2021 ICCV OCTOBER 11-17 VIRTUAL

Pose Correction for Highly Accurate Visual Localization in Large-scale Indoor Spaces

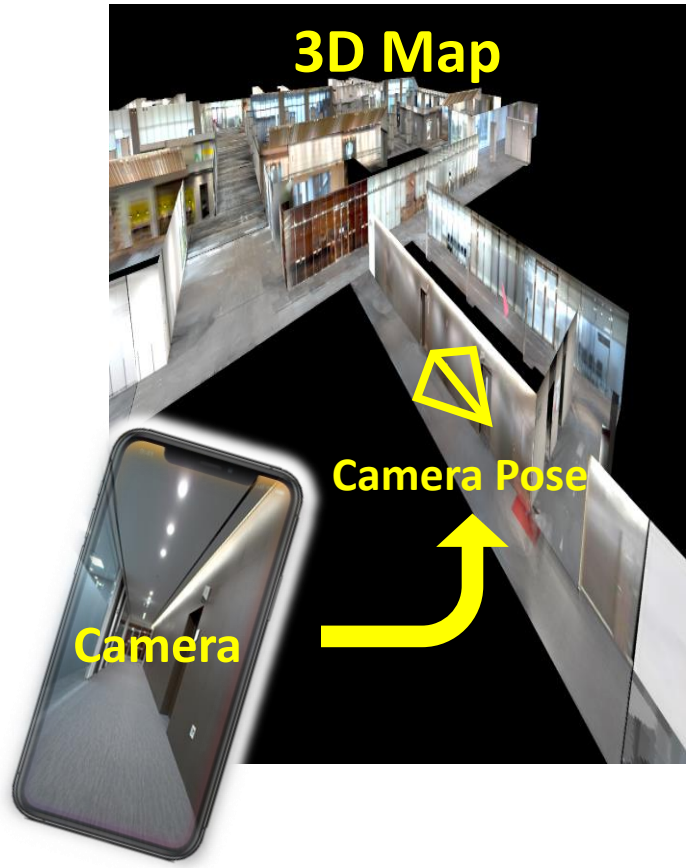
Janghun Hyeon*[1], Joohyung Kim*[1], Nakju Doh[1,2]
Korea University[1], TeeLabs[2]

Equal Contribution*

KOREA UNIVERSITY

TEE LABS

Visual Localization

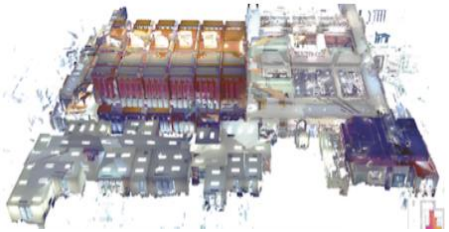


Goal

- Long-term visual localization

  - Moving object
  - Dynamics
  - Illumination Changes

- Large-scale Indoor spaces



  - InLoc: 25,287 m^2
  - Featureless
  - Sparse Databases

Hierarchical Methods

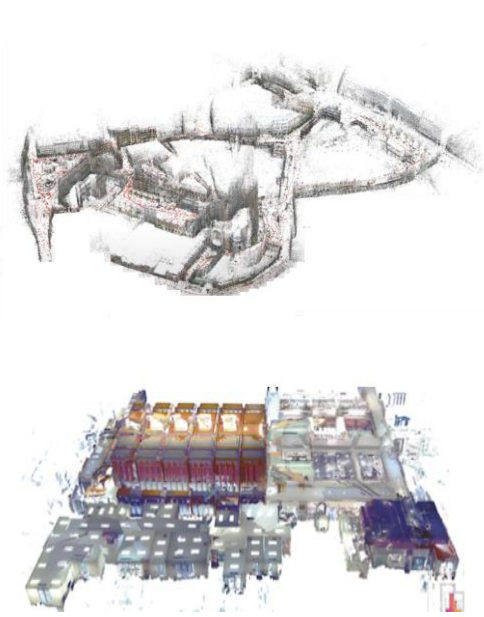
Various Visual Localization methods

Absolute pose regression-based method

Relative pose regression-based method

Scene coordinate regression method

Hierarchical Method

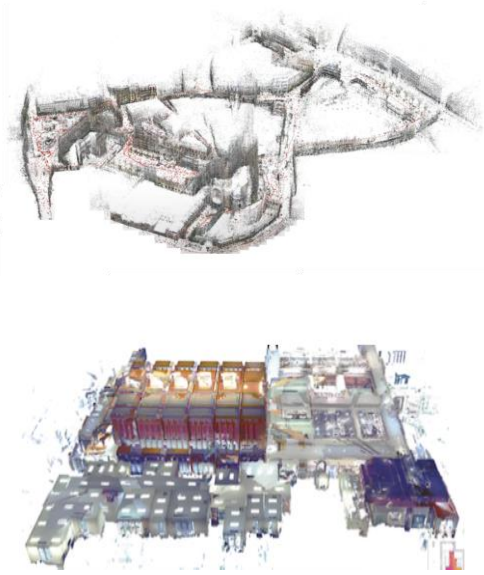


Large-scale Spaces

Hierarchical Methods

Various Visual Localization methods

- *Absolute pose regression-based method*
- *Relative pose regression-based method*
- *Scene coordinate regression method*
- *Hierarchical Method*



