Robust Place Recognition of Robot in Urban Canyon Using Neural Network

Jongwon Lee (jongwon.lee@kaist.ac.kr)

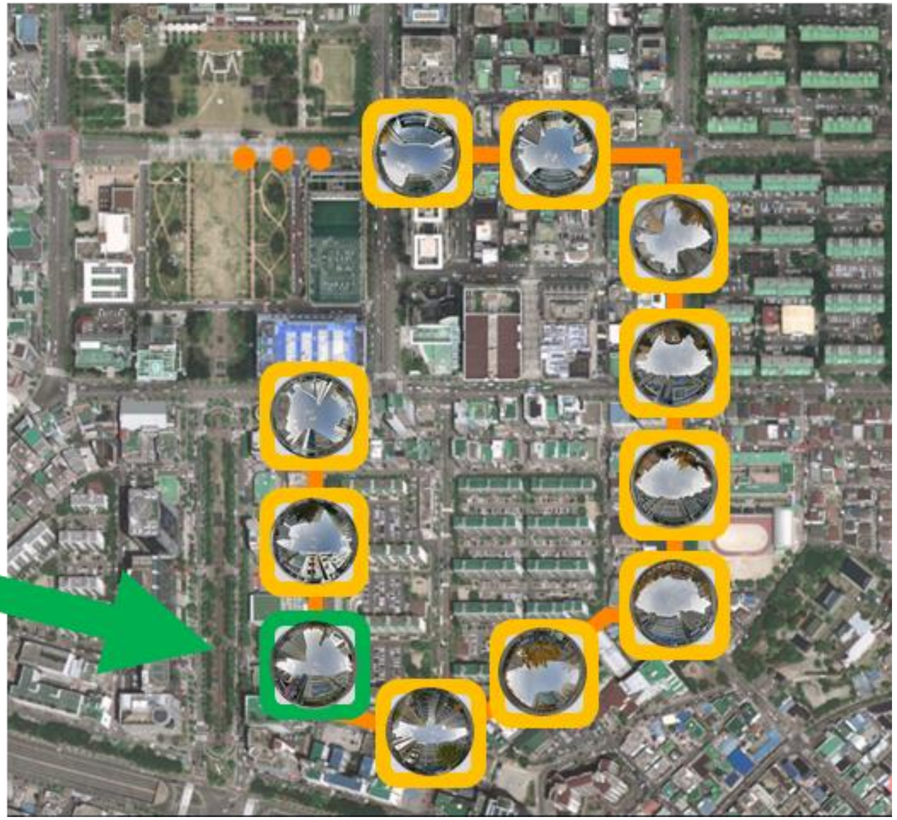
Ayoung Kim* (ayoungk@kaist.ac.kr)

Korean Advanced Institute of Science and Technology

Abstract

Query (Sep. 2018)

Trained (Nov. 2014)

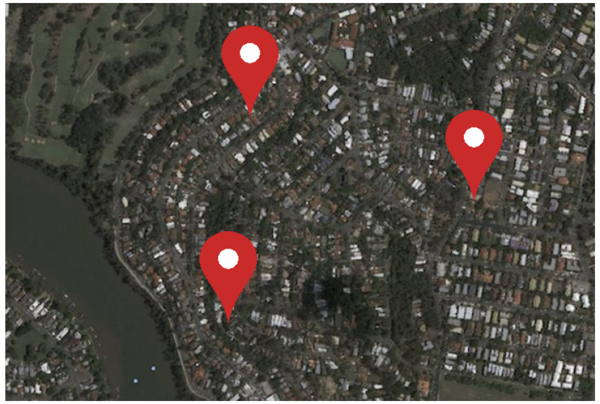


Motivation

Visual Place Recognition
in Urban Environment

(Kim et al. 2017)

Observed Scene



Find *where I am*

Challenges in
Revisiting the Same Place

(Sunderhauf et al. 2015)

