



# Visual Place Recognition under Substantial Appearance Changes using Event-based Data

**Department of Software Convergence**

2017103749 이현기

지도교수 정지영교수님

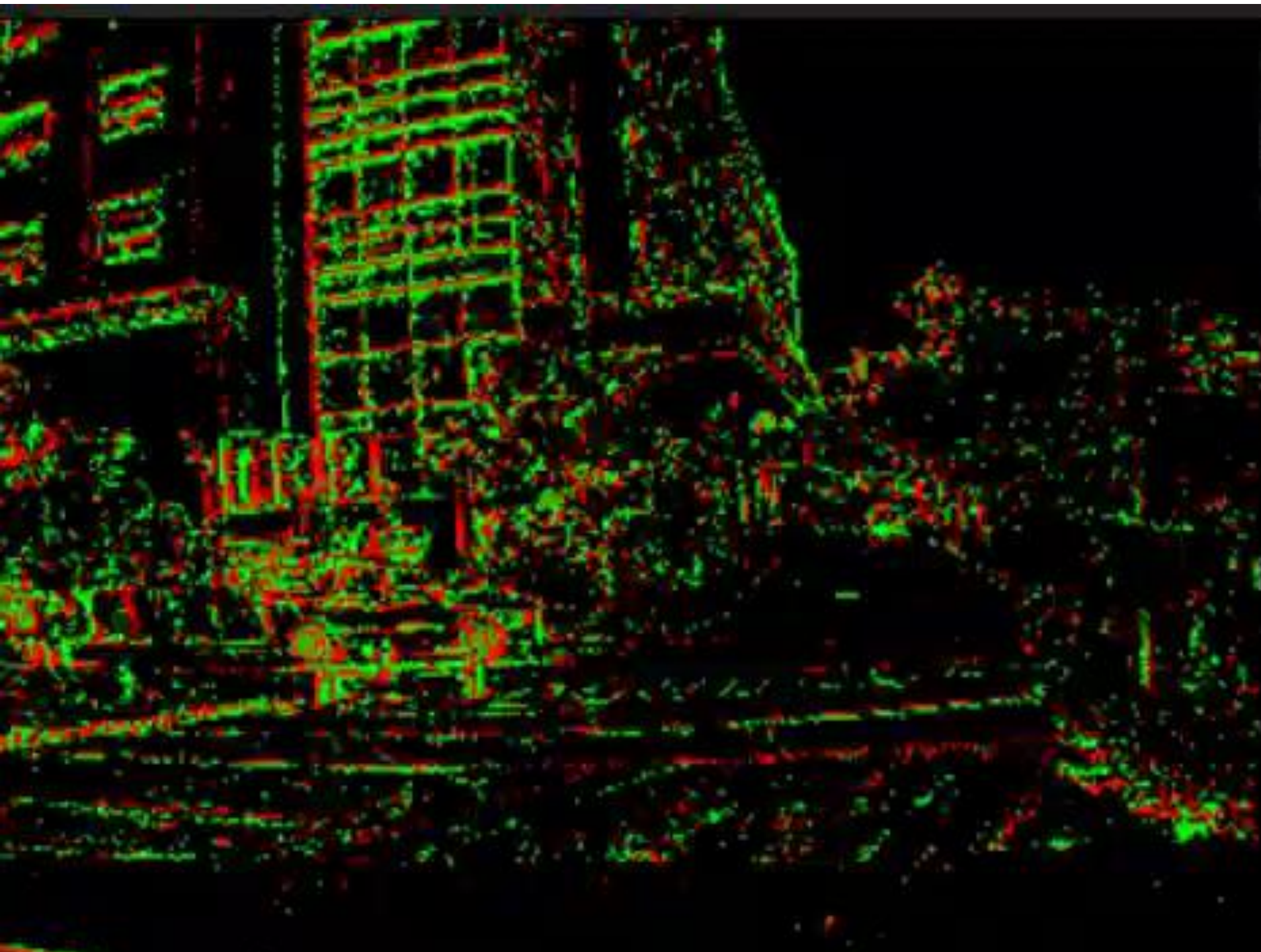
# 이벤트 카메라

- DVS240[1]과 같은 이벤트 카메라(Event Camera)는 시각 센서로, 빛의 밝기의 변화를 감지한다.
- 기존의 카메라와는 다르게, 각 픽셀의 값을 비동기적으로 업데이트한다.
- 따라서 이벤트 데이터는 개별 이벤트들의 집합(흐름,stream)으로 이루어지고, 각각의 이벤트는 발생 시간, 픽셀의 위치, 극성(빛의 밝기 변화에 대한 부호) 정보를 갖는다.