Large-scale Spaces

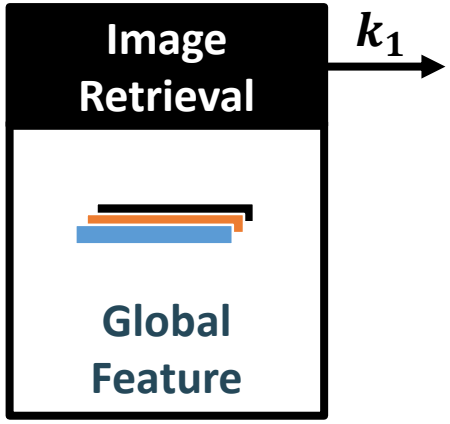


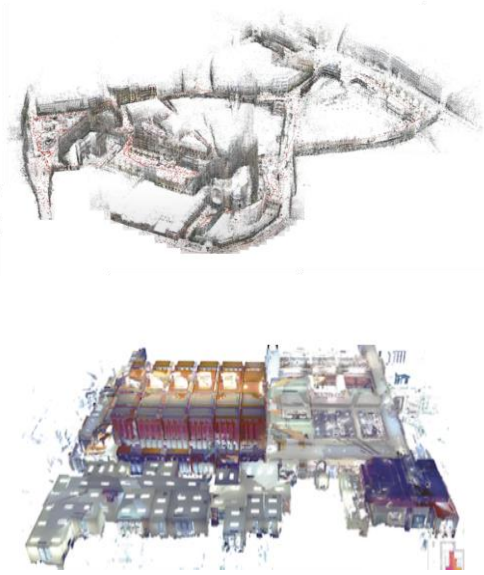
Image Retrieval

- **Uses global descriptors to retrieve top k_1 images from the database**

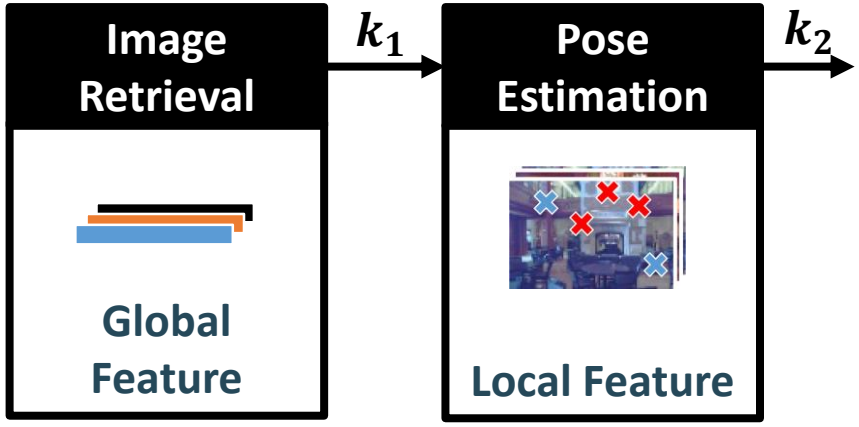
Hierarchical Methods

Various Visual Localization methods

- *Absolute pose regression-based method*
- *Relative pose regression-based method*
- *Scene coordinate regression method*
- *Hierarchical Method*



Large-scale Spaces



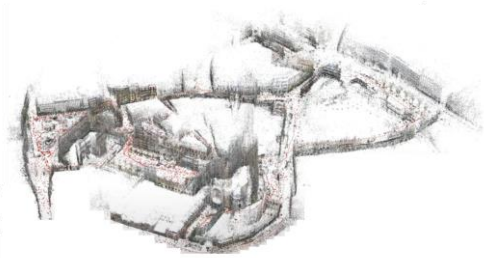
Pose Estimation

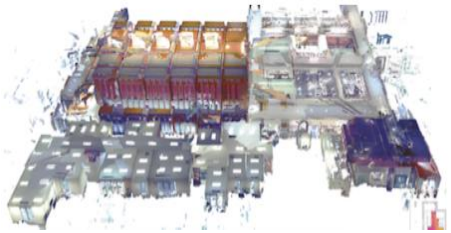
- **Finds 2D-3D correspondences and perform PnP algorithm in a RANSAC loop.**
- **Yields top k_2 candidates**

Hierarchical Methods

Various Visual Localization methods

- *Absolute pose regression-based method*
- *Relative pose regression-based method*
- *Scene coordinate regression method*
- *Hierarchical Method*





Large-scale Spaces

Image Retrieval → k_1 → **Pose Estimation** → k_2 → **Pose Selection**

Global Feature

Local Feature

3D Information

Pose Selection

- **Selects the best candidate as a final pose by a method such as pose verification**

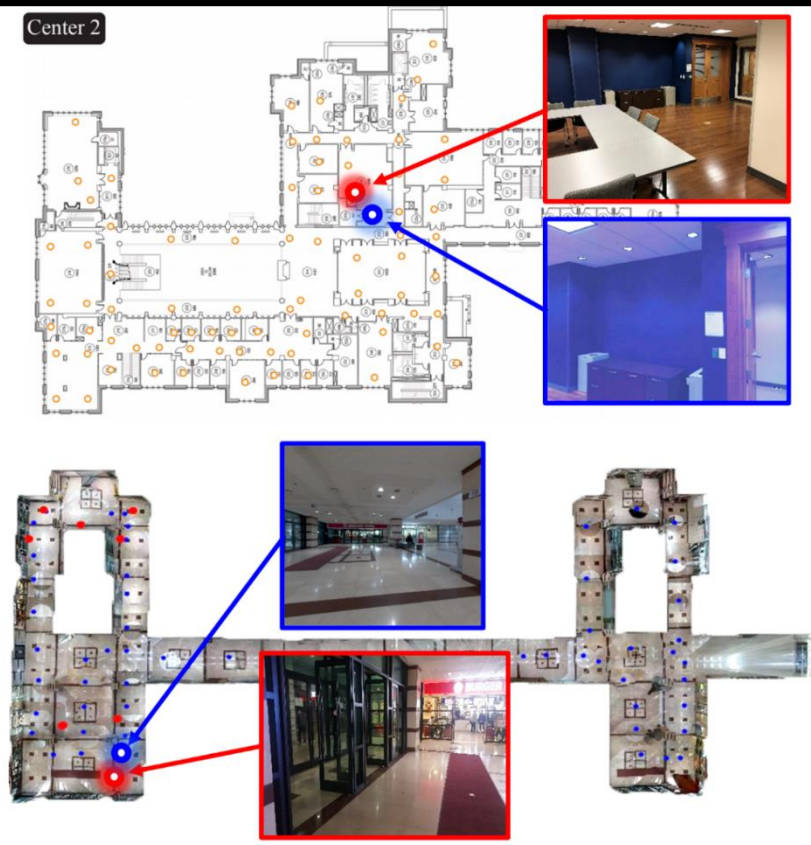
Challenge: Large-scale Indoor Space

Main challenge of hierarchical methods in large-scale indoor spaces
- Sparsity of camera locations in the database