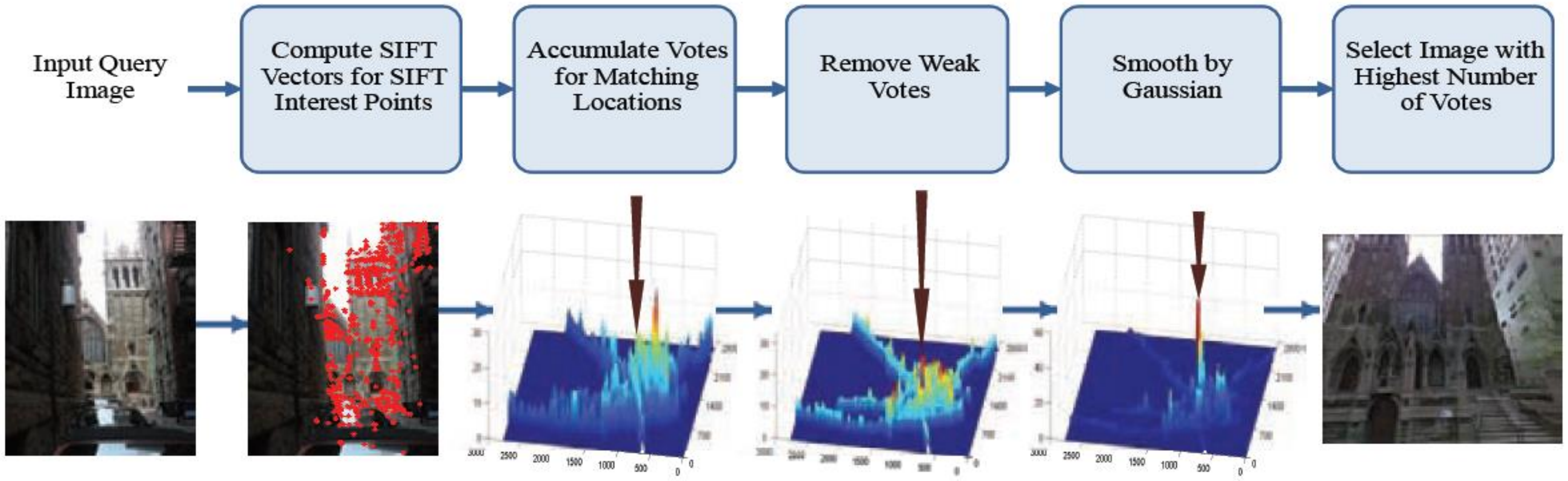
1. Short-term Appearance Change

2. Long-term Appearance Change

Related Works

1. Using global feature
   - GIST (Murillo 09, Singh 10)
2. Using local feature
   - Bag-of-Words (Cummins 08)
   - Hierarchical manner (Zamir 10)

(Zamir et al. 2010)



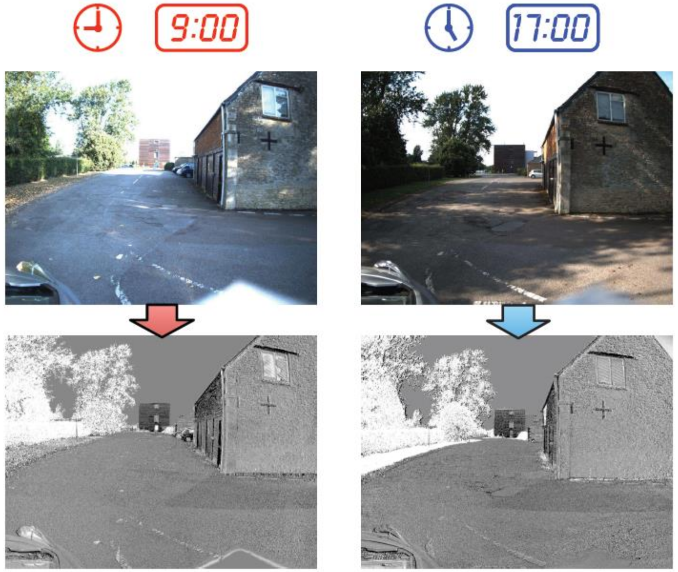
→ Not suitable for the long-term place recognition…
How to overcome spatio-temporal differences?

KAIST

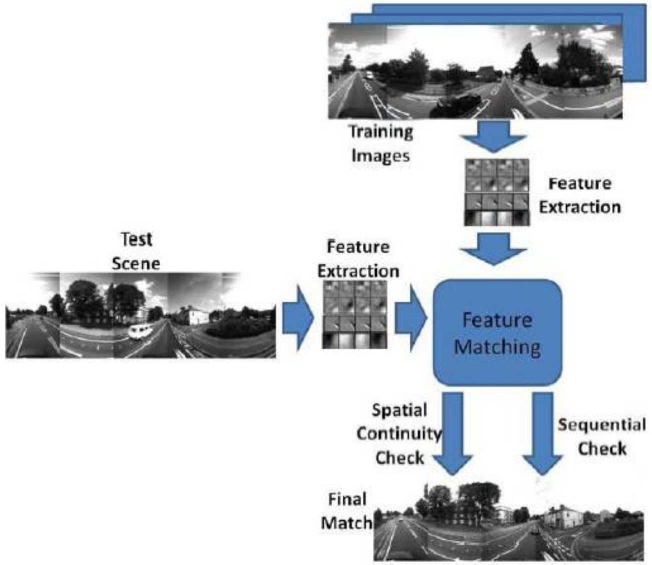
Related Works

1. Illumination-invariant imaging
   - ZNCC (Maddern 14, McManus 14)
2. The first place recognition based on a CNN model
   - Feature & Sequential matching (Chen 14)

(Maddern et al. 2014)



(Chen et al. 2014)