capture/frames  
- blocked

capture/imu  
- blocked

# Visual Place Recognition

- ‘시각 정보를 이용한 장소 인식’으로, 카메라와 같은 시각 센서를 통해 들어온 정보를 이용해, 모바일 로봇의 위치를 추정해내는 작업을 뜻한다.
- SLAM(Simultaneous Localization And Mapping)분야나 자율주행 등의 분야에서 쓰인다.



# 연구 동기

- 기존의 연구들은 다양한 환경 조건에서(조도, 날씨, 시간 등) 이미지를 보고 장소를 매칭하는 데에 어려움을 겪는다.
- 이를 해결하기 위해 다양한 연구가 이루어져 왔고, BoW[2]나 적외선을 이용한 방법[3], 인공 신경망을 이용한 낮-밤 관계를 학습시키는 방법[4] 이외에도 SeqSLAM[5], Graph-Based[6]와 같은 매칭 알고리즘을 이용한 후처리 방법 등의 연구가 활발히 진행되고 있다.
- 이벤트 카메라는 일반 카메라에 비해 높은 감도를 가지고 있고, 환경의 영향을 적게 받기 때문에 이벤트 카메라를 이용해 효과적인 장소 인식을 진행할 수는 없을까 라는 생각을 하게 되었고, 해당 연구를 진행하게 되었다.



# 연구 목표

1. 이벤트 카메라를 이용하여 시각적인 장소 인식이 가능함을 보인다.
2. 매칭 정확도 90% 이상
3. 기존의 일반 카메라에 적용 가능한 알고리즘을 이벤트 카메라에도 적용 가능함을 보인다.
4. 이벤트 카메라를 place recognition에 사용하였을 때 일반 카메라에 비해 성능이 나음을 보인다.

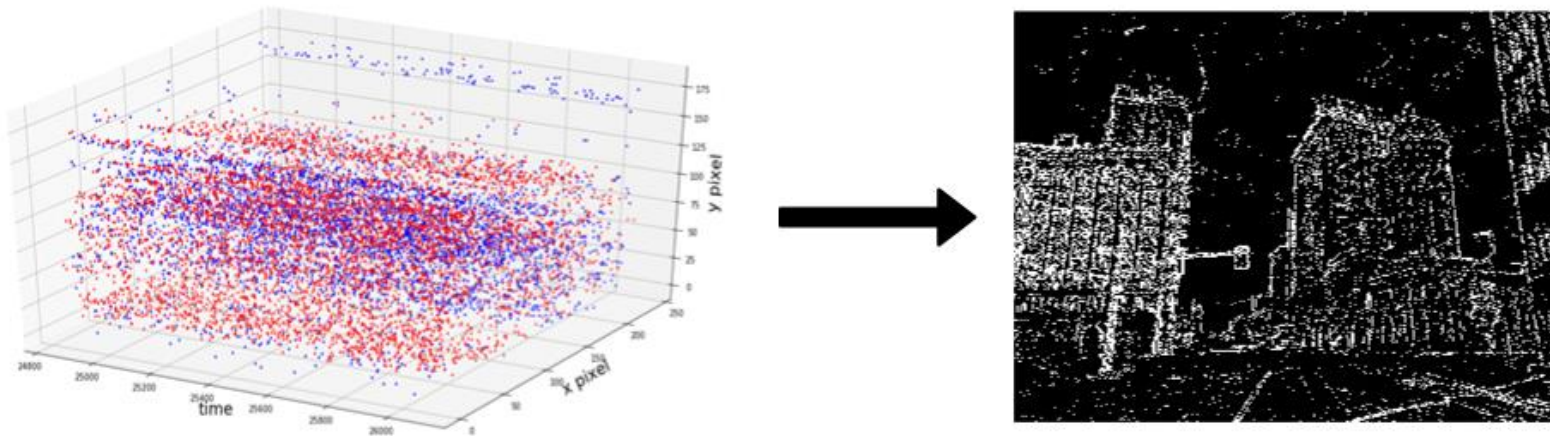




# 연구 방법

## 1. 이벤트-이미지 변환 (Event-Image Conversion)

- 이벤트 데이터는 비동기적으로 주어지기 때문에, CNN네트워크의 입력으로 주기 위해서는 2차원 이미지의 형태로의 변환이 필요하다.