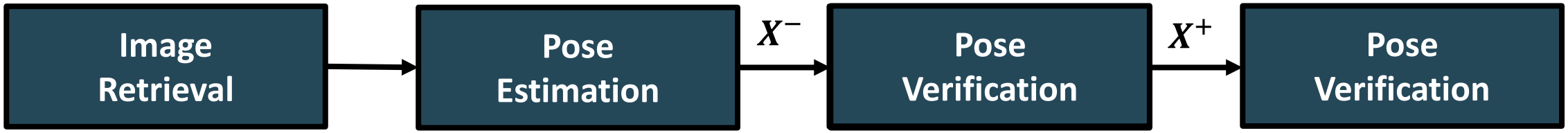
Dataset	# of Camera Locations	# of DB images	Area
7-Scenes	26,000	26,000	$31.5m^3$
12-Scenes	240,002	240,002	$521m^3$
M-site	**720**	25,920	**$12,557m^2$**
InLoc	**277**	9,972	**$25,287m^2$**

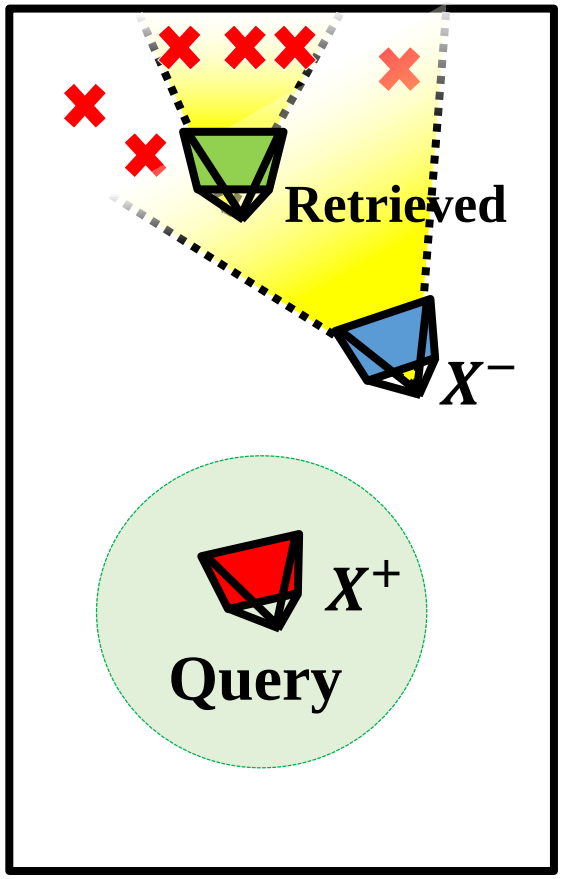
- Difficult to construct a database with densely captured images in large-scale indoor spaces
- Retrieved images may be captured far from the query pose

View-Difference Problem



Pose Correction





Pose Estimation

Local Feature (Query)

Local Feature (Database)

View-Difference Problem

Pose Correction

Local Feature (Query)

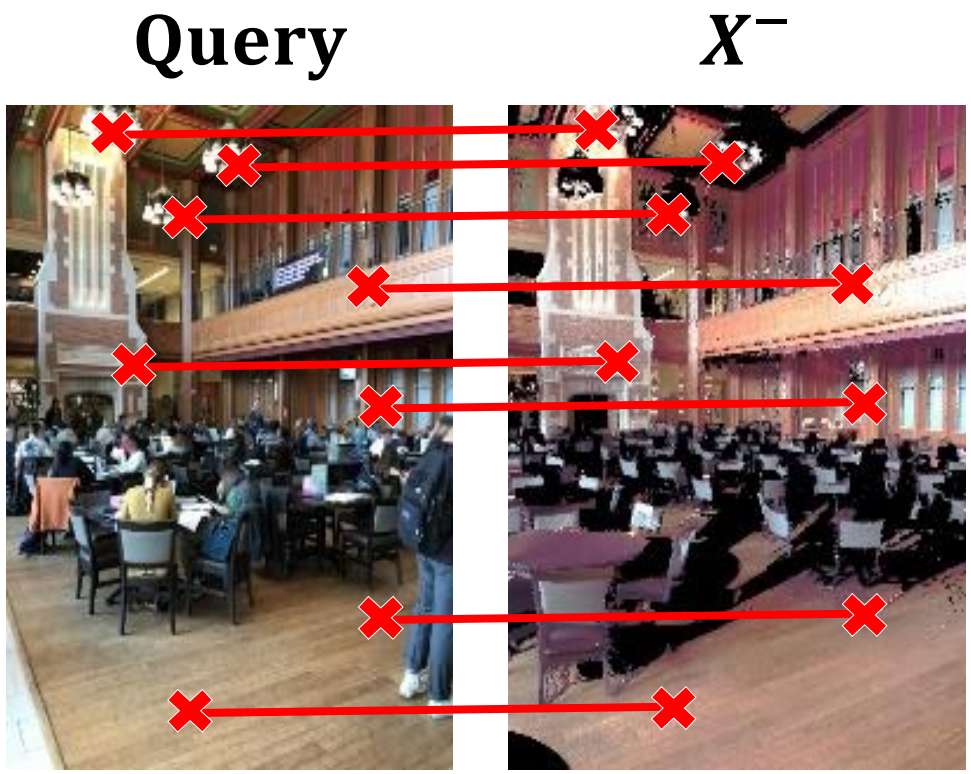
Local Feature (X^-)

