Visual Localization using SIFT features
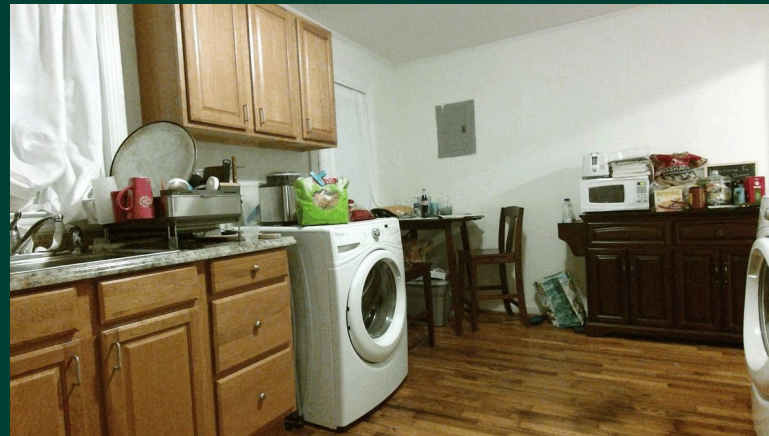
Sulabh Shrestha

Problem Statement

- Figuring out where the agent is in an environment
  - Location
  - Pose
- Using RGB images captured by the agent
- Assuming a database of reference images
  - Known locations/pose

Data

- Active Vision Dataset [1]
  - Indoor environment
  - Locations sampled at 30 cm
  - RGBD taken 30 degrees apart per location

[1] P. Ammirato, P. Poirson, E. Park, J. Košecká and A. C. Berg, "A dataset for developing and benchmarking active vision," ICRA 2017

Approach: Creating feature database

- Reference data
  - Hand picked representative locations
- For each reference node (location)
  - For each image taken here
    - Calculate SIFT keypoints and descriptors
    - Store top N
- Data-structure
  - Features → Image-name
- Query on all other images
- Ground truth
  - Nearest node
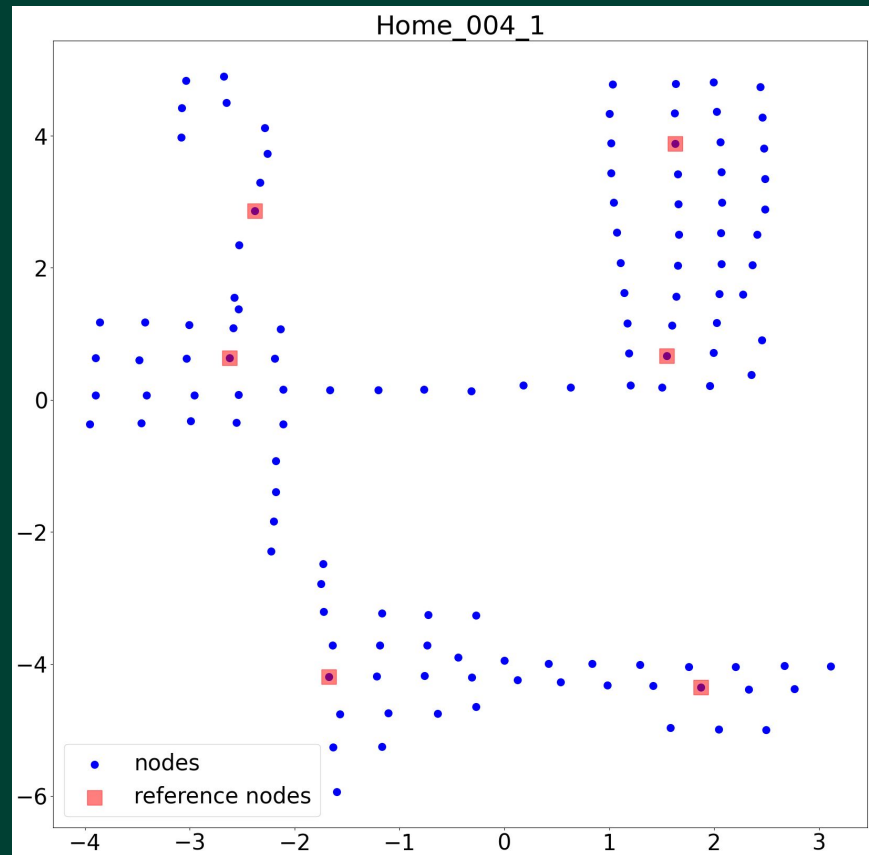
Approach: Creating feature database

Approach: Creating feature database

Approach: Query

- Calculate SIFT keypoints and descriptors
- Match with each keypoint in database
  - Euclidean distance
  - Sort in ascending order
  - Match to closest reference node
    - in feature space

Approach: Enhancements

- SIFT descriptors are **histograms**
  - Euclidean distance is not a good measure
  - Hellinger or χ^2 measure is better
  - Use approximation i.e. Root-SIFT[1]
    - L1 norm each descriptor
    - Square-root each element
    - Finding Euclidean is approximation to χ^2
- Not all matches are correct
  - Filter using max threshold for distance

