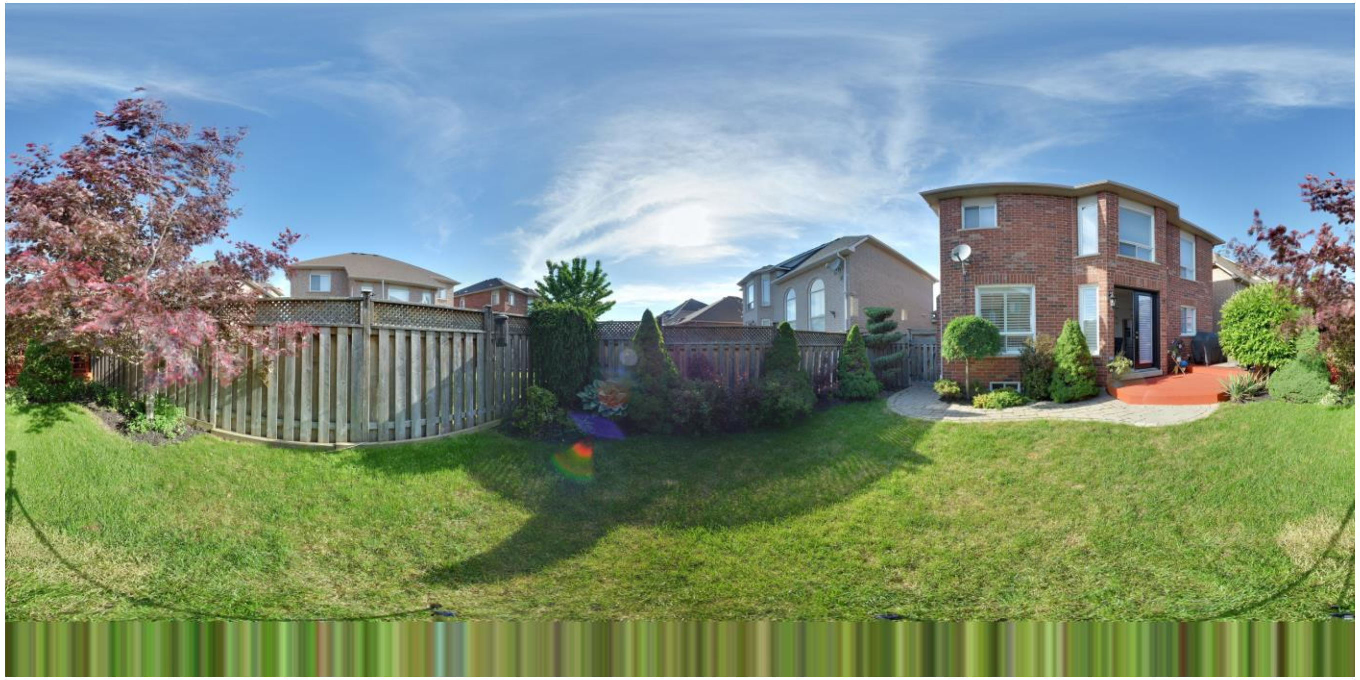
ROBUST POSE ESTIMATION FOR SPHERICAL PANORAMA

Presenter: Shuai Wang

Advisor: Dr. Chen Wang

University at Buffalo The State University of New York

What's a spherical panoramic image?

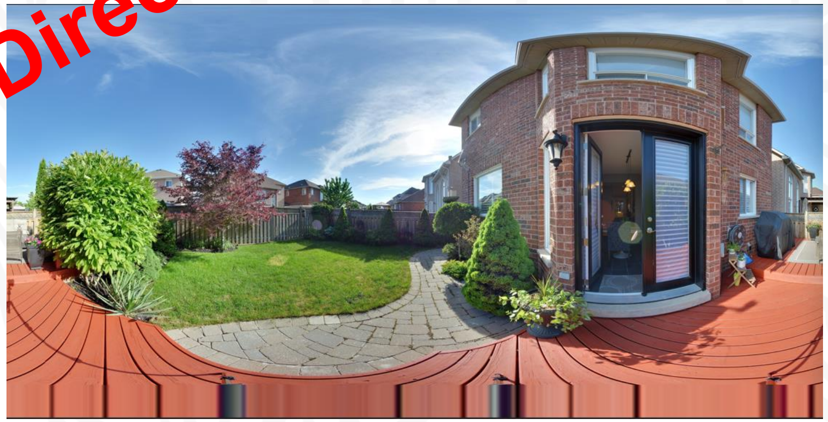


What's virtual tour?

Note: The video comes from the Google earth.

What's virtual tour?

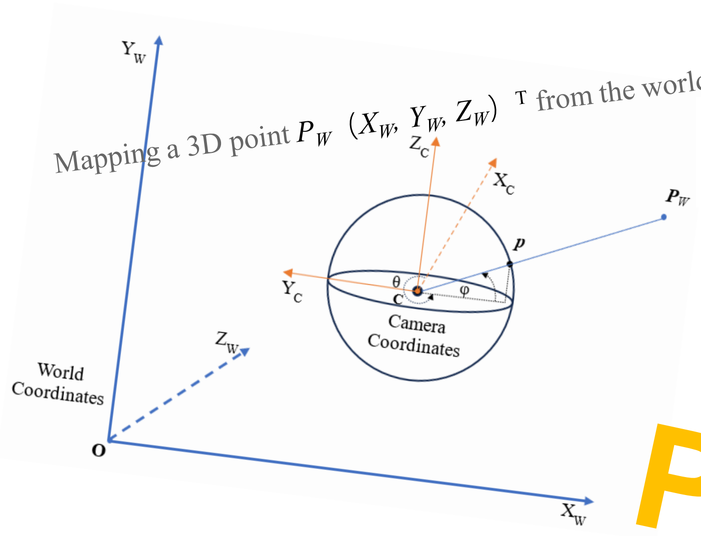
What’s virtual tour?