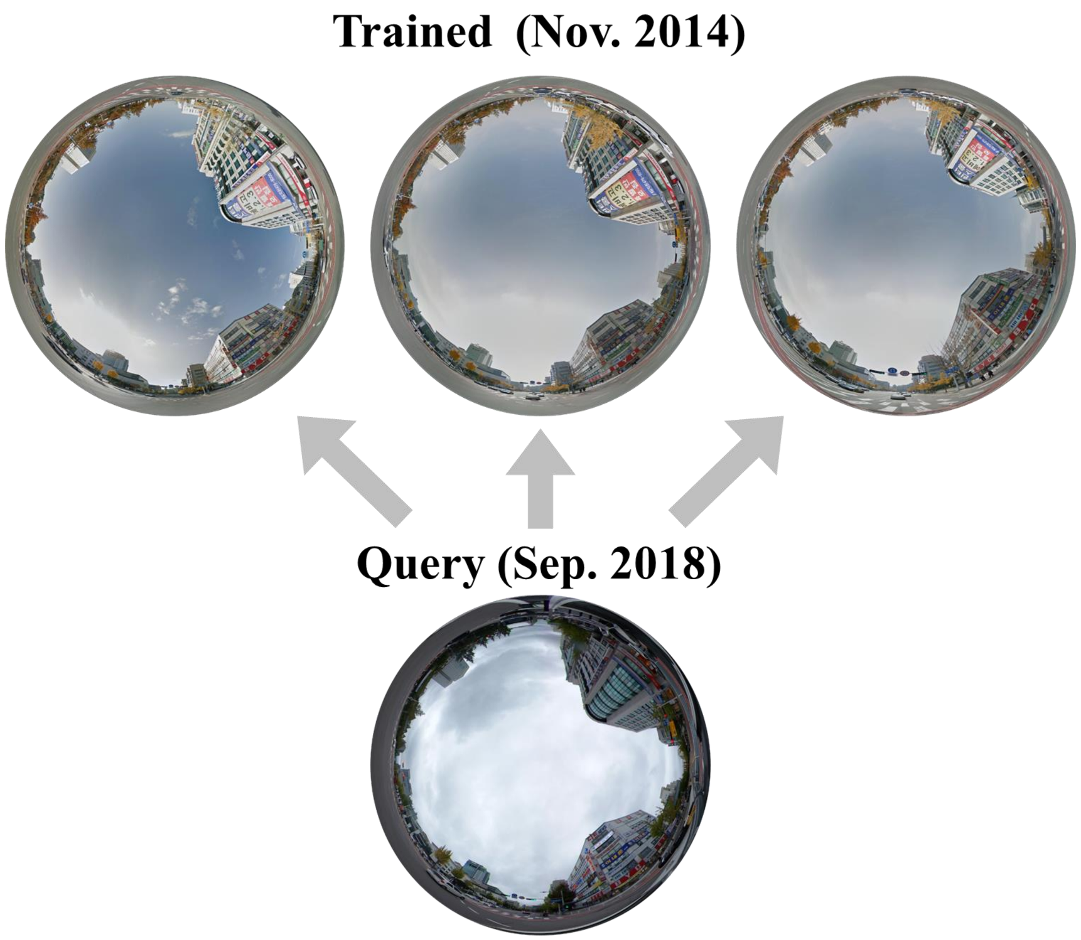


→ Combine and take advantage of both methods
to overcome long-term (*seasons and years*) changes