이벤트 데이터는 4차원 데이터( $x, y, t, p$ )로, CNN에 적용하기 위해서는 2차원 또는 3차원 데이터로 변환해야 한다.



# 연구 방법

## 2. 네트워크 학습 (Training Network)

장소 식별을 위한 네트워크는 기본적인 CNN 네트워크를 사용하였다. LeNet[7]을 참고하여 네트워크를 구성하였고, 아래의 표는 CNN의 파라미터를 설명한다.

Layer	Value
Input Image	$R_x \times R_y \times C$
MP(ReLu(Conv1))	input_channel= $C$ , output_channel=6
Batch_Norm	channel=6, eps= $1e-05$ , momentum=0.1
MP(ReLu(Conv2))	input_channel=6, output_channel=16
Batch_Norm	channel=16, eps= $1e-05$ , momentum=0.1
MP(ReLu(Conv3))	input_channel=16, output_channel=32
ReLu(FullyConnected1)	in_features=7040, out_features= $2 \times N$
ReLu(FullyConnected2)	in_features= $2 \times N$ , out_features= $1.5 \times N$
ReLu(FullyConnected3)	in_features= $1.5 \times N$ , out_features= $N$

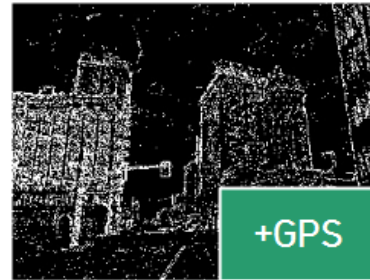


# 연구 방법

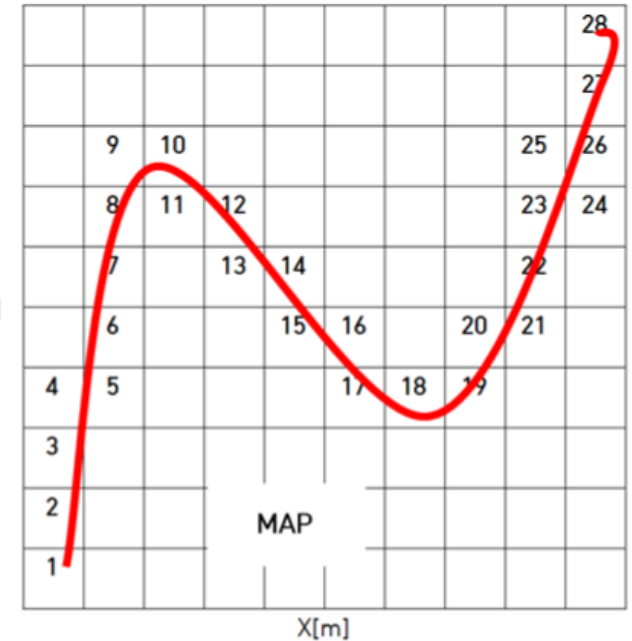
## 3. 장소 추정 및 라벨링

(Localization and Class Labeling)

맵 전체를 일정한 간격으로 나누고,  
이벤트 이미지와 그에 해당하는  
GPS 데이터가 들어오면 로봇이 지나  
는 맵의 구역을 알 수가 있다. 각각의  
구역이 고유한 class가 된다.



Classification



TRAJECTORY

# 연구 방법

## 4. 매칭 알고리즘 (Matching)

매칭 정확도 향상을 위해 연속된 이미지 사이의 구속 조건을 이용하여 Sliding Window Filter 기법을 적용하였다.