Direction of the motion!

Pose Estimation

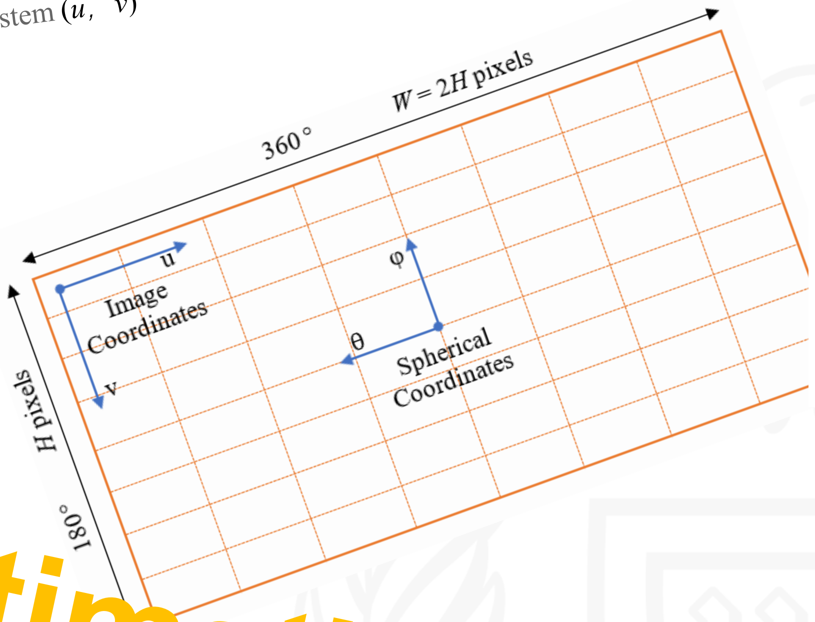
Mapping a 3D point $P_W\ (X_W, Y_W, Z_W)^T$ from the world coordinates to the 2D image coordinate system $(u, v)^T$



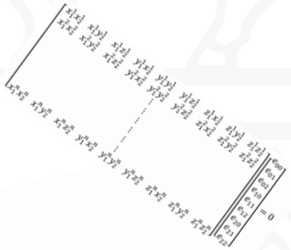
$$P_C = R\,P_W + t$$

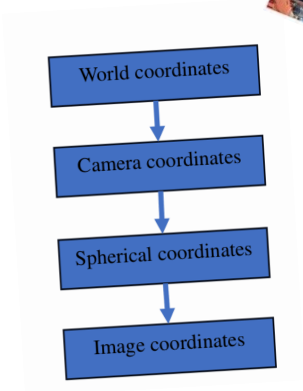
$$p = \frac{P_C}{\|P_C\|} = \frac{R\,P_W + t}{\|R\,P_W + t\|}$$

$$\begin{cases} \varphi = \sin^{-1} z \\ \theta = \tan^{-1} \frac{y}{x} \end{cases}$$

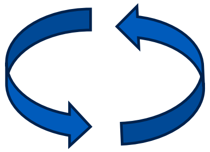


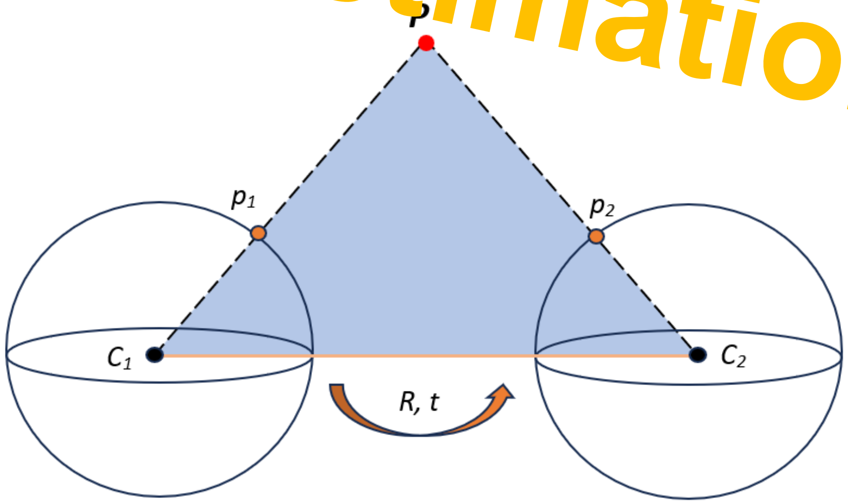
$$\begin{cases} u = \dfrac{W\,(\pi - \theta)}{2\pi} \\ v = \dfrac{H\,(\pi - 2\varphi)}{2\pi} \end{cases}$$

