

열화상 영상의 Image Translation을 통한 Pseudo-*RGB* 기반 장소 인식 시스템

Pseudo-*RGB*-based Place Recognition through Thermal-to-*RGB* Image Translation

이승현¹·김태주²·최유경[†]

Seunghyeon Lee¹, Taejoo Kim², Yukyung Choi[†]

Abstract: Many studies have been conducted to ensure that Visual Place Recognition is reliable in various environments, including edge cases. However, existing approaches use visible imaging sensors, *RGB* cameras, which are greatly influenced by illumination changes, as is widely known. Thus, in this paper, we use an invisible imaging sensor, a long wave length infrared camera (LWIR) instead of *RGB*, that is shown to be more reliable in low-light and highly noisy conditions. In addition, although the camera sensor used to solve this problem is an LWIR camera, but since the thermal image is converted into *RGB* image the proposed method is highly compatible with existing algorithms and databases. We demonstrate that the proposed method outperforms the baseline method by about 0.19 for recall performance.

Keywords: Visual Place Recognition, Thermal, Image Translation, Image Retrieval

1. 서론

지도 서비스, 자율주행 등의 위치 기반 서비스를 이용하기 위해서는 차량 혹은 기기의 현재 위치 정보가 필요하다. 장소 인식을 위해 GPS (Global Positioning System)를 활용하는 방식이 일반적이지만, 빌딩 숲이나 지하, 터널 등과 같이 인공위성과의 통신이 차단되는 경우에는 GPS를 대체할 장소 인식 시스템이 요구된다. 따라서, 최근에는 어떠한 상황에도 장소 인식이 가능하도록 하기 위해 와이파이, LiDAR, 영상 정보 등을 활용하는 연구¹⁻³⁾가 이루어지고 있다. 특히, 영상 기반 장소

인식(Visual Place Recognition) 연구는 영상 센서의 범용성과 접근성이라는 장점을 앞세워 많은 연구가 진행되고 있다.

영상 기반 장소 인식³⁻⁷⁾은 위치 정보를 알고 싶은 쿼리영상 (Query)과 사전에 구축된 데이터베이스(DB) 속의 영상들을 비교하여 DB 중 가장 유사한 영상의 위치정보를 현재 위치로 활용한다. 실제 상황에서 영상 기반의 장소 인식 시스템을 항상 활용하기 위해서는 주야간, 날씨, 계절의 변화에도 장소 인식이 강인하게 작동하여야 한다. 그러나 *RGB* 영상을 DB와 쿼리로 이용하는 기존의 장소 인식 방법론은 *RGB* 영상의 특성 상 빛이 적은 밤이나 안개와 같이 가시광선이 제대로 투과되지 못하는 환경에서 잘 작동하지 않는다. 이러한 *RGB* 영상의 문제를 해결하기 위해 저조도 환경에서 촬영된 *RGB* 영상을 딥러닝 모델을 이용하여 정상적인 영상으로 렌더링하는 연구⁸⁾, *RGB* 영상의 구조적 정보에 집중하여 장소 인식을 수행하는 연구⁹⁾, 야간 *RGB* 영상을 낮의 *RGB* 영상으로 변환하는 연구¹⁰⁾ 등 입력 데이터의 표현력을 개선하기 위한 연구가 활발히 진행되고 있다. 그러나 위의 방법론들은 극심한 저조도 상황과 같이 정보량이 적은 영상에 대해서는 정확한 예측이 어렵다.

또 다른 방법으로 원적외선 파장대를 포착하여 야간에도 강인한 열화상 카메라를 활용할 수 있다. 이러한 장점 덕분에

Received : Nov. 1. 2022; Revised : Nov. 22. 2022; Accepted : Nov. 23. 2022

※ This project was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1F1A1076987 and NRF-2020M3F6A1109603)

1. Undergraduate Student, Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Korea (shlee@rcv.sejong.ac.kr)

2. Integrated M.S./ PhD Program Student, Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Korea (tjikim@rcv.sejong.ac.kr)

† Associate Professor, Corresponding author: Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Korea (ykchoi@rcv.sejong.ac.kr)

컴퓨터비전 및 로봇틱스 분야에서는 RGB 영상과 열화상 영상을 결합하여 이미지의 표현력을 높이는 연구^[11], 열화상 영상을 RGB 영상과 열화상 영상의 공통의 도메인으로 임베딩하여 유사도를 비교하는 연구^[12] 등이 이루어지고 있다. 본 연구는 RGB 영상이 가지는 한계를 보완하기 위해 조도 및 환경 변화에 강인한 열화상 영상을 쿼리 영상으로 이용하는 방법을 고안한다.

