 가정한다. 이와 같이 학습 데이터가 주어져 있다는 가정은 장소 인식 분야에서 보다 높은 성능을 위하여 널리 쓰이고 있는 가정이다^{[4], [6], [7]}.

CAE는 원래 입력과 출력을 같게 하여 학습을 수행한다. 하지만 본 논문에서는 입력에 상관없이 출력을 원하는 환경의 이미지 $\mathbf{y} \in \mathbb{R}^{m \times n}$ 로 주고 변화한 환경의 이미지를 생성하도록 학습을 수행한다. 즉, Encoding 함수를 $f_\theta : \mathbf{x} \rightarrow \mathbf{z}$, Decoding 함수를 $g_\phi : \mathbf{z} \rightarrow \mathbf{y}$ 라고 정의하고, 학습은 다음과 같이 식 (2)를 만족하도록 파라미터를 학습한다.

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} \sum_{k=1}^n (L(\mathbf{y}^{(k)}, g_\phi f_\theta(\mathbf{y}^{(k)})) + L(\mathbf{y}^{(k)}, g_\phi f_\theta(\mathbf{x}^{(k)}))) \quad (3)$$

즉 학습이 완료되면 이 네트워크는 입력 이미지의 환경에 상관없이 원하는 조건의 이미지를 만들어 내는 결과를 갖는다. 따라서 서로 다른 두 환경의 이미지도 유사한 조건의 이미지로 만들어서 비교할 수 있기 때문에 환경의 영향을 최소화할 수 있다. 특히 딥러닝 구조는 이와 같이 서로 다른 두 환경에서 얻은 이미지 사이의 비선형 변환을 추정하는데 유리하기 때문에 기존의 다른 학습 방법보다 높은 성능을 기대할 수 있다.

이미지를 유사한 환경의 이미지로 생성하여 비교하는 방법도 있지만 이미지 정보를 압축한 descriptor를 사용하면 비교 연산 시간도 줄이면서도 데이터의 핵심 정보만 사용하기 때문에 노이즈에 강인하다는 장점이 있다. CAE는 [Fig. 2]에서 볼

수 있듯이 Encoding 함수를 거친 입력이 저차원의 벡터 \mathbf{z} 로 압축되기 때문에 이 벡터를 이미지 descriptor로 이용할 수 있다. 서로 다른 환경의 이미지를 압축한 descriptor $\mathbf{u} \in \mathbf{R}^d$ 와 $\mathbf{v} \in \mathbf{R}^d$ 를 얻었다고 하자. 이 때 두 벡터 사이의 거리는 값이 $[-1,1]$ 사이의 값을 가지는 코사인 거리(cosine distance)를 이용하였다.

$$D = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^d \mathbf{u}_i \mathbf{v}_i}{\sqrt{\sum_{i=1}^d \mathbf{u}_i^2} \sqrt{\sum_{i=1}^d \mathbf{v}_i^2}} \quad (4)$$

이와 같이 두 벡터 사이의 거리를 계산한 후 두 이미지 시퀀스가 주어졌을 때 가장 가까운 벡터를 찾아 두 이미지가 같은 장소라고 인식할 수 있다. Descriptor 간의 거리 유사도로부터 장소를 인식하는 방법은 가장 간단한 방법이고 이 방법은 필터 기반의 로봇 위치 추정^[4]이나 시퀀스 기반의 장소 인식 방법^[6]과 같은 추정 등과 같은 다른 추정 정보와 결합할 수 있다. 이러한 방법과 결합된다면 보다 환경 변화에 강인한 장소 인식을 수행할 수 있다.

4. 실험 결과

4.1 실험 환경

제한한 방법을 검증하기 위해 실제 환경 변화가 뚜렷한 데이터셋 이미지를 이용하여 실험을 수행하였다. 사용한 데이터셋은 Nordland 데이터셋^[7]으로, 노르웨이를 가로지르는 기차가 약 728km 거리를 주행하면서 촬영한 이미지이다. 이 데이터셋은 1년에 걸쳐 봄, 여름, 가을, 겨울에 촬영한 이미지를 모두 가지고 있으며 기차 트랙 위에서 주행하면서 얻은 이미지이기 때문에 카메라 시점의 변화 없이 환경의 변화만 있다. 따라서 다른 조건은 통제된 채 변화하는 환경에서의 장소 인식 성능을 테스트하기 적합한 데이터이다. 본 논문에서는 4계절 전부를 이용하지 않고 환경의 변화가 뚜렷한 봄과 겨울에서 얻은 이미지를 가지고 실험을 수행하였다. 각각 3200장의 이미지를 가지고 있으며 해상도는 32×64 의 크기에 컬러 채널을 가지고 있어서 총 $32 \times 64 \times 3$ 의 차원을 가지고 있다.