Reorganize Visible Local Features

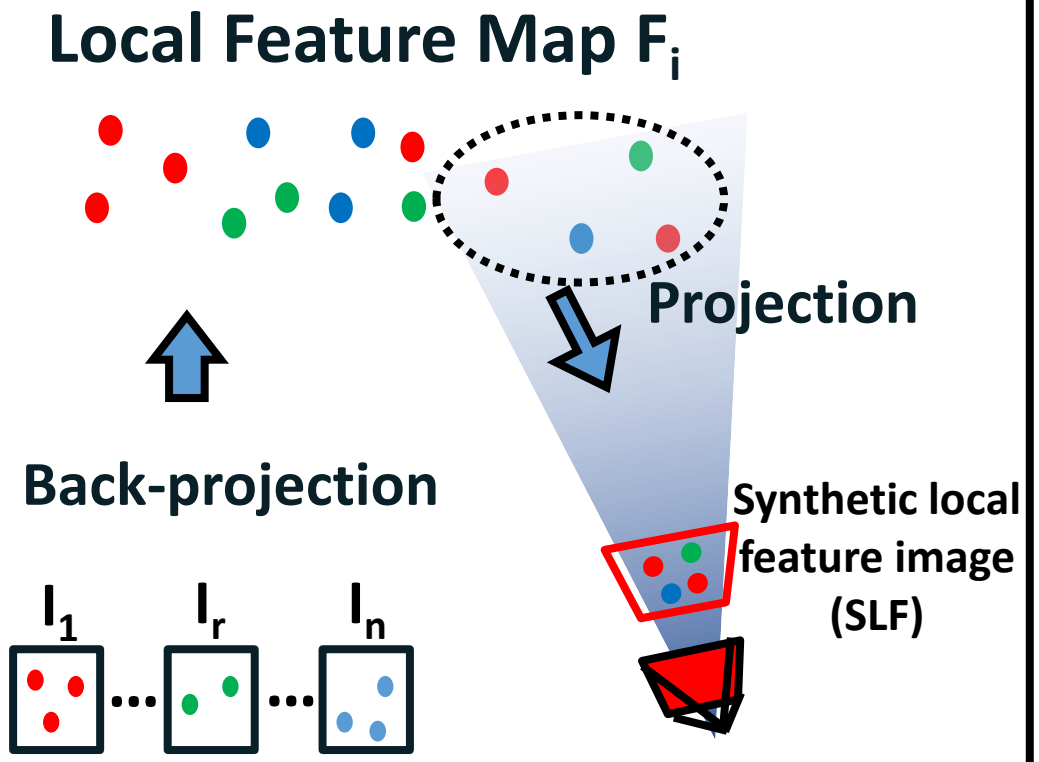
Pose Correction

Properties for Pose Correction

Proximity

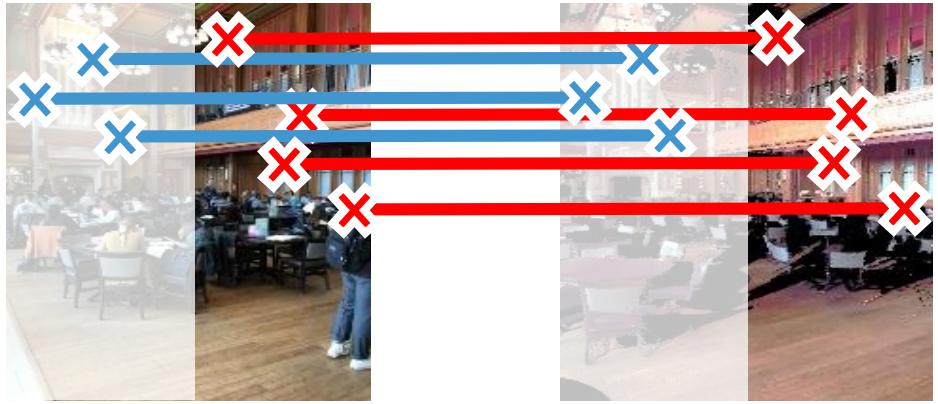


Abundance of Features



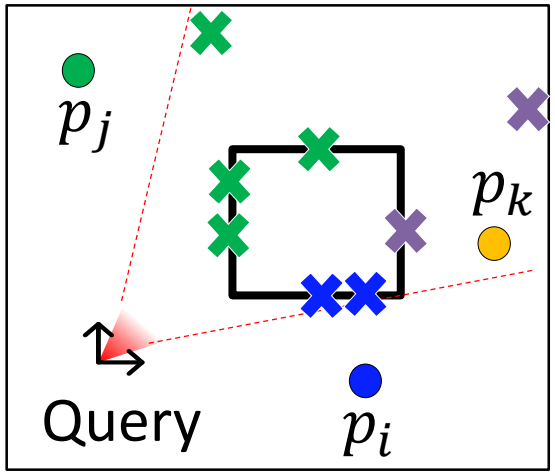
Extended Pose Correction

Divided matching



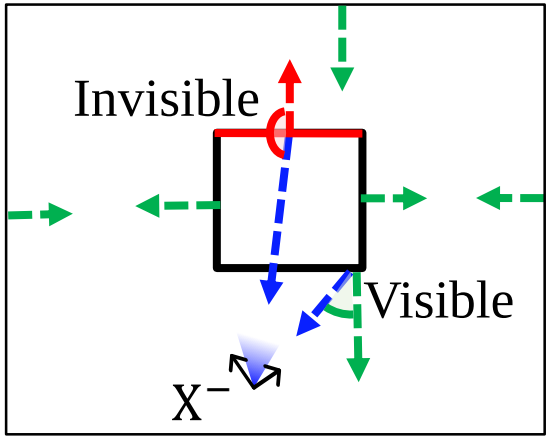
Query Image SLF Image

Inter-pose matching