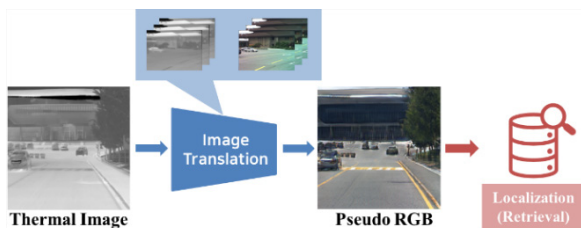
열화상 영상은 RGB 영상과 다른 파장의 정보를 포착하기 때문에 영상이 가진 정보도 다르다. 두 영상의 유사도를 비교해야 하는 영상 기반 장소 인식의 신뢰도를 확보하기 위해서는 DB 영상과 쿼리 영상이 동일한 도메인으로 구성되어 있어야 한다. 하지만 대부분의 DB는 RGB 영상으로 구축되어 있어 RGB DB에 열화상 영상을 쿼리로 활용하기에는 어려움이 있다. 이 문제를 해결하기 위해 열화상 영상에 대한 별도의 DB를 구축하는 방법이 있으나, 이는 시간과 비용 측면에서 비효율적이다. 또한, 열화상 영상과 RGB 영상을 함께 쿼리 이미지로 이용할 경우 두 도메인별로 유사도를 구해야 하므로 용량이 두 배로 커지고, 연산량의 증가로 속도가 느려진다는 문제가 발생한다. 따라서 본 연구는 열화상 영상을 위한 새로운 DB를 구축하지 않는 것을 목표로 장소 인식 시스템을 디자인한다.

우리는 열화상 영상을 RGB 도메인으로 변환하여 장소 인식을 수행하는 시스템을 제안한다. 또한, 실험을 통해 RGB 영상을 쿼리로 이용하는 방법론이 저조도 및 주변 환경 변화에 의해 영향을 받는다는 것을 확인하고, 제안한 시스템을 활용할 경우 장소 인식 성능이 개선됨을 보인다.

2. 열화상 영상을 활용한 Place Recognition

2.1 시스템 개요

[Fig. 1]은 열화상 영상을 쿼리 영상으로 이용하는 장소 인식 과정을 도식화한 것이다. 먼저, RGB 영상으로 구성된 DB와 유사도를 비교하기 위해 시스템에 입력되는 열화상 영상을 Image Translation 기법을 이용하여 RGB 도메인으로 변환하고, 이렇게 얻어진 영상을 Pseudo RGB라 한다. 이후, Pseudo RGB 영상과 DB 내의 RGB 영상을 비교하여 가장 유사한 영상



[Fig. 1] Place recognition system pipeline using Thermal imaging

을 찾는다. 이때, 유사한 영상을 정렬하기 위해 각 영상을 특징점 주변 영역에서 추출한 기술자(descriptor)로 표현한 뒤, 쿼리 영상과 DB 영상 간의 유클리디안 거리를 기준으로 거리가 가까울수록 유사한 영상으로 판단한다. 장소 인식의 성능 지표는 recall로, 유사도를 기준으로 내림차순으로 정렬한 뒤 k개의 상위 영상(Top-k) 중 실제로 촬영된 위치와 일정한 거리 임계값(threshold) 내의 영상들의 비율을 나타낸 값이다. 이때 각 영상에는 위치정보가 붙어있어 영상의 실제 위치 정보를 알 수 있다. 다음 절은 각 단계에 대한 자세한 설명을 포함한다.