실험에 쓰인 네트워크 구조는 [Table 1]과 같다. 입력 이미지의 차원이 크지 않기 때문에 Encoder 부분에서는 convolutional layer와 max pooling layer 2층과 fully-connected 1층, 그리고 \mathbf{z} 로 압축된 후 Decoder 부분에서 fully-connected 1층, 그리고 deconvolutional layer 2층 후 출력 이미지를 복원하는 구조로 되어 있다. 네트워크의 층 수와 노드 개수는 데이터셋에 따라 더 높은 성능을 위해 달라질 수 있다.

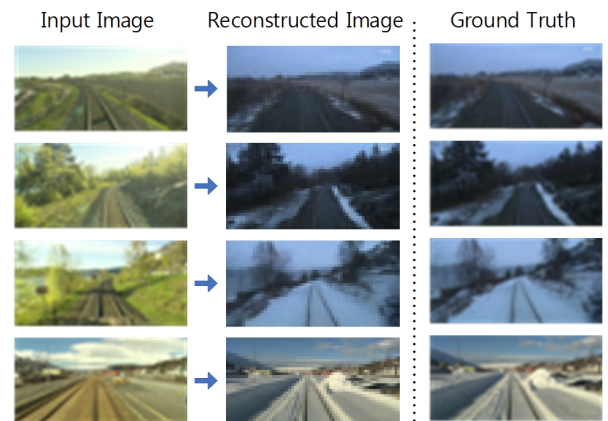
[Table 1] Output size of each layer

	Layers	Name	Dimensions
Encoder	Convolutional layers	input	$32 \times 64 \times 3$
		conv1	$32 \times 64 \times 48$
		pool1	$16 \times 32 \times 48$
		conv1	$16 \times 32 \times 96$
		pool2	$8 \times 16 \times 96$
Decoder	Fully conneted layers	fc1	2048×1
		z	1024×1
	De-convolutional layers	fc1_decoded	2048×1
		deconv2	$16 \times 32 \times 96$
		deconv1	$32 \times 64 \times 48$
		output	$32 \times 64 \times 3$

4.2 이미지 생성 실험

먼저 입력 이미지와 출력 이미지를 다르게 학습하였을 때 이 네트워크가 실제로 이미지 복원을 수행할 수 있는지 실험을 수행하였다. 입력으로는 봄에 찍은 이미지들을 주고 출력으로는 그에 대응되는 겨울에 찍은 이미지들을 주어서 학습을 수행하였다. 그 후 학습된 네트워크에 학습에 쓰이지 않은 입력 이미지를 주었을 때 복원된 겨울 이미지가 실제 겨울 이미지와 유사하게 나타나는지 실험을 수행하였다.

그 결과 [Fig. 3]과 같이 복원 이미지는 실제와 매우 유사한 이미지가 생성되는 것을 관찰할 수 있었다. 복원된 이미지가 실제 이미지보다 뿌옇게 나오는 blur되는 현상이 관찰되긴 하지만, 겨울의 특징인 짙은 푸른 색의 하늘과 바닥에는 하얀 눈이 생성되



[Fig. 3] The reconstruction test of the proposed method. Using the trained CAE, the input image can be transformed to reconstructed image. Although the input images are not in the training set, we can find that the reconstructed images are similar to real images

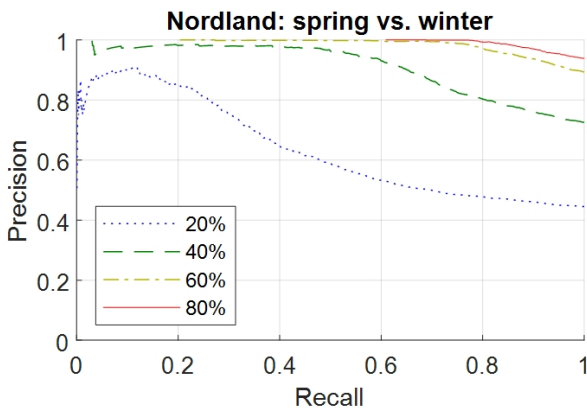
고, 나무의 색도 보다 어두운 색으로 변한 것을 볼 수 있었다. 봄 이미지에는 없던 환경이 CAE를 통해 새로운 환경을 생성한 것이다. 이는 생성 모델(generative model)인 Auto-encoder의 특징이기도 하다. 이 특징을 이용하면 봄 이미지를 겨울 이미지로 바꾸어서 비교할 수 있는 것처럼 서로 다른 환경도 유사한 환경으로 변환시킨 후 비교할 수 있다. 변환시킨 이미지 사이에 이미지 feature를 비교하면 변화하는 환경에서도 보다 정확한 장소 인식을 수행할 수 있다.