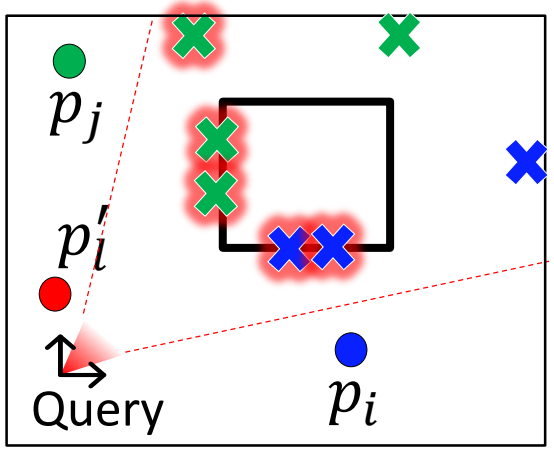
Feature matching using more features (features from $p_i, p_j,$ and p_k)

Filtering Process



Remove invisible (occluded) features using normal distances

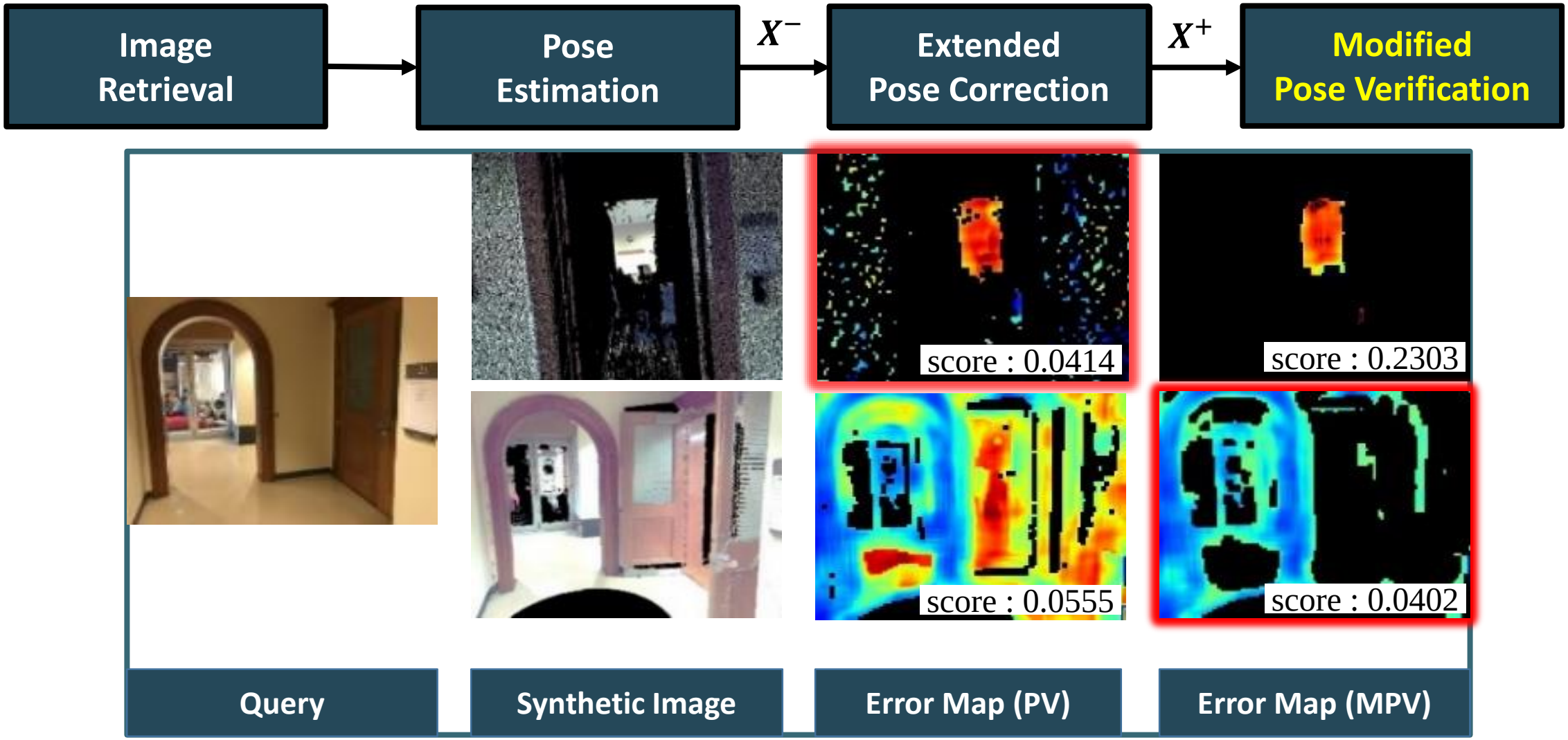
Point normal filtering



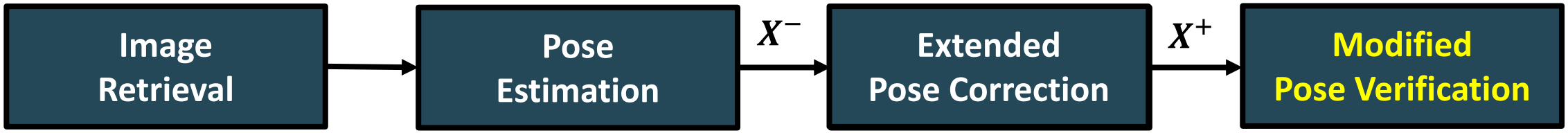
- Filtering visible features at virtual positions.
- Pose Correction using the closest virtual position.