2.2 열화상 영상을 RGB 영상으로 변환

가시광선 영역의 정보를 이용하는 RGB 영상과 적외선 영역의 정보를 이용하는 열화상 영상은 서로 다른 정보를 가지고 있으므로, 두 도메인 사이에는 간격이 존재한다. 검색 기반의 장소 인식 방식은 단일 영상의 특징점을 찾고 특징점의 주변 픽셀 정보를 서술하는 기술자를 이용하여 유사도를 측정하는 방식으로, 서로 다른 정보를 가진 영상으로부터 동일한 방식으로 추출한 기술자 비교를 통해 얻은 유사도 측정 결과는 신뢰하기 어렵다. 이를 실험적으로 확인하기 위해 RGB로 구성된 DB에 대한 열화상 영상의 검색 성능을 측정하였다. 정량적 결과는 [Table 1]에 정리하였다. 동일한 도메인의 영상에 대한 Top-1의 recall 성능은 0.9662로 매우 높았지만, 다른 도메인의 영상에 대해서는 0.2545로 성능 하락 폭이 큰 것을 확인할 수 있다.

이처럼 두 도메인 사이의 간격에 의해 발생하는 성능 하락 문제를 해결하기 위해 본 연구는 열화상 영상인 쿼리 영상을 RGB 영상으로 변환한다. 변환된 Pseudo RGB 영상을 만들기 위해 소스 이미지를 타겟 도메인의 영상으로 변환하는 연구인 Image Translation을 적용하였고, 이때 생성 모델을 활용한다. 기존의 딥러닝 기반 Image Translation 방법론은 영상의 해상도를 줄였다 늘리는 Encoder-Decoder 구조로 인해 정보 손실이 일어나 변환된 영상의 구조적 정보 및 국소적(local) 정보를 잃게 된다는 문제점이 있다. 이를 해결하기 위해 본 논문에서는 계층적인 네트워크 구조를 적용하며, 고해상도의 특징 맵을 병렬적으로 연산하여 디테일한 정보를 유지하는 방법론인 HRFormer^[13]를 적용한다.

[Table 1] Retrieval performance in two domains (RGB/ Thermal) for a DB configured in RGB (d = 25 m)

DB: RGB		recall@1	recall@5	recall@10	recall@20
Query	RGB	0.9332	0.9489	0.9587	0.9617
	Thermal	0.1967	0.3348	0.4204	0.5323
	Ours	0.5548	0.7342	0.7988	0.8514

Image Translation 모델을 학습하기 위해, 동일한 장면에 대한 RGB 영상과 열화상 영상이 존재하는 KAIST dataset^[14]을 이용하였다. 낮에 촬영된 3040개의 쌍을 학습 데이터로 이용하였으며, 장소 인식에 쿼리로 사용되는 1332장의 영상에 대해 추론을 하였다. 모델은 열화상 영상을 입력하고 그에 대응하는 RGB 영상과 유사하게 이미지를 만들도록 생성 모델을 학습하였다. [Fig. 2]는 학습한 Image Translation 모델을 이용하여 변환한 Pseudo RGB이다. 위의 2개 행은 낮에 촬영된 영상이고 아래의 2개 행은 밤에 촬영된 영상이다. 왼쪽부터 순서대로 열화상 영상, Pseudo RGB, RGB 영상이며, 야간에는 RGB 영상의 표현력이 떨어진다는 것을 확인할 수 있다. 모델의 추론 결과들을 통해 계층적 구조로 이루어진 HRFormer를 활용함으로써 방지턱, 횡단보도와 같이 영상의 디테일한 부분도 잘 변환하는 것을 확인할 수 있었다. 한 가지 흥미로운 점은 야간에 촬영된 영상을 학습 과정에서 보지 않았음에도 불구하고 야간에 촬영된 열화상 영상을 낮 영상처럼 생성한다는 점이다. 이는 장소 인식 관점에서 DB와의 유사도를 계산할 때 야간의 열악한 조도 환경을 극복할 수 있음을 의미한다.