Procedure

Find the match between a query and a corresponding image

→ Feature matching is done by *end-to-end supervised learning*



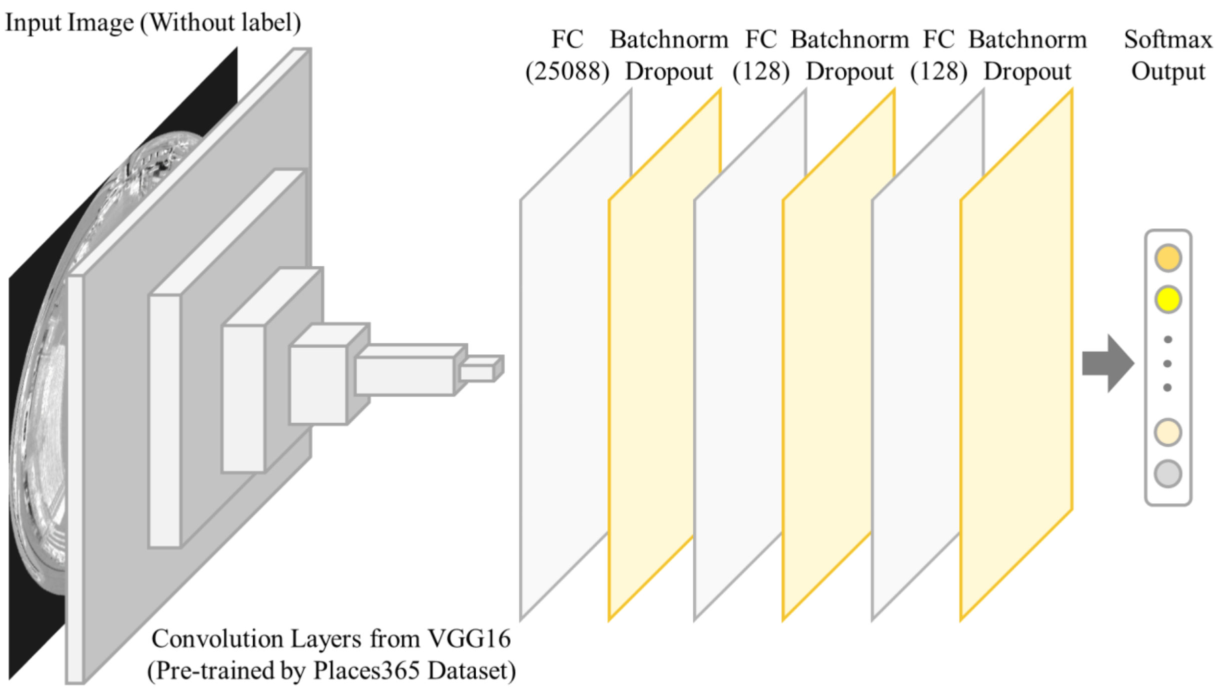
Procedure

▪ Neural Network Design

A. Convolution layers
   - Based on VGG16, pre-trained by Places365 Dataset

B. Dense layers
   - Optimal depth and width are found by heuristic approach

C. Prediction
   - Print out five predicted scenes which might be matched with the query image



Procedure

Use *fisheye image* instead of generic image taken in urban environment

1. Independent to the direction
2. Wide field of view → Contains less-variant visual contexts (e.g. skylines, buildings…)

Generic images
(Temporally variant)

Fisheye image
(Temporally invariant)

KAIST

Procedure

▪ Data Acquisition

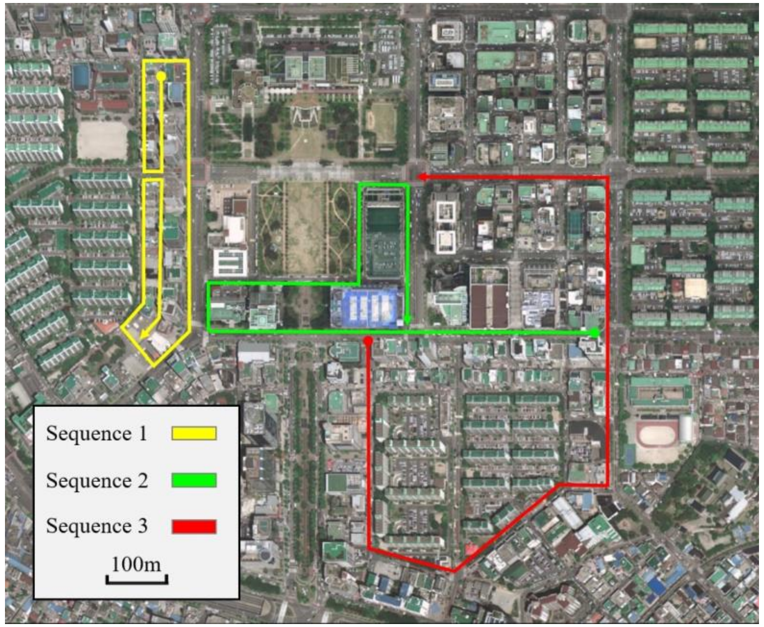
Three different paths in an urban area (Daejeon, Korea)

B. Training images

- Download panorama images from Google Streetview and convert them into spherical form (2014.11)

C. Test images

- Taken by an omni-directional camera (2018.09)



Sequence type	Length [km]	Number of images
Sequence 1	1.64	137
Sequence 2	1.47	136
Sequence 3	1.61	132

* Image data augmentation: Representative method utilized to increase the performance of deep-learning network by giving variant onto training dataset such as rotating, warping, re-coloring and so on.