[1] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval" CVPR 2012
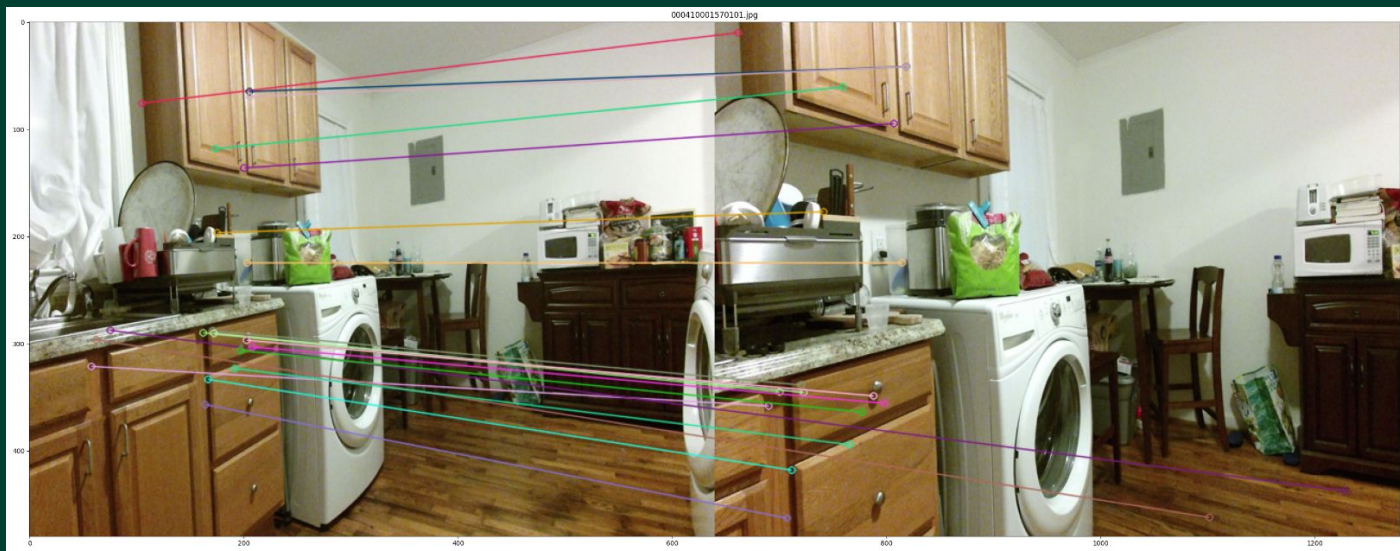
Results

- Inference with cv2.BFMatcher / numpy   = ~50 min
- Inference with pytorch-gpu matching   = ~5 min

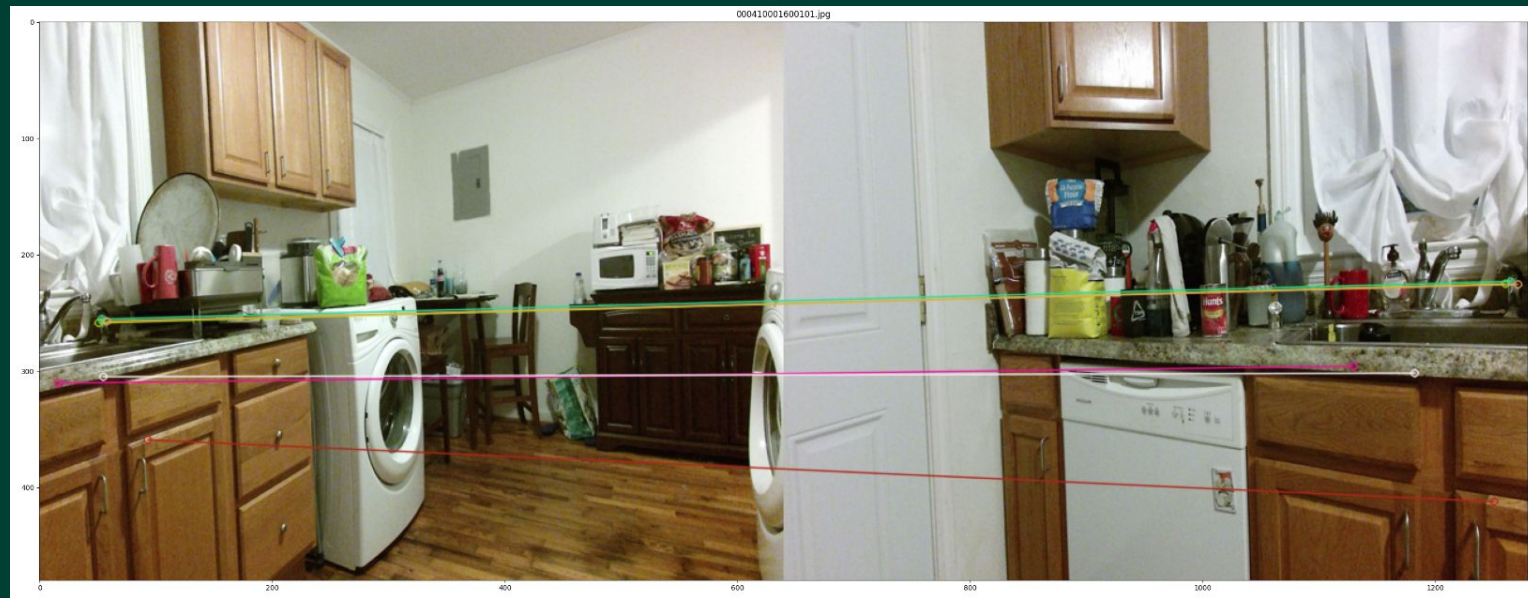| Feature type	Distance Threshold	Top-N Ref Key-points per node	Vocabulary size	Mean Accuracy	Mean Average Precision
SIFT	–	–	~54k	75.71	63.39
ROOT-SIFT	–	–	~54k	77.89	65.94
ROOT-SIFT	0.4	–	~54k	84.49	73.93
ROOT-SIFT	0.3	–	~54k	**85.29**	**75.08**
ROOT-SIFT	0.2	–	~54k	82.11	70.07
ROOT-SIFT	0.3	5000	~29k	80.29	68.94

Results

- Match in similar viewpoints

Results

- Match in extreme viewpoint change

Results

- Bad matches

000410001620101.jpg