[Fig. 2] Qualitative results of Thermal-to-RGB translation. The first and second rows were taken during the day, and the third and last rows were taken at night. On the left is Thermal, in the middle is Pseudo RGB, and on the right is RGB

2.3 Pseudo RGB를 이용한 장소 인식

Image Translation을 통해 변환한 Pseudo RGB 영상을 이용하여 장소 인식을 수행하는 단계이다. 우리는 장소 인식에서 많이 사용되는 NetVLAD^[4]를 이용한다. NetVLAD는 Pseudo RGB 영상과 DB 내의 단일 영상을 각각 CNN 모델을 통과시켜 이미지의 매칭 포인트가 아닌, 이미지 전반에 걸친 기술자 (Dense Descriptor)를 추출한다. 이후, 기술자들에 클러스터링을 적용하여 구한 K개의 클러스터의 중심점을 Visual Words라 하고, 각 기술자와 Visual Words 사이의 차이를 구하여 벡터로 영상을 표현한다. 벡터로 표현된 영상은 유클리드 거리를 이용하여 가장 유사한 영상부터 정렬하고, Top-k개의 영상을 이용하여 현재 위치를 알아내게 된다.

장소 인식에 사용한 데이터셋은 Image Translation과 마찬가지로 KAIST dataset^[14]을 이용하였다. 이때 DB영상과 Query로 사용된 데이터는 동일 장소에서 따로 촬영된 시퀀스를 이용하였고, recall 성능을 구할 때 NetVLAD 논문에서 활용한 25 m를 임계 값으로 하여 Top-k개의 영상의 참과 거짓을 판단하였다.

3. 실험

3.1 야간 상황에 대한 장소 인식 실험

열화상 카메라를 이용한 장소 인식이 밤에 잘 작동하는지 확인하기 위해, 밤 시간대에 촬영된 RGB와 열화상 영상에 대한 장소 인식 성능을 [Table 2]에 정리하였다. 표를 통해 확인할 수 있듯이 밤에 촬영된 RGB 영상을 이용하여 장소 인식을 수행할 경우 Top-1에 대한 recall 성능은 0.1009로 매우 낮은 성능을 보이는 것을 확인할 수 있었다. 반면 RGB 도메인으로 변환한 이미지로 장소 인식을 수행할 경우 0.3892로 RGB나 Thermal 영상을 그대로 사용할 때보다 성능이 확연히 개선되었다. 이를 통해 본 논문에서 제안한 방식을 활용할 경우 야간에 더욱 강인하게 장소 인식이 가능함을 확인하였다.

3.2 이상 상황에 대한 장소 인식 실험

가시광선 파장을 이용하는 RGB 카메라는 안개가 낄 경우 가시광선이 통과하지 못하여 주변 환경에 대한 정보를 얻을

[Table 2] Nighttime Image Retrieval Performance (d = 25 m)

DB: RGB		recall@1	recall@5	recall@10	recall@20
Query	RGB	0.1009	0.1856	0.2505	0.3495
	Thermal	0.2018	0.3730	0.4829	0.5838
	Ours	0.3892	0.5784	0.6793	0.7694

수 없고, 카메라 렌즈에 물이 맺히거나 서리가 끼는 등 일반적인 지 못한 상황이 발생할 수 있다. 이러한 상황에도 장소 인식이 가능하도록 하기 위해 RGB 영상에 비해 날씨 변화에 강한 열화상 영상을 쿼리 영상으로 활용하여 장소 인식을 수행하고자 한다. RGB 영상에 주변 환경에 의한 변화가 촬영된 경우에 장소 인식 시스템이 영향을 받는지를 검증하기 위해 우선 영상에 노이즈가 발생할 수 있는 상황을 아래와 같이 5가지로 정의하였다.

- **Motion Blur:** 조도가 낮아질 경우 RGB 카메라는 받는 빛의 양을 늘리기 위해 셔터스피드를 길게 늘이고, 이에 따라 blur가 생긴 이미지가 촬영될 수 있다.
- **Spatter:** 비가 오거나 기온 차로 인해 차량 앞 유리 또는 카메라 렌즈에 습기가 맺혀 물방울이 같이 촬영될 수 있다.
- **Snow:** 눈이 올 경우 촬영되는 RGB 영상에 내리는 눈이 함께 촬영될 수 있다.
- **Fog:** 카메라 주변에 불이 나거나, 온도 차이로 인해 연기가 발생할 수 있다.
- **Frost:** 겨울철 주변의 수증기가 얼어 카메라나 차량 전면에 서리가 생길 수 있다.