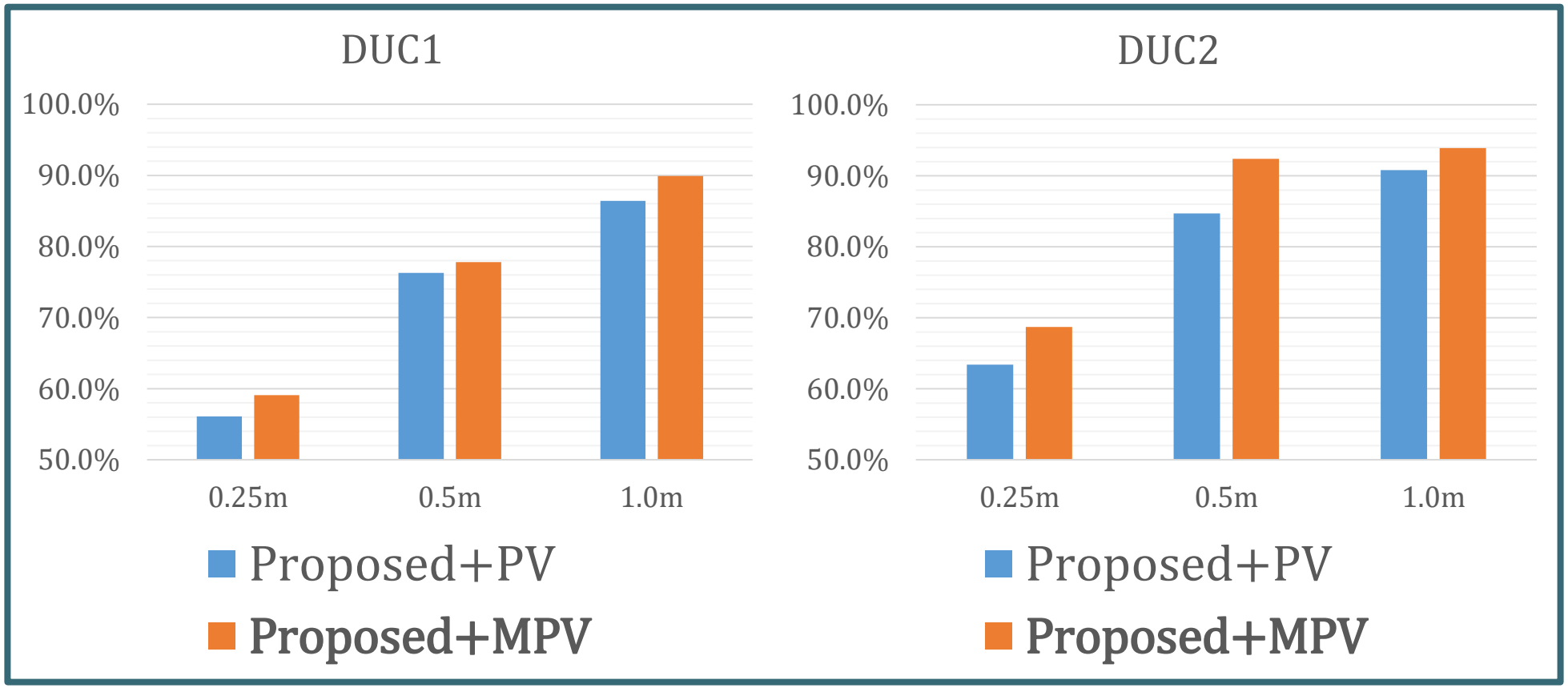
Virtual Local Feature Map

Modified Pose Verification



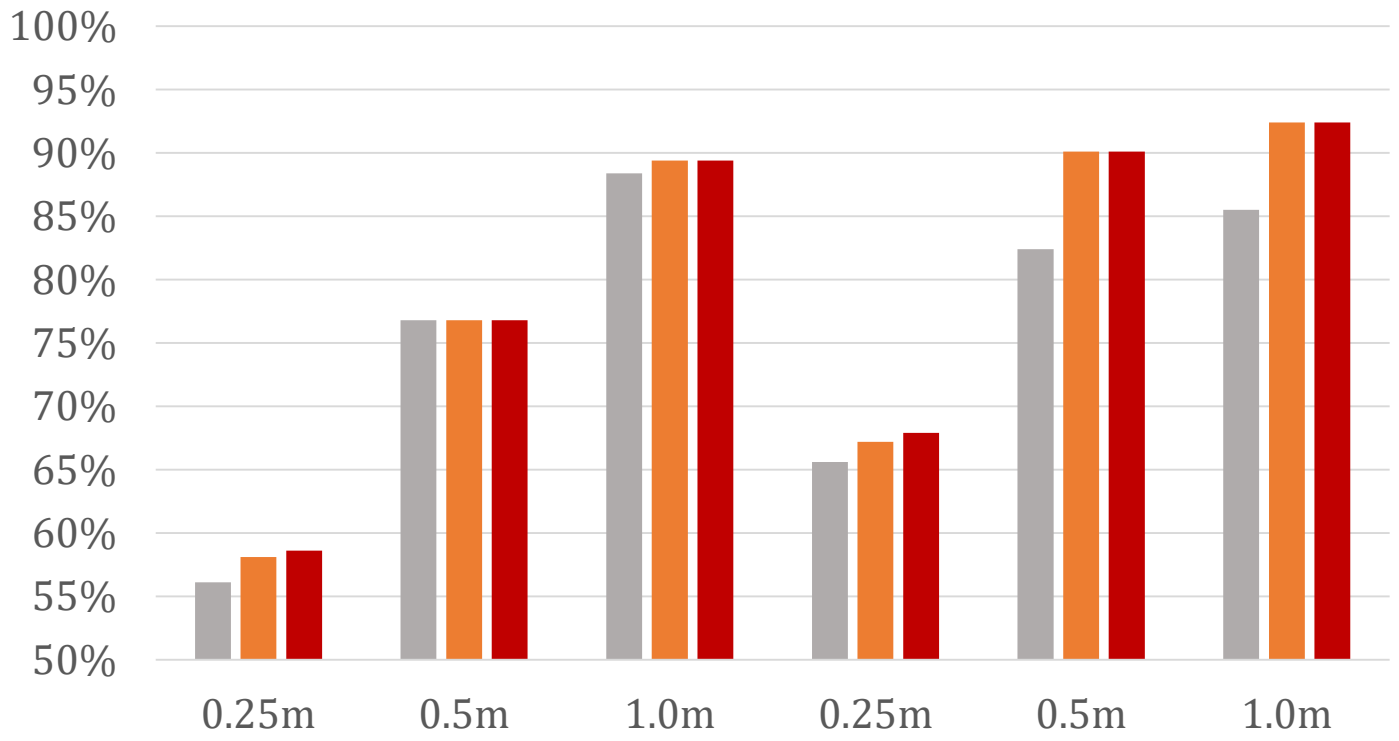
Modified Pose Verification





Experimental Results

- **Pose Correction = Pose update + Reranking**
  - **(a) Baseline**