Procedure

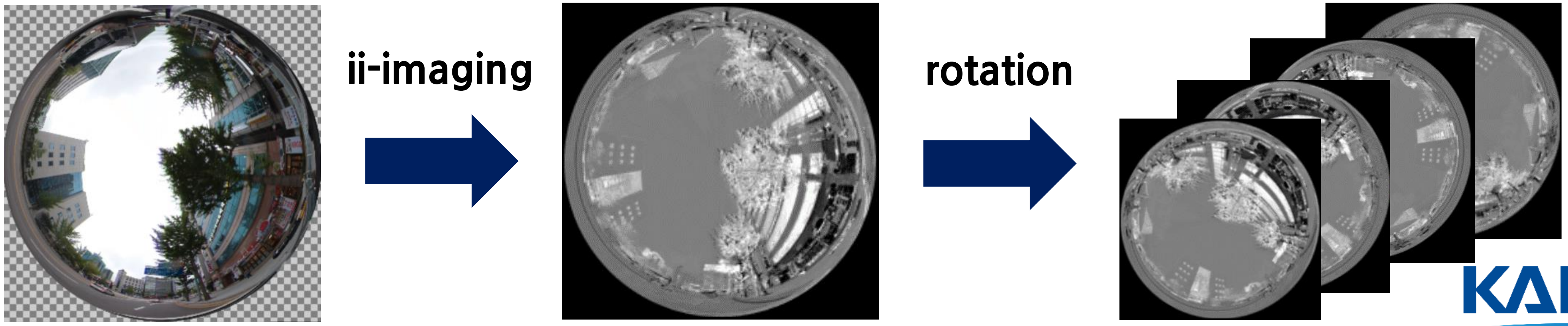
▪ Image Augmentation*

A. Illumination-invariant imaging

- Project RGB channel of original image onto single illumination invariant channel
- Robust to color and illumination variance of image dataset
- $I = \log R_2 - \alpha \log R_1 - (1-\alpha) \log R_3$ where R_1, R_2, and R_3 are three color channels
- α ranges from 0.1 to 0.9 at 0.1 intervals in training, and set to be 0.5 in test

B. Rotation

- Reduce directional variance
- Rotated at 30 degrees from 0 to 360 degrees



* Top-N accuracy: correct class gets to be in the Top-N probabilities for it to count as “correct”

Result

Top-1 and Top-5 Accuracies*

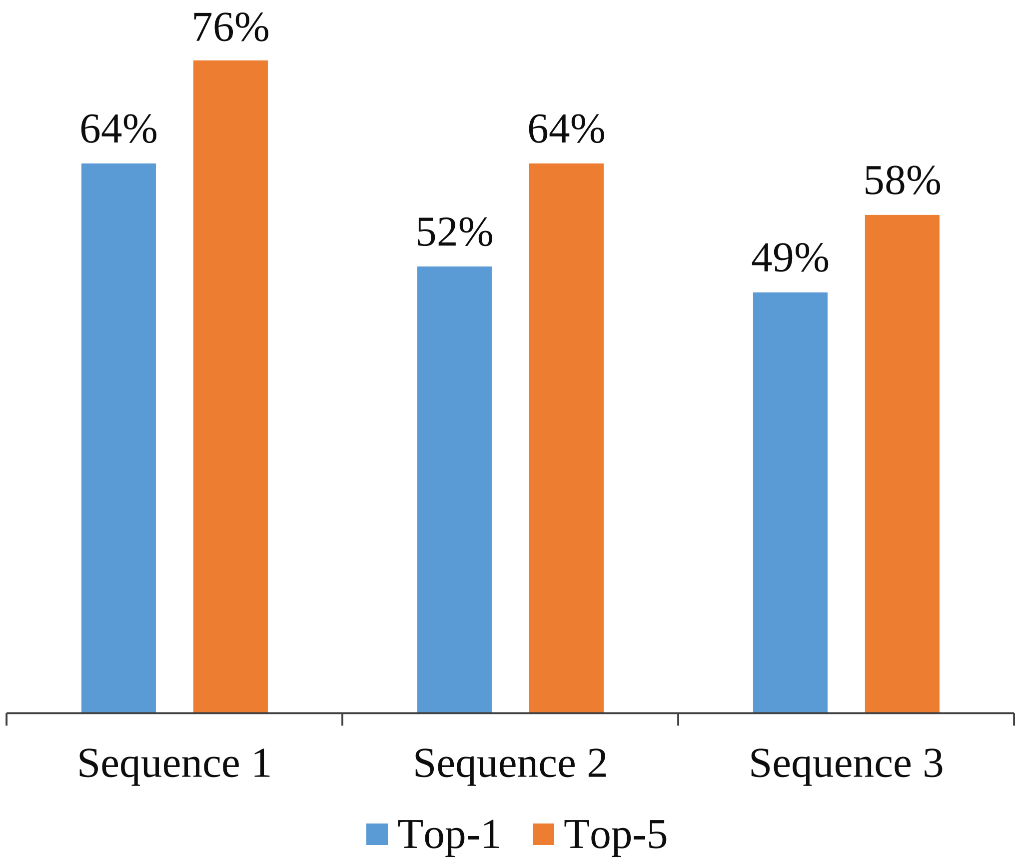
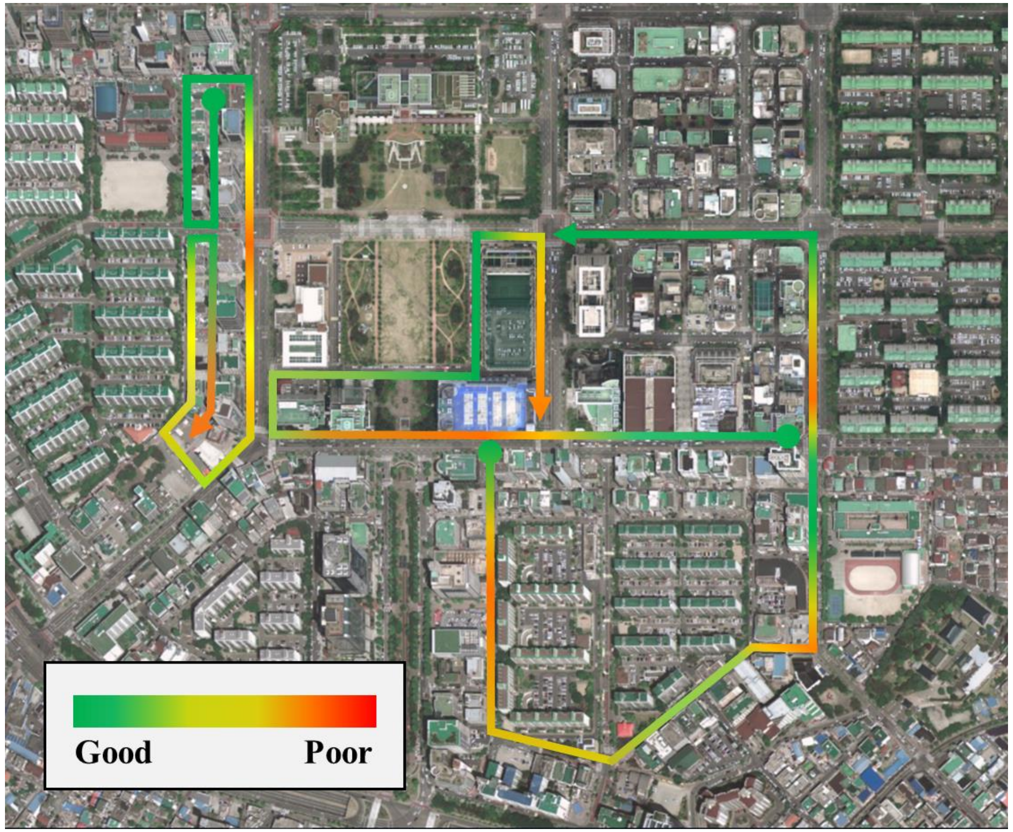