기존의 데이터 셋에는 위의 5가지 상황을 모두 포함하며 영상의 촬영 위치가 태그된 데이터셋이 존재하지 않는다. 이에 컴퓨터비전 분야에서 이상 상황을 모사하는 데 많이 사용되는 ImageNet-C^[15]의 합성 데이터 생성 방식을 이용하여 위의 5가지 상황을 표현하였다. 이때, RGB 영상이 잘 작동하지 않는 상황에서의 열화상 영상의 대체 가능성을 확인하고자 하는 목적에 맞춰 Fog에 대한 강도를 높여 연기가 자욱한 상황을 모사하였다. 이상 상황에 대한 장소 인식을 수행하여 [Table 3]에 결과를 정리하였다.

[Table 3]을 통해 확인할 수 있듯이 다양한 상황을 모사한 RGB 영상에 대해 장소 인식을 수행할 경우 정확도가 크게 낮아진다. 이에 비해 Thermal 영상을 RGB로 변환한 뒤에 장소 인식을 수행할 경우 RGB 영상에 비해 개선된 recall을 얻을 수

[Table 3] Image Retrieval Performance for RGB images that mimic noisy environments ($d = 25$ m)

DB: RGB		recall@1	recall@5	recall@10	recall@20
Query	Motion blur	0.3791	0.5222	0.6099	0.7003
	Spatter	0.4309	0.6059	0.6809	0.7523
	Snow	0.4165	0.5871	0.6698	0.7485
	Fog	0.4287	0.5976	0.6652	0.7432
	Frost	0.4602	0.6134	0.6929	0.7628
	Ours	0.5548	0.7342	0.7988	0.8514

있다. 이를 통해 열화상 영상을 이용한 장소 인식을 진행한다면 일반적이지 않은 상황에서도 강한 장소 인식 시스템을 가질 수 있음을 확인할 수 있다.

4. 결 론

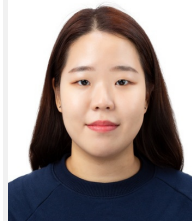
본 연구는 조도 및 주변 상황에 크게 영향을 받는 RGB 카메라를 보완하기 위해 조도 변화에 강한 열화상 카메라를 이용하여 장소 인식 시스템을 구축하였다. 이때, 열화상 영상에 대한 장소 인식을 위해 DB를 새롭게 구축하지 않고, 기존에 구축된 DB를 활용하여 장소 인식이 가능함을 실험적으로 확인하였다. 또한, 시간과 날씨 변화에 대응하기 위해 열화상 영상을 이용 가능함을 확인하였고, 제안한 장소 인식 방법론을 활용하면 연기가 자욱하거나 습기가 많은 사고 및 재난 현장에서 활용할 수 있을 것으로 기대된다.

References

- [1] F. Zafari, A. Gkeliias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, 2019, DOI: 10.1109/COMST.2019.2911558.
- [2] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. McCullough, and A. Mouzakitis, "A Survey of the State-of-the-Art Localization Techniques and Their Potentials for Autonomous Vehicle Applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829-846, 2018, DOI: 10.1109/JIOT.2018.2812300.
- [3] X. Zhang, L. Wang, and Y. Su, "Visual place recognition: A survey from deep learning perspective," *Pattern Recognition*, vol. 113, May, 2021, DOI: 10.1016/j.patcog.2020.107760.
- [4] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016, DOI: 10.1109/CVPR.2016.572.
- [5] G. Berton, C. Masone, and B. Caputo, "Rethinking Visual Geo-localization for Large-Scale Applications," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, Jun., 2022, DOI: 10.1109/CVPR52688.2022.00483.
- [6] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla, "24/7 place recognition by view synthesis," *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, DOI: 10.1109/CVPR.2015.7298790.
- [7] J. Hyeon, J. Kim, and N. Doh, "Pose correction for highly accurate visual localization in large-scale indoor spaces," *IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, Oct., 2021, DOI: 10.1109/ICCV48922.2021.01567.
- [8] E. Ershov, A. Savchik, D. Shepelev, N. Banić, M. S. Brown, R. Timofte, K. Koščević, M. Freeman, V. Tesalin, D. Bocharov, I.