  - **(b) Pose update (w/o Reranking)**
  - **(c) Pose update with Reranking**

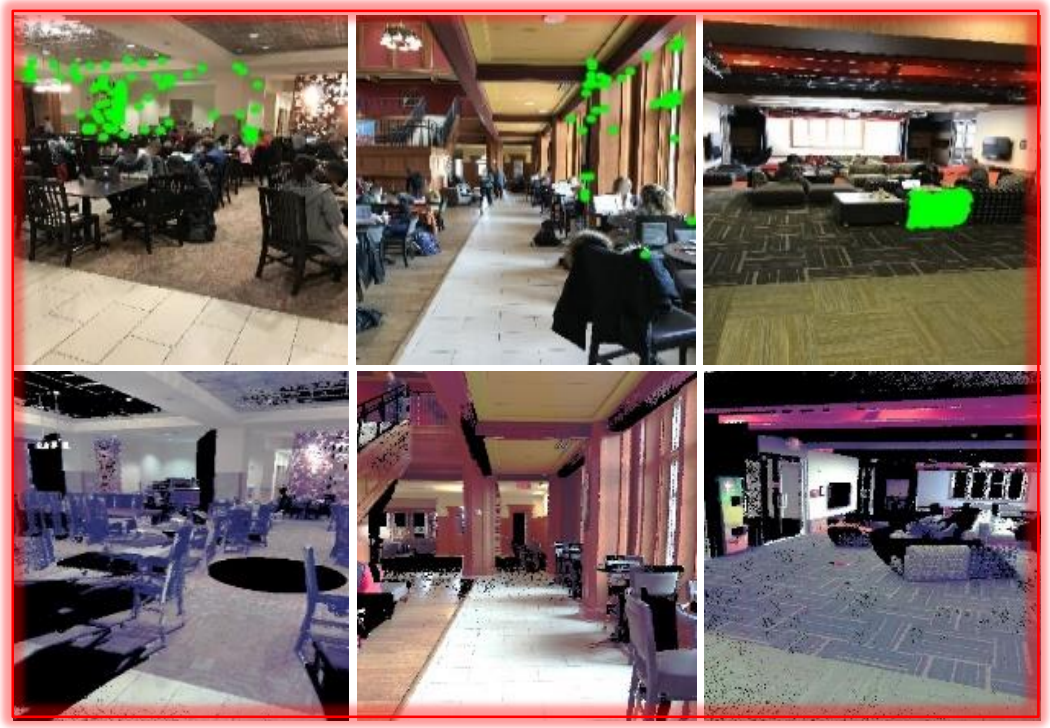


Qualitative Comparison

- **Baseline**

Query

Rendered view



- **Repetitive patterns, dynamic objects**
- **Transitional and rotational errors**