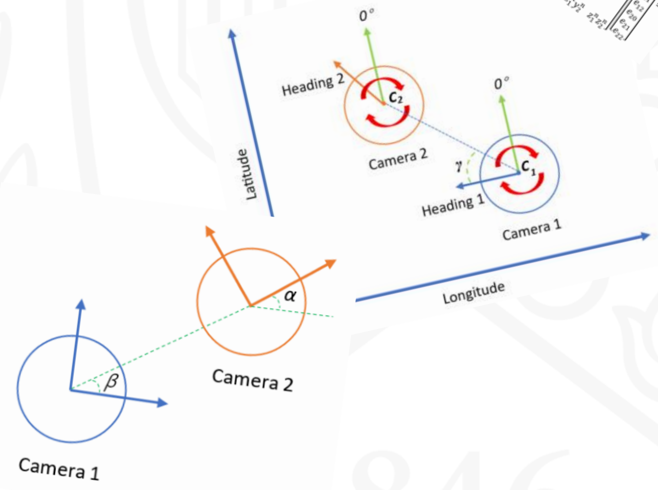




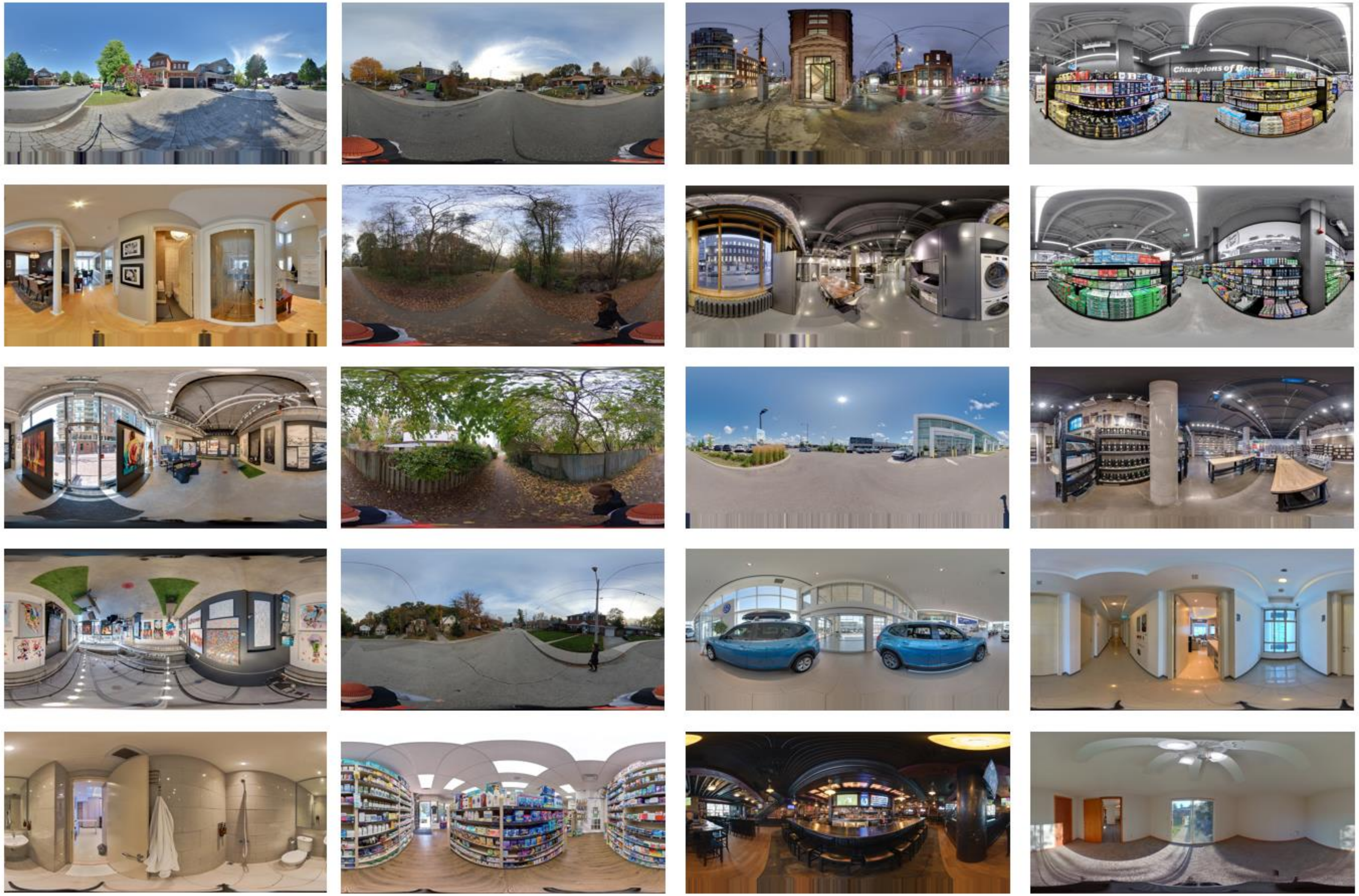
$$p_1^T E\, p_2 = 0$$







The environments in virtual tour



Diverse environments

Large baseline

Featureless regions

Visual ambiguity

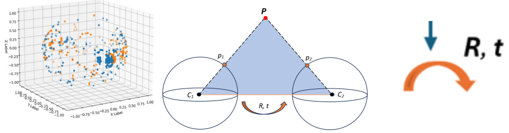
…

Challenges we faced

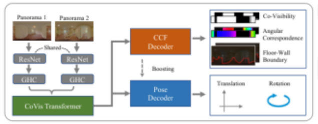
Existing methods

Manual Annotation

Classic feature based



Learning based