---

### Algorithm 1: Matching algorithm

---

**Result:** Match\_vector

Match\_vector[0] = index(max(p[0])), i = 1;

**while**  $i \leq N_q$  **do**

    previous = Match\_vector[i-1];

    local\_max = index(max(p[i][previous-K:previous+K]));

**if**  $p[i][local\_max] > \tau_2$  **then**

        Match\_vector[i] = index(local\_max);

        i += 1;

**else**

        global\_max = index(max(p[i]));

**if**  $p[i][global\_max] > \tau_1$  **then**

            Match\_vector[i] = index(global\_max);

            i += 1;

**else**

            Match\_vector[i] = -1;

            i += 1;




**end**

**end**

**end**

---

# 데이터셋

	GTA[8][9]	Oxford Robot Car[10]	KH-Campus
			
타입	가상	실제	실제
장소	GTA V5 인게임	영국 옥스포드 대학교	경희대학교 국제캠퍼스
이벤트 데이터	이벤트 카메라 시뮬레이터[11]	이벤트 카메라 시뮬레이터	이벤트 카메라 DVXplore Lite
데이터셋 종류	맑은 낮, 야간, 우천 (2개의 경로)	맑은 낮, 야간	맑은 낮, 저녁, 밤
카메라 시점 오차	없음	큼	작음



# 실험1 - 컬러 이미지와의 성능 비교

- 첫 번째 실험은 이벤트 카메라의 이미지가 일반적인 컬러 이미지에 비해 해당 지역의 특징 정보를 더 잘 담고 있어 매칭 성능이 더욱 높아짐을 보이기 위한 실험이다.
- GTA 데이터셋과 Oxford Robot Car 데이터셋을 이용하였고 train 데이터셋으로 맑은 날 데이터셋을, test 데이터셋으로 비오는 날과 야간의 데이터셋을 이용하였다.
- 이벤트카메라 이미지와 일반 카메라 이미지 모두에 대해 진행하였다.



# 실험2 - 기존 방법들과의 성능 비교

- 기존의 이미지-장소 매칭 방법에는 SeqSLAM이나 Graph-Based알고리즘 등이 있는데, 이러한 알고리즘과 우리의 알고리즘에 이벤트 카메라 이미지를 적용하여 Precision-Recall선도를 그림으로써 성능을 비교하였다.