- **Pose Correction**

- **Better matching inliers**
- **Lesser transitional and rotational errors**

Experimental Results

- **Ablation studies**

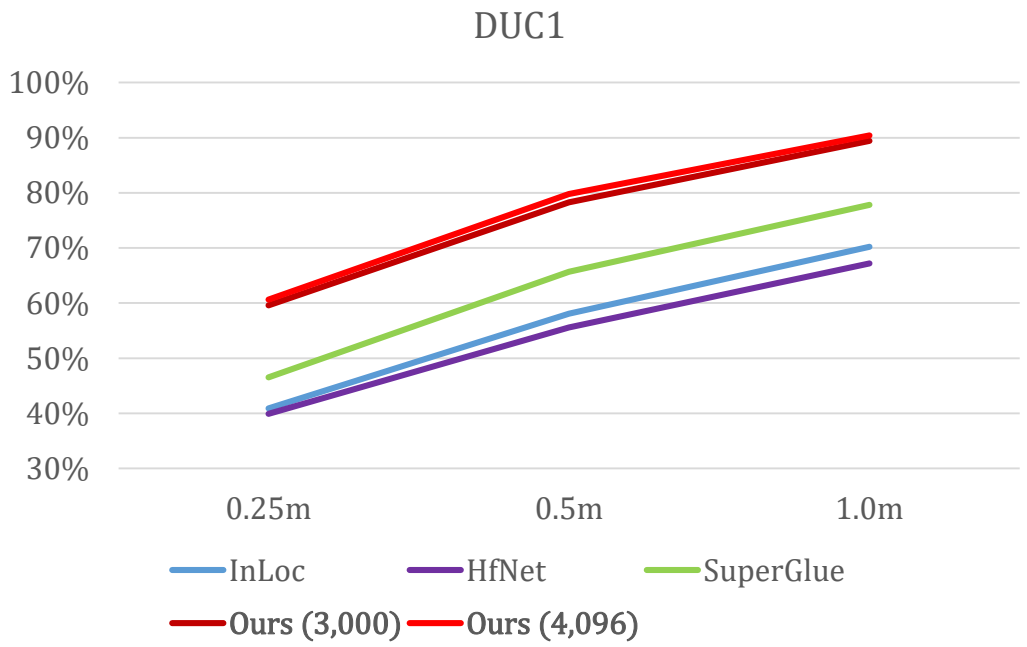
	idx	Error [10°]	DUC1			DUC2		
			0.25m	0.5m	1.0m	0.25m	0.5m	1.0m
w/o inter-pose	(a-1)	F	58.6	76.8	89.4	67.9	90.1	92.4
	(a-2)	Div	60.1	75.8	89.4	69.5	91.6	92.4
	(a-3)	Div-N	59.6	80.8	89.4	67.2	90.8	91.6
	(a-4)	Div-SG	59.6	77.8	88.9	66.4	90.8	91.6
	(a-5)	Div-N-SG	59.6	80.8	89.4	67.2	90.8	91.6
w/ inter-pose	(b-1)	F	57.1	79.8	88.9	66.4	87.8	91.6
	(b-2)	Div	59.6	80.3	89.9	71.0	90.1	90.8
	(b-3)	Div-N	59.1	79.3	89.9	71.0	91.6	91.6
	(b-4)	Div-SG	59.6	79.8	88.9	69.5	90.1	90.1
	(b-5)	Div-N-SG	60.6	79.3	89.4	70.2	90.1	90.1
w/ VLF map	(c-1)	F	58.1	78.3	90.4	69.5	89.3	92.4
	(c-2)	Div	60.1	79.3	90.9	68.7	91.6	92.4
	(c-3)	Div-N	59.1	77.8	89.9	68.7	92.4	93.9
	(c-4)	Div-SG	60.6	77.8	89.9	70.2	92.4	93.9
	(c-5)	Div-N-SG	59.6	78.3	89.4	71.0	93.1	93.9

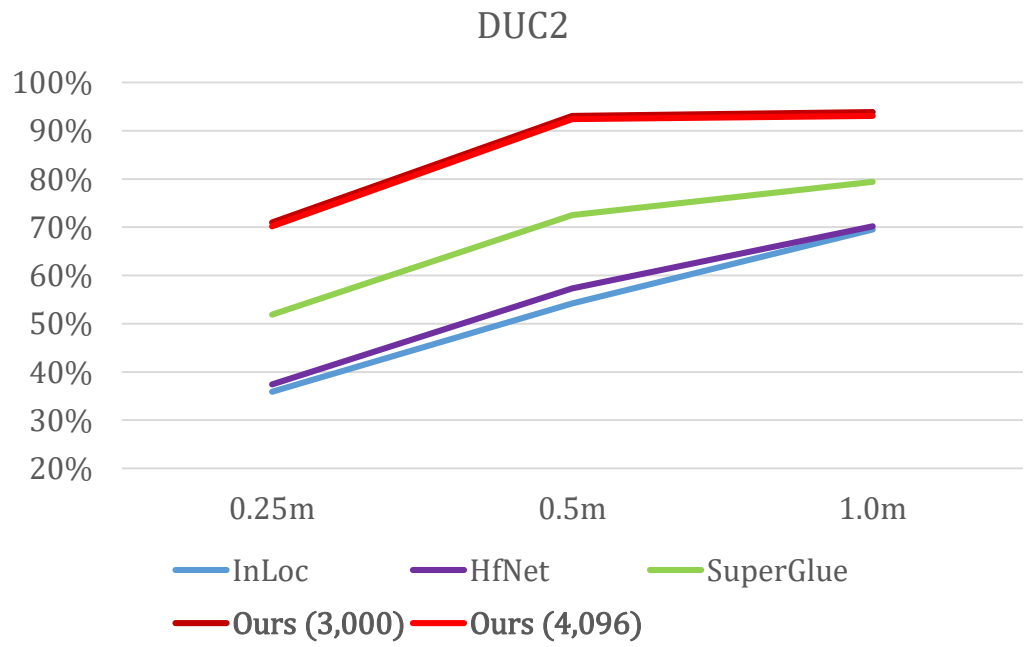
Error [m, 10°]	DUC1			DUC2		
	0.25	0.5	1.0	0.25	0.5	1.0
(a) Baseline +PV	53.0	76.8	85.9	61.8	80.9	87.0
(b) Baseline +MPV	56.1	76.8	88.4	65.6	82.4	85.5
(c) Proposed +PV	56.1	76.3	86.4	63.4	84.7	90.8
(d) Proposed +MPV	59.1	77.8	89.9	68.7	92.4	93.9
(e) Baseline +SGPV	56.1	73.7	83.8	58.0	77.1	83.2
(f) Baseline +SGMPV	57.1	74.7	87.4	63.4	79.4	84.0
(g) Proposed +SGPV	59.1	77.8	89.9	68.7	92.4	93.9
(h) Proposed +SGMPV	59.6	78.3	89.4	71.0	93.1	93.9

Experimental Results

- **State-of-the-art performance**

Error (10°)	DUC1			DUC2		
	0.25m	0.5m	1.0m	0.25m	0.5m	1.0m
InLoc	40.9	58.1	70.2	35.9	54.2	69.5
HfNet	39.9	55.6	67.2	37.4	57.3	70.2
SuperGlue	46.5	65.7	77.8	51.9	72.5	79.4
Ours (3,000)	59.6	78.3	89.4	71.0	93.1	93.9
Ours (4,096)	60.6	79.8	90.4	70.2	92.4	93.1





2021 ICCV OCTOBER 11-17 VIRTUAL

Thanks for listening!

Project page: https://github.com/JanghunHyeon/PCLoc