Diverse environments

Large baseline

Featureless regions

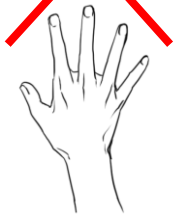
Visual ambiguity

…

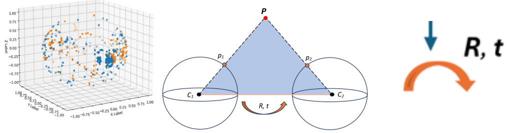
Challenges we faced

Existing methods

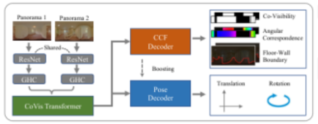
~~Manual Annotation~~



Classic feature based



Learning based



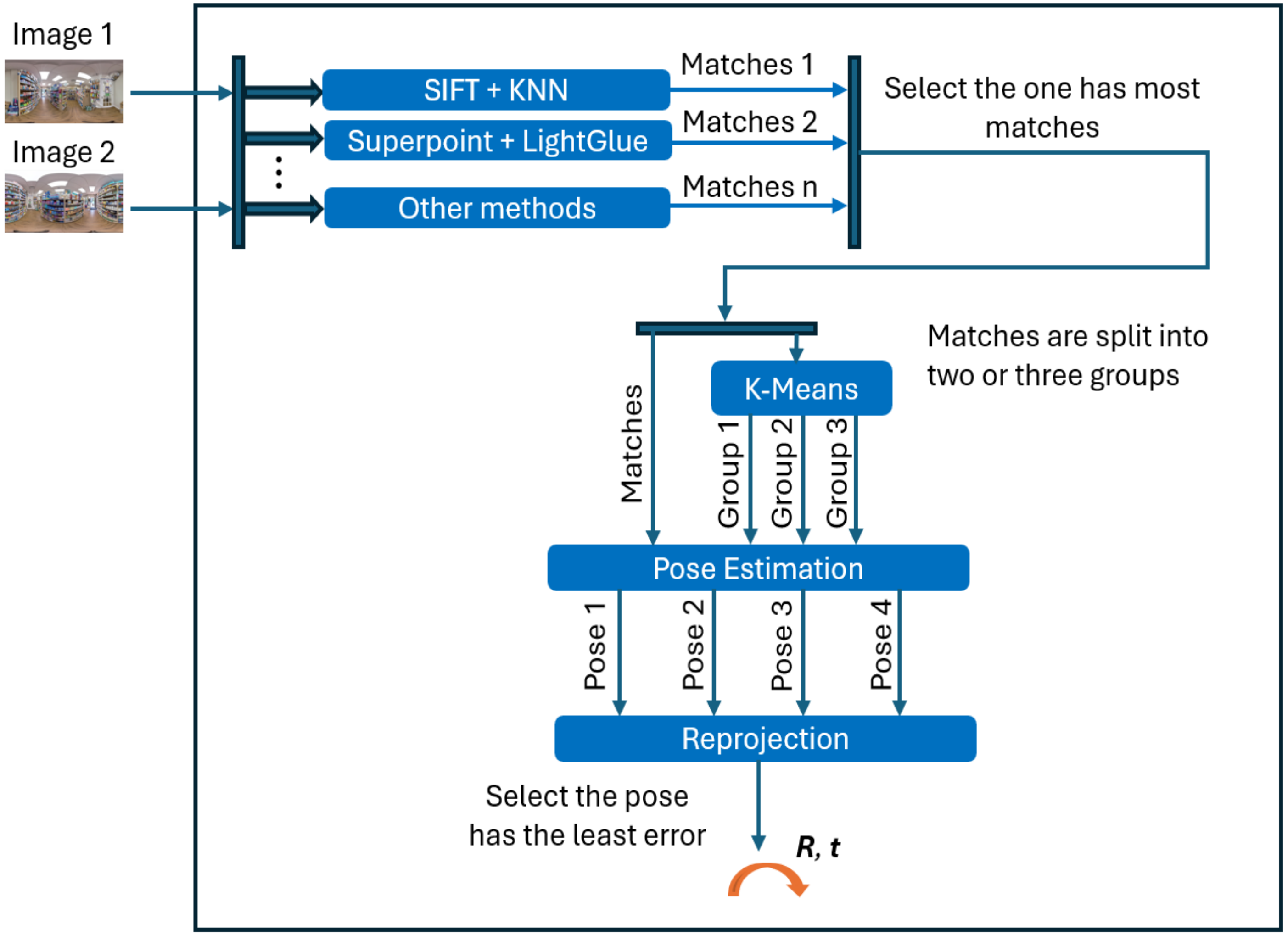
The Research of pose estimation on perspective images VS spherical images