4.3 장소 인식 성능 실험

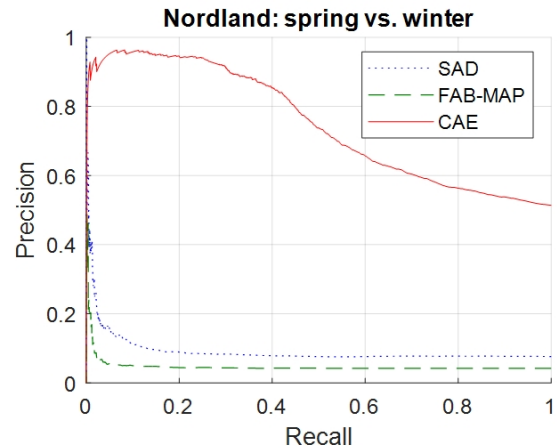
학습된 네트워크를 이용하여 실제 장소 인식 성능이 높아지는지 살펴보기 위하여 실험을 수행하였다. 장소 인식 성능은 일반적으로 Precision-recall curve를 통해 평가할 수 있다¹¹⁾. 장소 인식 성능은 Precision과 Recall이 모두 1에 가까울수록 높은 성능을 가지지만, False positive에 보다 더 민감하다. 따라서 Precision 성능을 우선시하여 최대한 높인 후에 Recall 성능이 높을수록 보다 우수한 장소 인식 성능을 가지고 있다.

이제 Precision-recall curve로 장소 인식을 검증하기 전에 먼저 Training data의 비율에 따라 이 성능이 어떻게 바뀌는지를 살펴보았다. 그 결과는 [Fig. 4]와 같다. Training data의 비율이 높을수록 보다 많은 데이터를 가지고 학습을 수행하기 때문에 그 환경에 대한 정보를 더 많이 가질 수 있다. 따라서 더 높은 장소 인식 성능을 가진다.

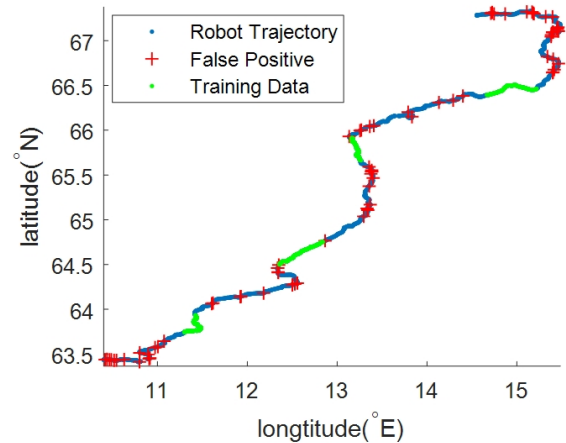
이제 Precision-recall curve로 장소 인식을 검증하기 전에 먼저 Training data의 비율에 따라 이 성능이 어떻게 바뀌는지를 살펴보았다. 그 결과는 [Fig. 4]와 같다. Training data의 비율이 높을수록 보다 많은 데이터를 가지고 학습을 수행하기 때문에 그 환경에 대한 정보를 더 많이 가질 수 있다. 따라서 더 높은 장소 인식 성능을 가진다.



[Fig. 4] Precision-recall curve results of the Nordland spring-winter dataset. The higher the ratio of the training data, the better the performance of the place recognition



[Fig. 5] The comparison of the precision-recall curve results of the Nordland spring-winter dataset. The results showed that our method outperforms than other methods



[Fig. 6] The place recognition results of the robot trajectory. The blue line is the robot trajectory, and the green dots are the locations of the training data. Locations marked with a red cross are false positives

제안한 알고리즘이 다른 알고리즘보다 우수한 장소 인식을 가지는지 살펴보기 위하여 Precision-recall curve를 그려서 비교하였다. 결과는 [Fig. 5]와 같다.