Illustration of test results

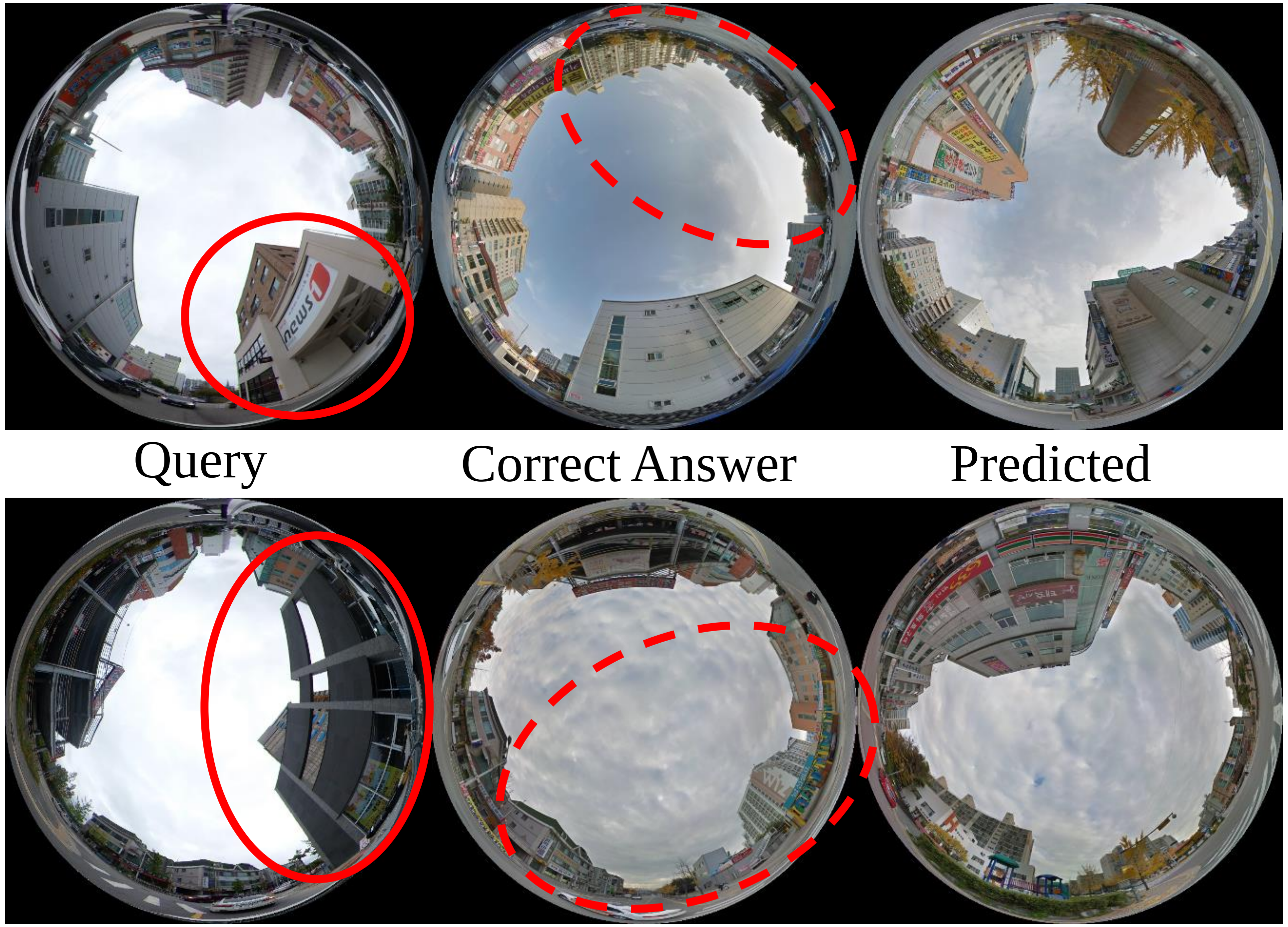


* Query: Taken in 2018.09. by omni-image camera
* Correct answer & Predicted: Taken in 2014.11. by Google Streetview

Result