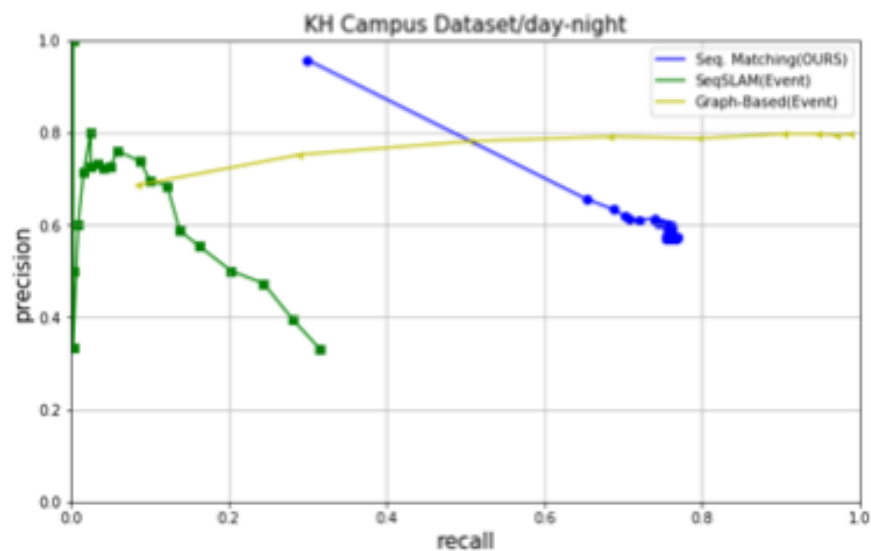
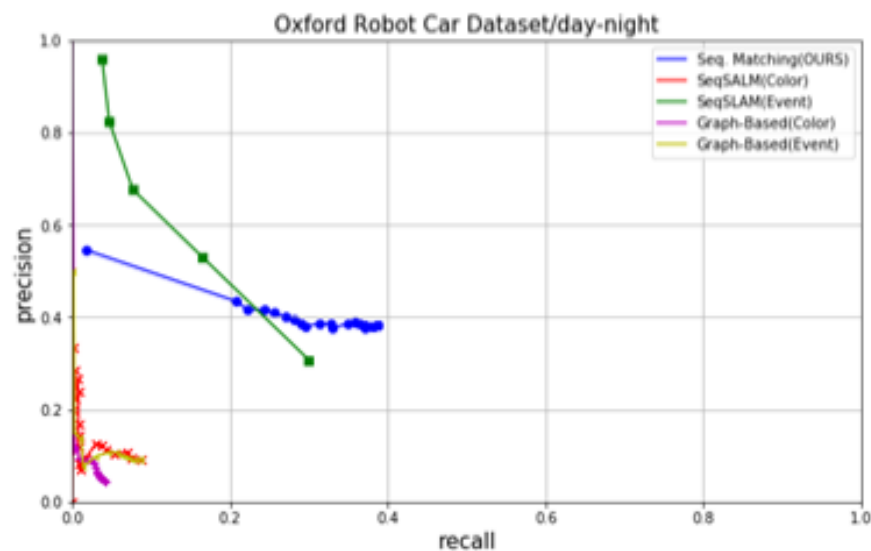
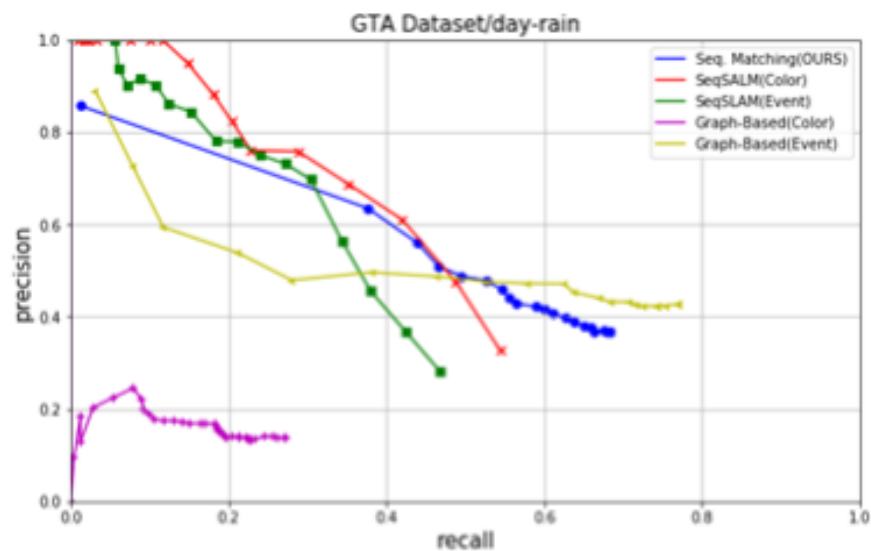
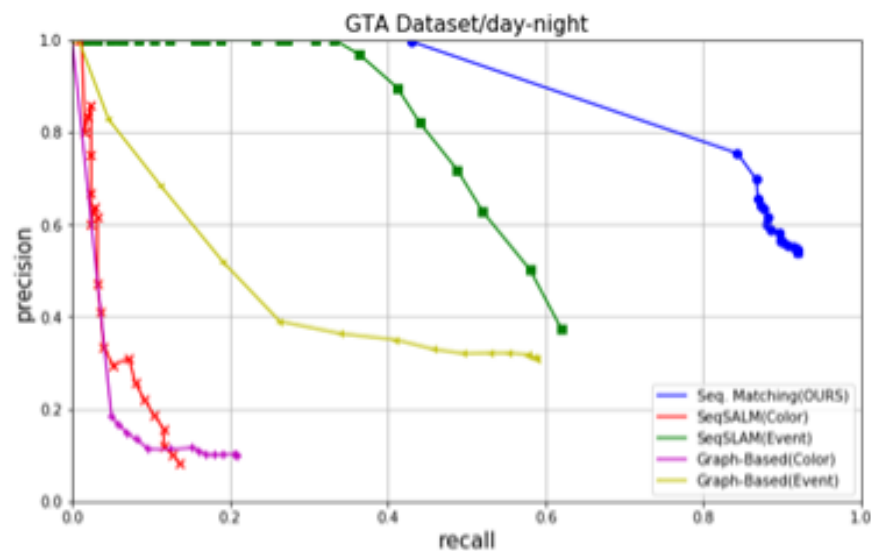
# 실험1 결과 - 컬러 이미지와의 성능 비교

Table 2: The accuracy while using the event images compared to color images. The superscript 1 and 2 of the dataset means the trajectory type of the GTA dataset.