Perspective images

VS

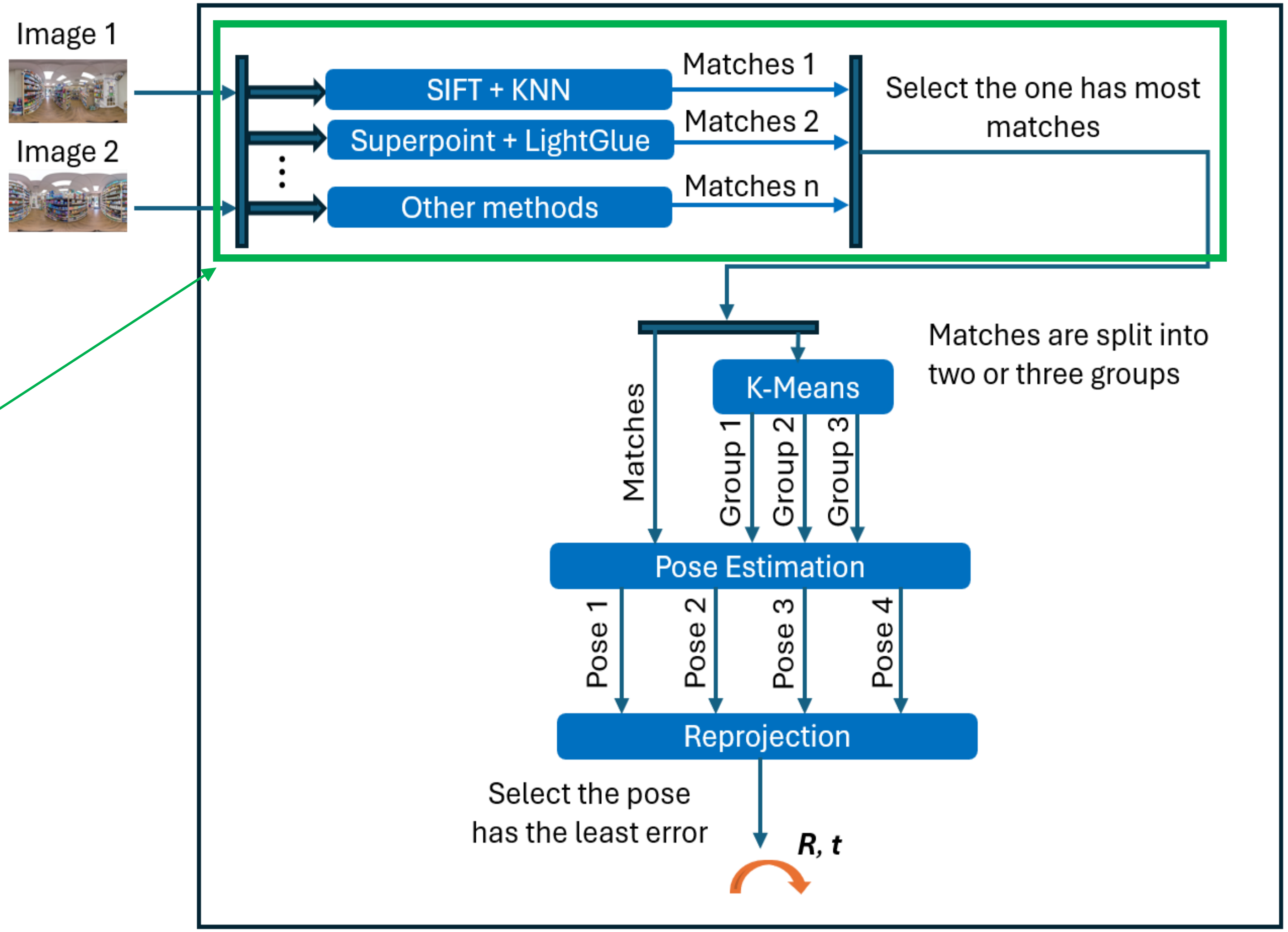
Spherical images

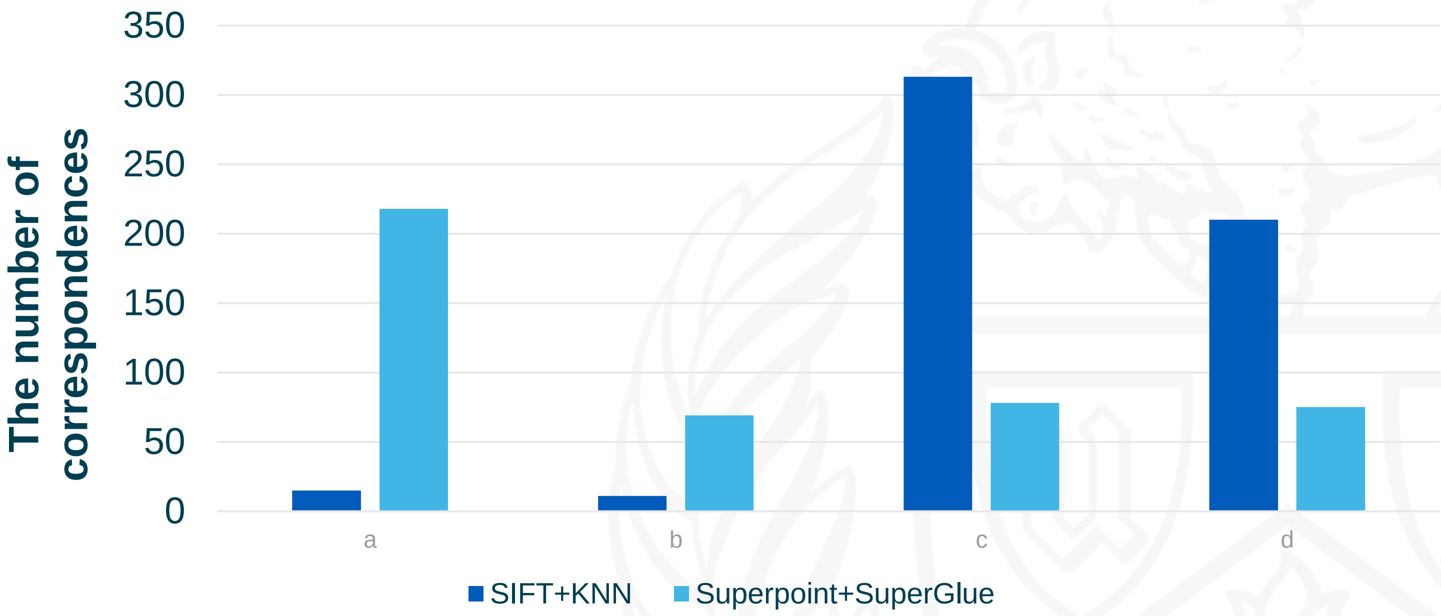
Proposed Methodology



Keypoints detection and matching

Multi-method





The highest number of the correspondences usually get the good result.

Test results – compare single methods

Accuracy (%) for different acceptance thresholds in various indoor and outdoor environments

Method	5° ↑		10° ↑		15° ↑		20° ↑		25° ↑	
	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor
SIFT+KNN	39.19	22.96	90.54	39.58	93.24	51.45	94.59	58.31	94.59	62.53
SuperPoint+ LightGlue	41.90	27.18	70.27	48.55	77.02	64.38	79.73	73.88	79.73	80.21

Test results – compare single methods

Accuracy (%) for different acceptance thresholds in various indoor and outdoor environments

Method	5° ↑		10° ↑		15° ↑		20° ↑		25° ↑	
	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor	indoor	outdoor
SIFT+KNN	39.19	22.96	90.54	39.58	93.24	51.45	94.59	58.31	94.59	62.53
SuperPoint+ LightGlue	41.90	27.18	70.27	48.55	77.02	64.38	79.73	73.88	79.73	80.21