▪ Unmatched Cases

A. Newly-built building

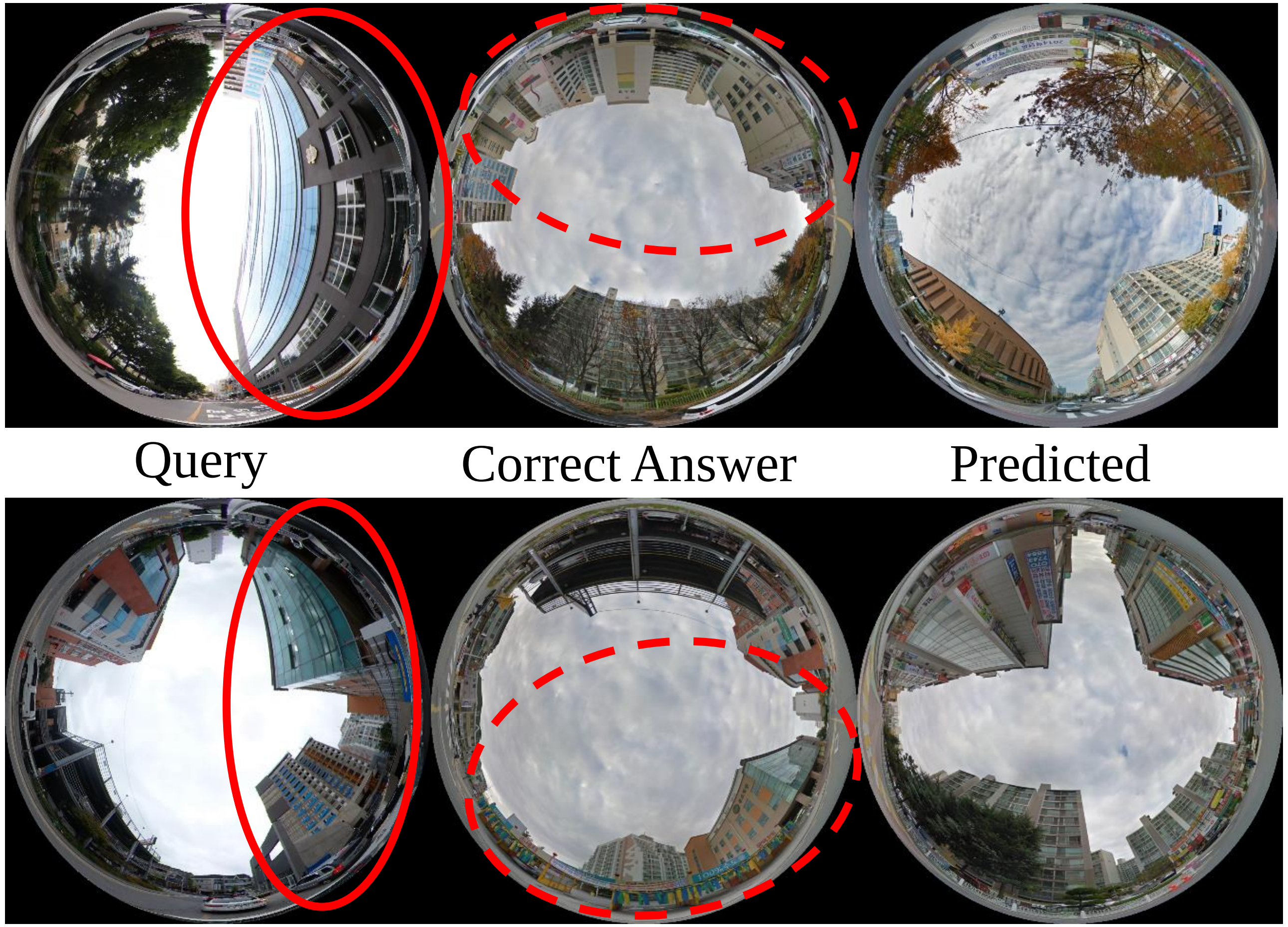


* Query: Taken in 2018.09. by omni-image camera
* Correct answer & Predicted: Taken in 2014.11. by Google Streetview