Global-Max				Sequence Matching			
Database	Query	Event	Color	Database	Query	Event	Color
clean <sup>1</sup>	night <sup>1</sup>	85.80	3.39	clean <sup>1</sup>	night <sup>1</sup>	93.20	3.79
clean <sup>1</sup>	rain <sup>1</sup>	75.42	31.36	clean <sup>1</sup>	rain <sup>1</sup>	81.34	38.22
clean <sup>1</sup>	clean <sup>2</sup>	55.58	34.07	clean <sup>1</sup>	clean <sup>2</sup>	91.28	55.35
clean <sup>1</sup>	night <sup>2</sup>	46.04	0.58	clean <sup>1</sup>	night <sup>2</sup>	77.44	21.27
clean <sup>1</sup>	rain <sup>2</sup>	33.26	19.77	clean <sup>1</sup>	rain <sup>2</sup>	56.74	46.16
SeqSLAM				Graph-Based			
Database	Query	Event	Color	Database	Query	Event	Color
clean <sup>1</sup>	night <sup>1</sup>	98.14	49.13	clean <sup>1</sup>	night <sup>1</sup>	85.49	24.77
clean <sup>1</sup>	rain <sup>1</sup>	98.14	93.21	clean <sup>1</sup>	rain <sup>1</sup>	88.55	34.62
clean <sup>1</sup>	clean <sup>2</sup>	32.33	76.98	clean <sup>1</sup>	clean <sup>2</sup>	10.93	10.93
clean <sup>1</sup>	night <sup>2</sup>	40.70	9.30	clean <sup>1</sup>	night <sup>2</sup>	25.35	6.05
clean <sup>1</sup>	rain <sup>2</sup>	36.74	50.47	clean <sup>1</sup>	rain <sup>2</sup>	37.91	6.28



# 실험2 결과





# 결론

- 이벤트 카메라를 이용한 장소 인식 방법을 새롭게 제시하였고, 이벤트 데이터를 이미지로 변환하여 CNN을 적용시켰다.
- CNN에서 출력된 확률 벡터를 이용하여 localization을 더욱 효과적으로 수행할 수 있었고 정확도를 높일 수 있었다.
- 첫 번째 실험에서, 우리는 동적 환경에서 이벤트 이미지가 기존의 컬러 이미지에 비해 더 강건함을 보였고, 두 번째 실험에서 우리는 기존의 방법들과 비교했을 때 비슷하거나 더 높은 성능을 보이는 것을 확인할 수 있었다.



# 후속 연구

- 이벤트 데이터를 CNN의 입력(2차원 이미지)로 변환하기 위해서는 상당한 양의 데이터 손실이 발생한다. 이를 보완하기 위해 생체-모방형(Neuromorphic) 비동기 네트워크인 SNN(Spiking Neural Network)을 사용한 연구를 진행하고 있고, 이벤트 데이터의 특성을 이용한 새로운 매칭 방법을 모색하고 있다.



- [1] P. Lichtsteiner, C. Posch, and T. Delbruck. A  $128 \times 128$  120 db 15  $\mu$ s latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits, 43(2): 566-576, 2008.
- [2] D. Galvez-López and J. D. Tardos. Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics, 28(5):1188-1197, 2012.
- [3] W. Maddern and S. Vidas. Towards robust night and day place recognition using visible and thermal imaging. 07 2012.
- [4] P. Neubert, N. Sünderhauf and P. Protzel, "Appearance change prediction for long-term navigation across seasons," 2013 European Conference on Mobile Robots, Barcelona, 2013, pp. 198-203, doi:10.1109/ECMR.2013.6698842.
- [5] M. J. Milford and G. F. Wyeth. Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In 2012 IEEE International Conference on Robotics and Automation, pages 1643-1649, 2012.
- [6] O. Vysotska and C. Stachniss. Lazy data association for image sequences matching under substantial appearance changes. IEEE Robotics and Automation Letters, 1(1): 213-220, 2016.
- [7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278-2324, 1998.
- [8] Grand theft auto 5. <https://www.rockstargames.com/V/>. Accessed: 2020- 04-01.
- [9] A.-D. Doan, A.M. Jawaid, T.-T. Do, and T.-J. Chin. G2D: from GTA to Data. arXiv preprint arXiv:1806.07381, pages 1-9, 2018.
- [10] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. The International Journal of Robotics Research (IJRR), 36(1):3-15, 2017. doi: 10.1177/0278364916679498. URL <http://dx.doi.org/10.1177/0278364916679498>.
- [11] H. Rebecq, D. Gehrig, and D. Scaramuzza. Esim: an open event camera simulator. In Conference on Robot Learning, pages 969-982, 2018.

# THANK YOU.