Test results – compare single method and combine method

Accuracy (%) for different acceptance thresholds in various environments

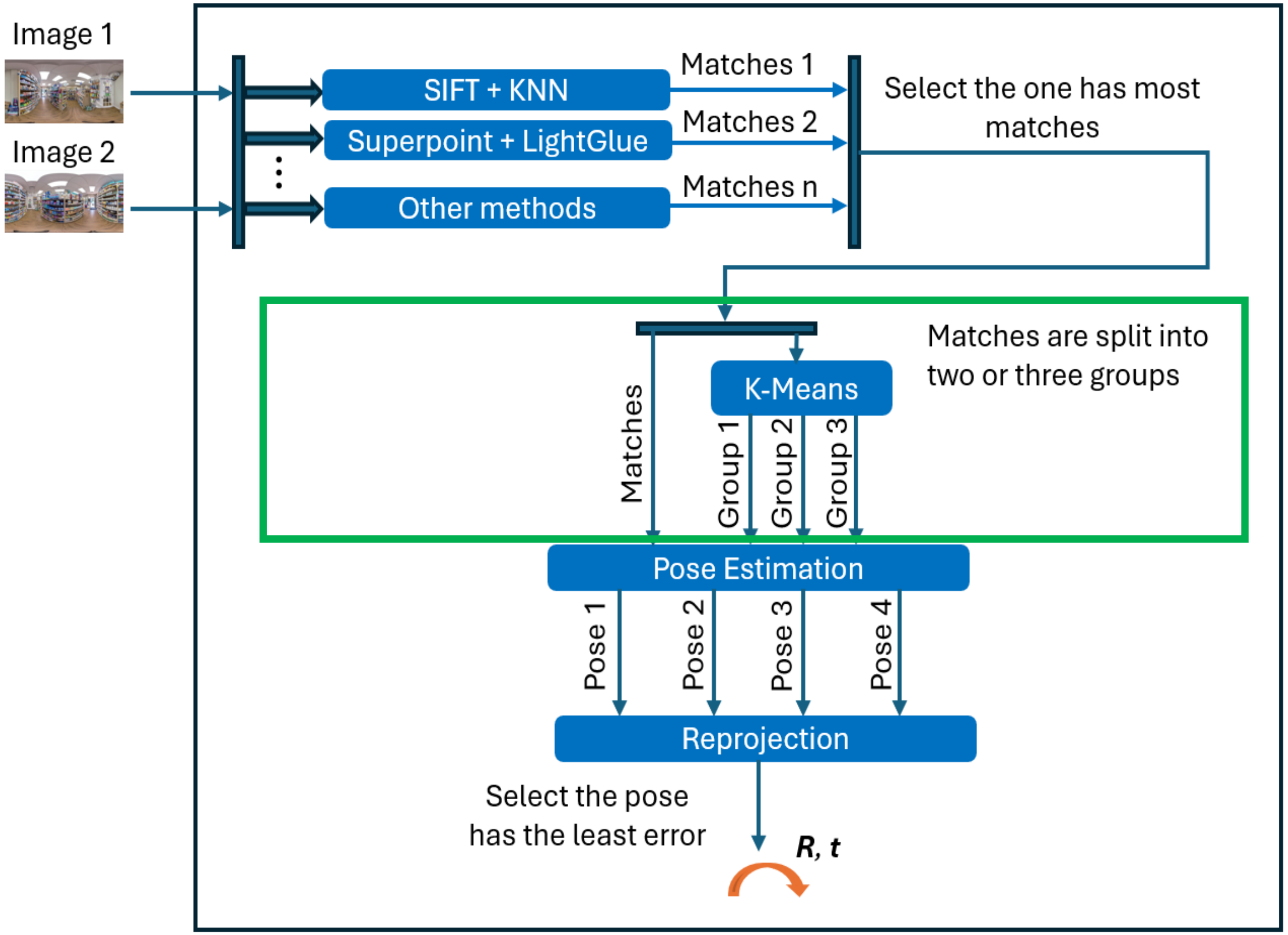
Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
SIFT + KNN	25.61	47.90	58.28	64.24	67.77
SuperPoint + LightGlue	29.58	52.10	66.45	74.83	80.13
Combine	30.68	55.63	69.09	78.15	84.77

Test results – compare single method and combine method

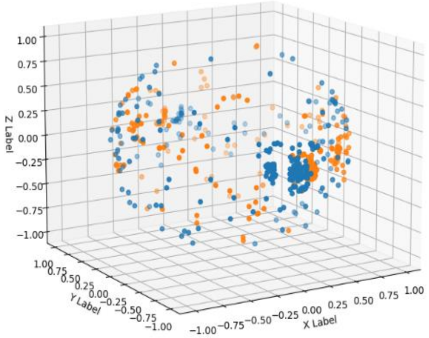
Accuracy (%) for different acceptance thresholds in various environments

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
SIFT + KNN	25.61	47.90	58.28	64.24	67.77
SuperPoint + LightGlue	29.58	52.10	66.45	74.83	80.13
Combine	30.68	55.63	69.09	78.15	84.77

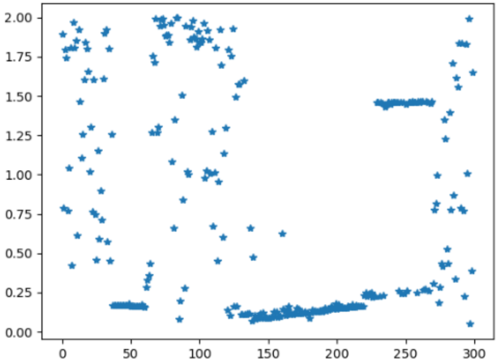
Preprocessing -reduce the outlier's rate