Result

▪ Unmatched Cases

A. Newly-built building



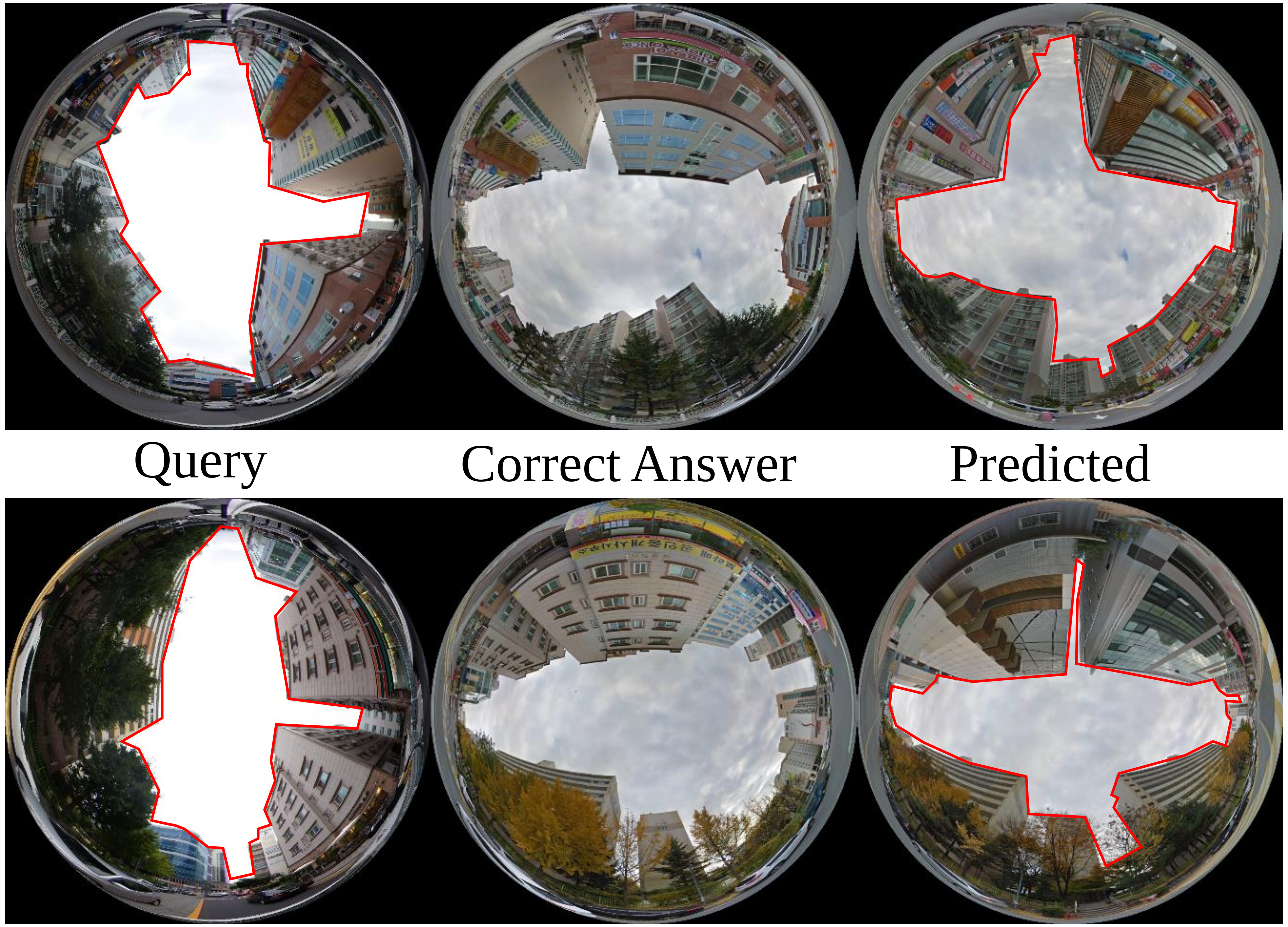
* Query: Taken in 2018.09. by omni-image camera

* Correct answer & Predicted: Taken in 2014.11. by Google Streetview

Result

- **Unmatched Cases**

B. Similar structure



* Query: Taken in 2018.09. by omni-image camera
* Correct answer & Predicted: Taken in 2014.11. by Google Streetview

Result

▪ Unmatched Cases

C. Lack of features

Query   Correct Answer   Predicted

KAIST

Discussion

- The suggested method classifies urban scenes well in spite of seasonal, illuminant, and viewpoint changes cause by *the four-year gap*.
- However, severe change of landscape during four years makes the task hard.
- Follow-up study on a proper far-long-term localization by spherical image should be conducted.

KAIST

Acknowledgement

- This work was supported by the Department of Civil and Environment Engineering, KAIST through ‘CEE-URP Program’ and by KAIST through ‘Undergraduate Research Participation Program.’

Questions

Video by omnidirectional camera, mounted on vehicle (2018.09)

KAIST