Semenkov, M. Subašić, S. Lončarić, A. Terekhin, S. Liu, C. Feng, H. Wang, R. Zhu, Y. Li, L. Lei, Z. Li, S. Yi, L. Han, R. Wu, X. Jin, C. Guo, F. Kinli, S. Menteş, B. Özcan, F. Kırac, S. Zini, C. Rota, M. Buzzelli, S. Bianco, R. Schettini, W. Li, Y. Ma, T. Wang, R. Xu, F. Song, W. Chen, H. Yang, Z. Huang, H. Chang, S. Kuo, Z. Liang, S. Zhou, R. Feng, C. Li, X. Chen, B. Song, S. Zhang, L. Liu, Z. Wang, D. Ryu, H. Bae, T. Kwon, C. Desai, N. Akalwadi, A. Joshi, C. Mandi, S. Malagi, A. Uppin, S. S. Reddy, R. A. Tabib, U. Patil, and U. Mudenagudi, "NTIRE 2022 challenge on night photography rendering," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, Jun., 2022, DOI: 10.1109/CVPRW56347.2022.00135.

- [9] T. Naseer, W. Burgard, and C. Stachniss, "Robust Visual Localization Across Seasons," *IEEE Transactions on Robotics*, vol. 34, no. 2, pp. 289-302, Apr., 2018, DOI: 10.1109/TRO.2017.2788045.
- [10] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. V. Gool, "Night-to-Day Image Translation for Retrieval-based Localization," *International Conference on Robotics and Automation*, Montreal, Canada, May., 2019, DOI: 10.1109/ICRA.2019.8794387.
- [11] V. Tsagaris and V. Anastassopoulos, "Fusion of visible and infrared imagery for night color vision," *Displays*, vol. 26, No. 4-5, pp. 191-196, Oct., 2005, DOI: 10.1016/j.displa.2005.06.007.
- [12] D. Han, Y. Hwang, N. Kim, and Y. Choi, "Multispectral Domain Invariant Image for Retrieval-based Place Recognition," *IEEE International Conference on Robotics and Automation*, Paris, France, 2020, DOI: 10.1109/ICRA40945.2020.9197514.
- [13] Y. Yuan, R. Fu, L. Huang, W. Lin, C. Zhang, X. Chen, and J. Wang, "Hrformer: High-resolution vision transformer for dense predict," *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, 2021, [Online], <https://proceedings.neurips.cc/paper/2021/hash/3bbfdde8842a5c44a0323518eec97cbe-Abstract.html>.
- [14] Y. Choi, N. Kim, S. Hwang, K. Park, J. Yoon, K. An, and I. Kweon, "KAIST Multi-Spectral Day/Night Data Set for Autonomous and Assisted Driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934-948, Mar., 2018, DOI: 10.1109/TITS.2018.2791533.
- [15] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," *Machine Learning*, 2019, DOI: 10.48550/arXiv.1903.12261.



이 승 현

2019~현재 세종대학교 지능기전공학과(학사)

관심분야: 컴퓨터비전, 로봇틱스, 장소 인식 및 인공지능



김 태 주

2021 세종대학교 지능기전공학과(학사)

2021~현재 세종대학교 지능기전공학과
(석박통합과정)

관심분야: 컴퓨터비전, 로봇틱스, 인공지능 및 물체인식



최 유 경

2006 숭실대학교 정보통신전자공학부(학사)

2008 연세대학교 전기전자공학부(석사)

2018 한국과학기술원전기전자공학과/로봇
학제전공(박사)

2018~현재 세종대학교 지능기전공학부
부교수

관심분야: 컴퓨터 비전, 로봇틱스