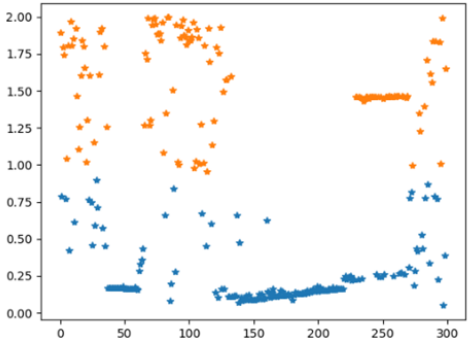
Correspondences



Spherical coordinates



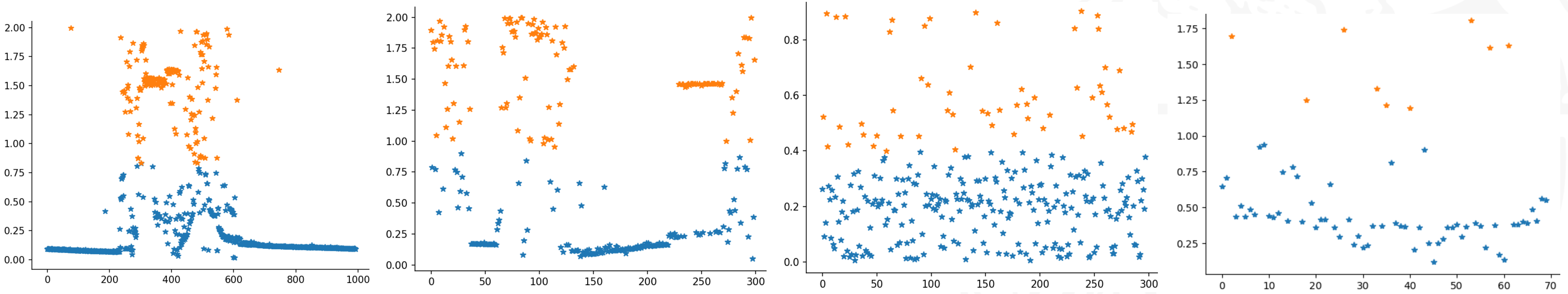
Distances



Clustering

The process of split the correspondence into groups

Some Split Results



Test results - Accuracy (%) for different acceptance thresholds in various environment

Indoor

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
Without preprocessing	43.24	89.19	91.90	91.89	93.24
After preprocessing	43.24	94.59	95.95	95.95	95.95

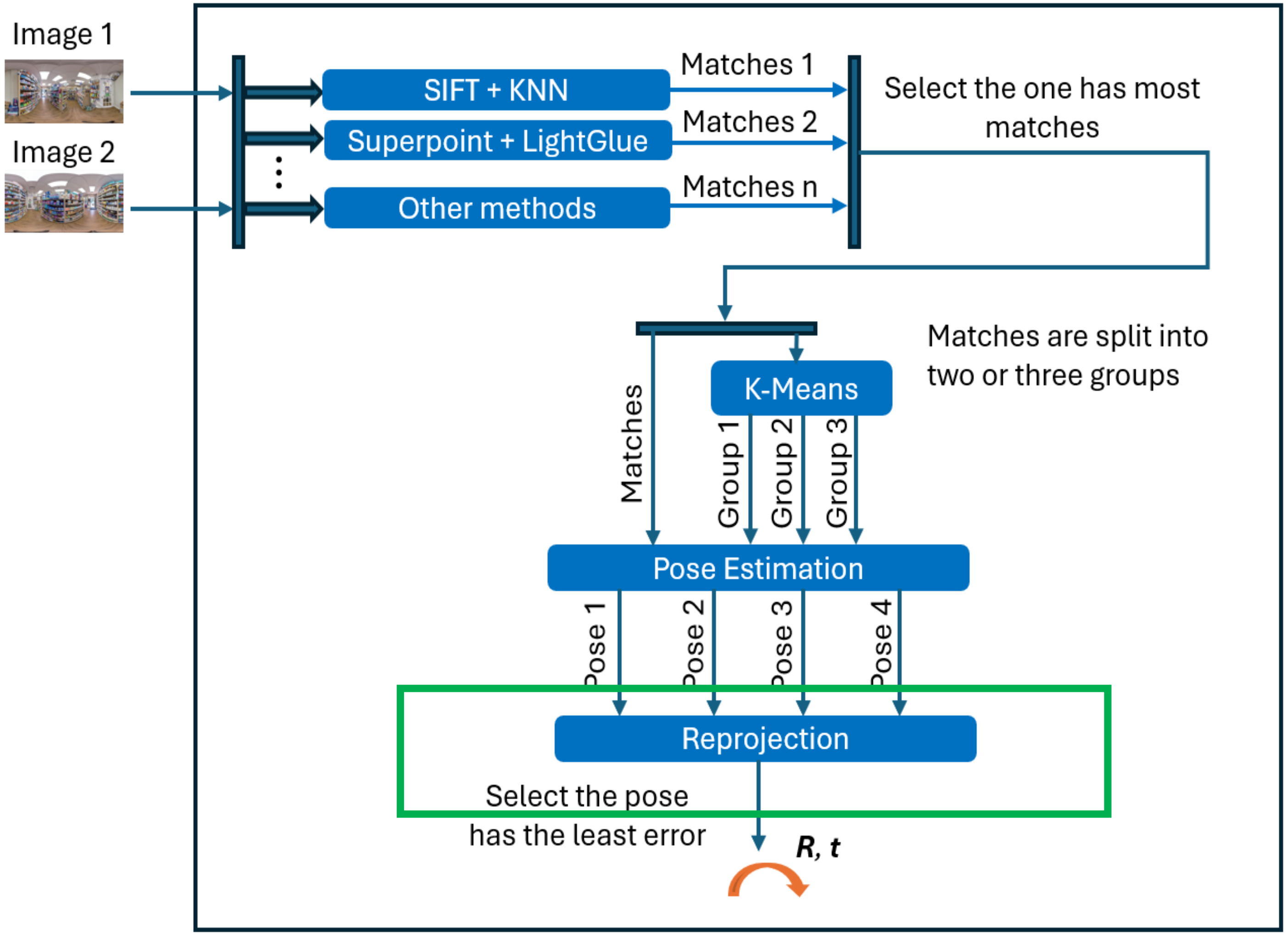
Outdoor

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
Without preprocessing	28.23	49.08	64.64	75.46	83.11
After preprocessing	30.61	49.34	67.28	78.89	84.96

Overall

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
Without preprocessing	30.68	55.63	69.09	78.15	84.77
After preprocessing	**56.73**	**56.73**	**71.96**	**81.68**	**86.76**

Different error function for RANSAC



New Error function for RANSAC

$$[\boldsymbol{p}_1 \quad -\boldsymbol{R}\boldsymbol{p}_2]\begin{bmatrix}\|\boldsymbol{P}_1\| \\ \|\boldsymbol{P}_2\|\end{bmatrix} = t \qquad (4-20)$$