비교한 알고리즘은 Local feature 기반의 FAB-MAP¹⁴⁾와 이미지를 저차원으로 변환 후 contrast enhancement를 통해 비교하는 SeqSLAM⁵⁾에서 쓰인 Sum-of-absolute-difference (SAD) 방식과 비교하였다. 제안하는 방법은 Training data의 비율을 20%로 고정하고 80%를 Test data로 이용하였다. Test 이미지 간은 평균적으로 약 200 m의 거리가 떨어져 있다. 비교하는 알고리즘은 정적인 환경에서는 우수한 장소 인식 성능을 가지지만 이번 실험과 같이 변화하는 환경에서는 Precision과 Recall이 모두 낮게 나타나는 것을 볼 수 있다. 반면 제안한 방법은 이전의 방법들보다 높은 Precision과 Recall을 가지기 때문에 변화하는 환경에서의 장소 인식 성능이 우수하다는 것을 확인할 수 있었다.

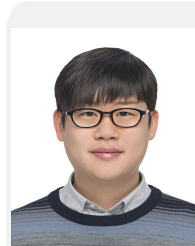
마지막으로 Training data의 비율이 20%일 때 로봇 경로상에서 장소 인식 성공 및 실패 여부를 나타낸 결과는 [Fig. 6]과 같다. 그림과 같이 변화하는 환경에서도 대부분의 위치에서 장소 인식이 성공함을 알 수 있다.

5. 결 론

본 논문은 환경 변화에서도 강인한 장소 인식을 위하여 딥러닝(deep learning) 구조인 CAE를 이용하였다. 기존의 입력과 출력을 같게 하여 학습을 수행하던 방식을 변형하여, 입력에 상관없이 원하는 조건의 이미지를 출력하도록 학습을 수행하였고, 이를 통해 유사한 조건에서 두 이미지를 비교할 수 있게 하였다. 또한 Auto-encoder의 특징을 활용하여 저차원 벡터를 descriptor로 활용하여 장소 인식을 수행하는 방법을 제안하였다. 그 결과 기존의 방법보다 높은 장소 인식 성능을 검증할 수 있었다.

References

- [1] S. Lowry, N. Sünderhauf, P. Newman, J. Leonard, D. Cox, P. Corke, and M. Milford, "Visual place recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1-19, Feb., 2016.
- [2] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," *European Conference on Computer Vision (ECCV)*, pp. 834-849, 2014.
- [3] R. Mur-Artal and J. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255-1262, Oct., 2017.
- [4] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100-1123, 2011.
- [5] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Fast and incremental method for loop-closure detection using bags of visual words," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1027-1037, Oct., 2008.
- [6] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," *IEEE International Conference on Robotics and Automation (ICRA)*, Saint Paul, MN, USA, pp. 1643-1649, 2012.
- [7] S. Lowry and M. J. Milford, "Supervised and unsupervised linear learning techniques for visual place recognition in changing environments," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 600-613, Jun., 2016.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov., 2004.
- [9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, Jun., 2008.
- [10] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *2011 IEEE International Conference on Computer Vision*, Barcelona, Spain, pp. 2548-2555, 2011.
- [11] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," *2012 IEEE International Conference on Robotics and Automation*, Saint Paul, MN, USA, pp. 1635-1642, 2012.
- [12] N. Sünderhauf and P. Protzel, "BRIEF-Gist - closing the loop by simple means," *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, CA, USA, pp. 1234-1241, 2011.
- [13] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of ConvNet features for place recognition," *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, pp. 4297-4304, 2015.
- [14] Z. Chen, A. Jacobson, N. Sünderhauf, B. Upcroft, L. Liu, C. Shen, I. Reid, and M. Milford, "Deep learning features at scale for visual place recognition," *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, Singapore, pp. 3223-3230, 2017.
- [15] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," *International Conference on Artificial Neural Networks*, pp. 52-59, 2011.
- [16] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv: 1312.6114 [stat. ML], 2014.
- [17] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, pp. 3371-3408, 2010.



오 정 현

2012 서울대학교 전기공학부(공학사)
2014 서울대학교 전기·컴퓨터공학부(공학석사)
2018 서울대학교 전기·정보공학부(공학박사)

관심분야: Place recognition, Persistent navigation, SLAM, Artificial intelligence, Deep learning



이 범 희

1978 서울대학교 전자공학과(공학사)
1980 서울대학교 전자공학과(공학석사)
1985 Computer Information & Control Engineering, University of Michigan, Ann Arbor. (공학박사)
1987~현재 서울대학교 전기·정보공학부 교수

관심분야: Place recognition, Persistent navigation, SLAM, Artificial intelligence, Deep learning