The error function computes the reproject error in camera coordinates according to equation (4-18).

$$\boldsymbol{error} = \boldsymbol{p}_1 * \|\boldsymbol{P}_1\| - (\boldsymbol{R}\boldsymbol{p}_2 * \|\boldsymbol{P}_1\| + \boldsymbol{t}) \qquad (4-23)$$

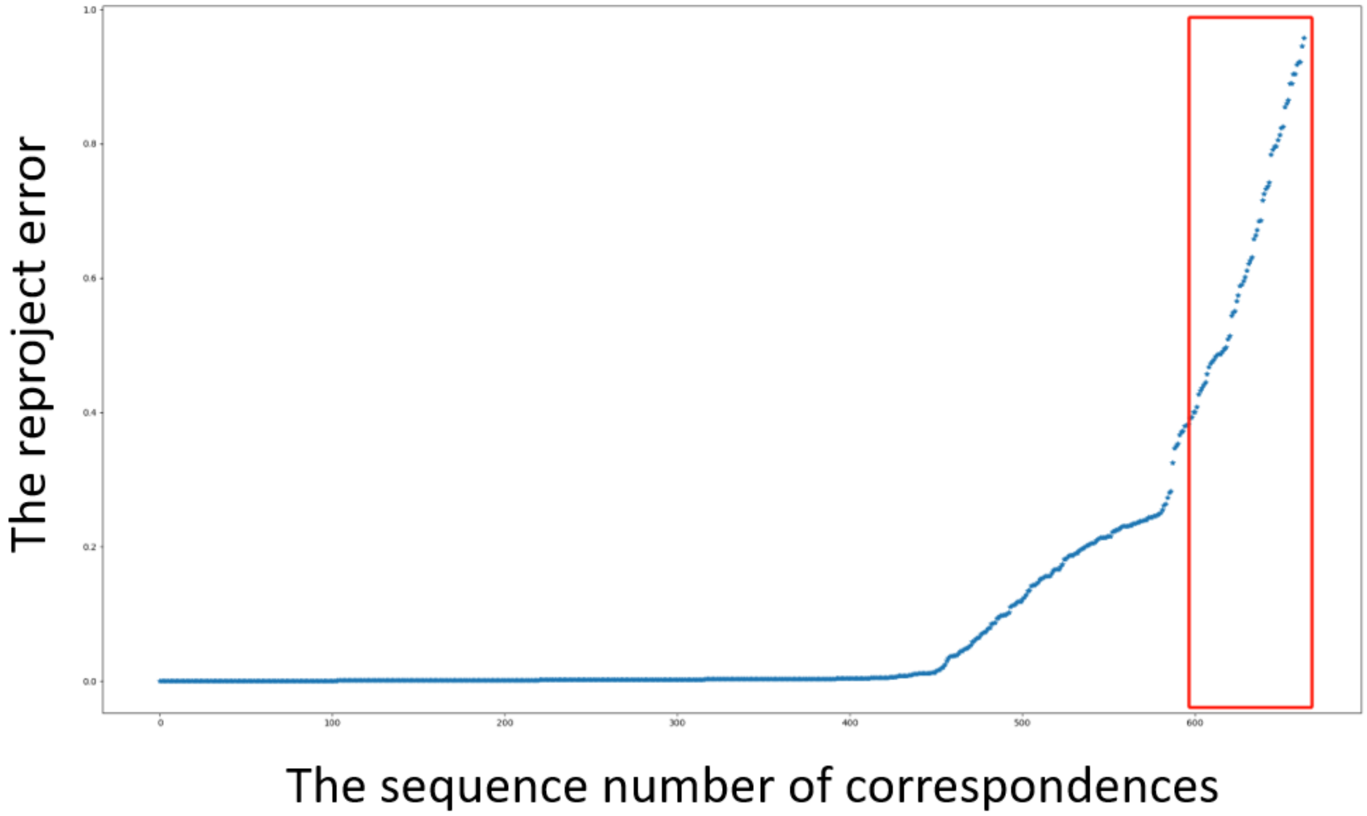
For every **R** and **t**, estimate its mean error for all the correspondences.

$$\textbf{Mean error} = \frac{1}{n}\sum_{i=0}^{n}(\boldsymbol{p}_1 * \|\boldsymbol{P}_1\|) - (\boldsymbol{R}\boldsymbol{p}_2 * \|\boldsymbol{P}_2\| + \mathbf{t}) \qquad (4-24)$$

The pose that has the least mean error is selected.

Argmin (Mean errors)

A small trick



Overall test results and conclusion

Indoor

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
SIFT + KNN	39.19	90.54	93.24	94.59	94.59
SuperPoint + LightGlue	41.90	70.27	77.02	79.73	79.73
Without preprocessing	43.24	89.19	91.90	91.89	93.24
Ours	43.24	94.59	95.95	95.95	95.95

Outdoor

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
SIFT + KNN	22.96	39.58	51.45	58.31	62.53
SuperPoint + LightGlue	27.18	48.55	64.38	73.88	80.21
Without preprocessing	28.23	49.08	64.64	75.46	83.11
Ours	30.61	49.34	67.28	78.89	84.96

Overall

Method	5° ↑	10° ↑	15° ↑	20° ↑	25° ↑
SIFT + KNN	25.61	47.90	58.28	64.24	67.77
SuperPoint + LightGlue	29.58	52.10	66.45	74.83	80.13
without preprocessing	30.68	55.63	69.09	78.15	84.77
Our	**56.73**	**56.73**	**71.96**	**81.68**	**86.76**

Acknowledgement

Thanks to my advisor, Dr. Chen Wang !

Thanks to all my colleagues in the SAIR lab!

Thanks to my friends during my time studying in UB!

Thanks to my family!

Please give advice!