Cluster-Wise Ratio Tests for Fast Camera Localization

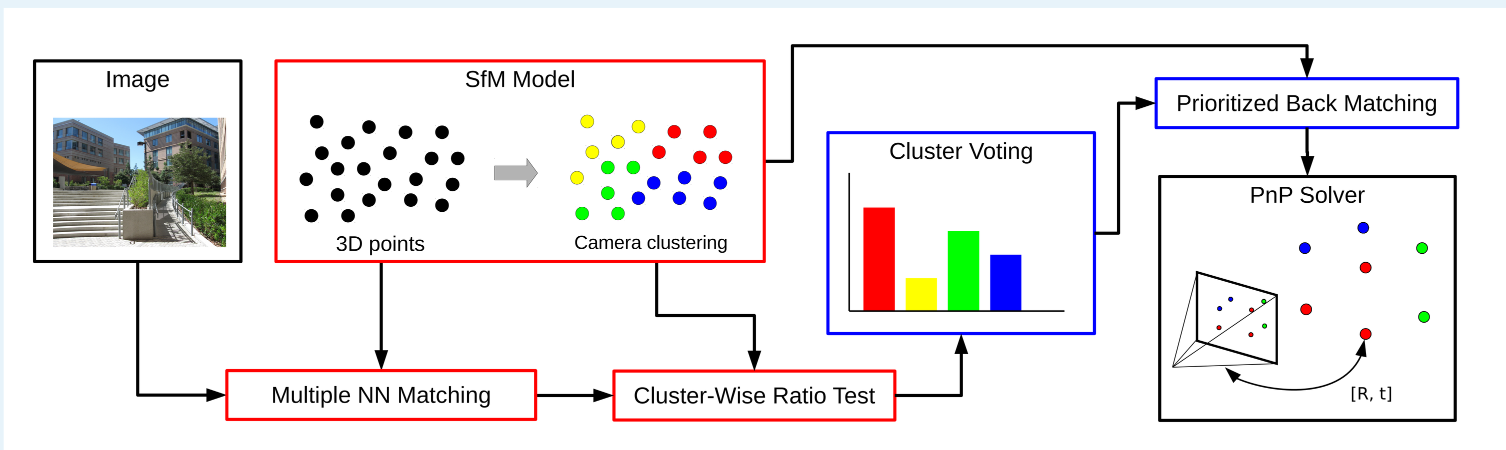
UCIRVINE

Raúl Díaz, Charless C. Fowlkes
School of Information and Computer Sciences
The Henry Samueli School of Engineering

1. Introduction

Feature point matching for camera localization suffers from scalability problems. As coverage grows, similar or repeated features become increasingly common. Hence widely used ratio-test becomes overly restrictive and rejects many good candidate matches.

We propose a simple voting strategy that uses conservative approximations to robust local ratio-tests. We compute them efficiently using approximate global k-nearest neighbor search for each query feature. We treat these forward matches as votes in camera pose space and use them to prioritize back-matching within candidate camera pose clusters, exploiting feature covisibility captured by the 3D model camera pose graph. We achieve excellent results on datasets with multiple global repeated structures.



2. Cluster-Wise Ratio Tests for Global Matching

Global matching of *m* queries and *N* observations can be approached in multiple ways:

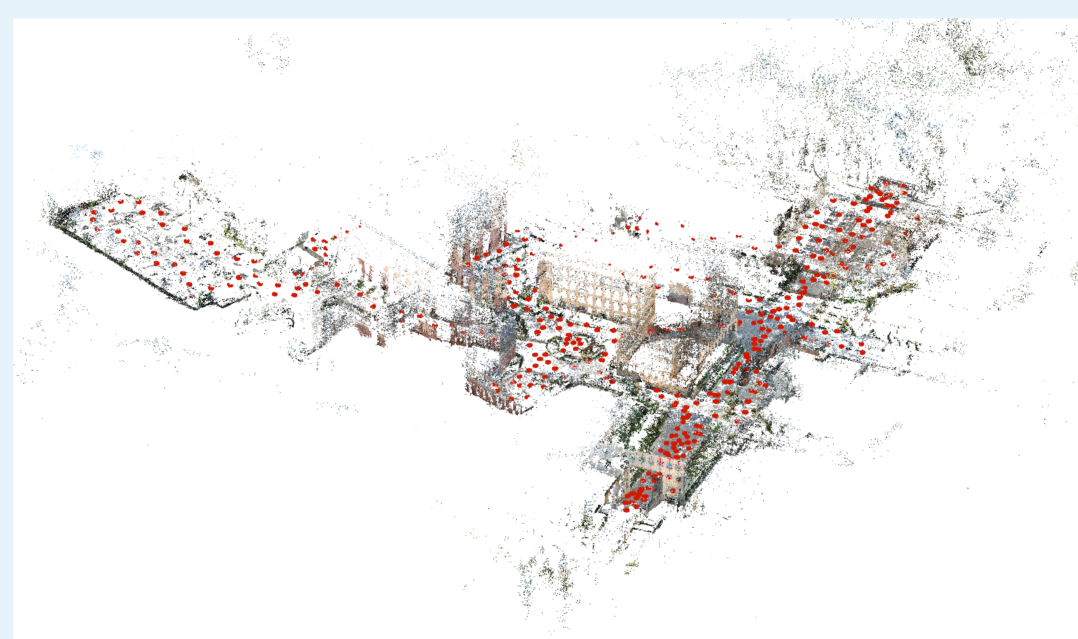
- *Forward match is fast* (*O(m·logN)*), but it performs poorly in large scale models.
- *Back matching* performs better at the expense of longer runtimes of *O(N·logm)*.
- *Exhaustive* forward matching of *c* local pose clusters improves standard forward matching, but runs slower: (*O(cm·log(N/c))*).

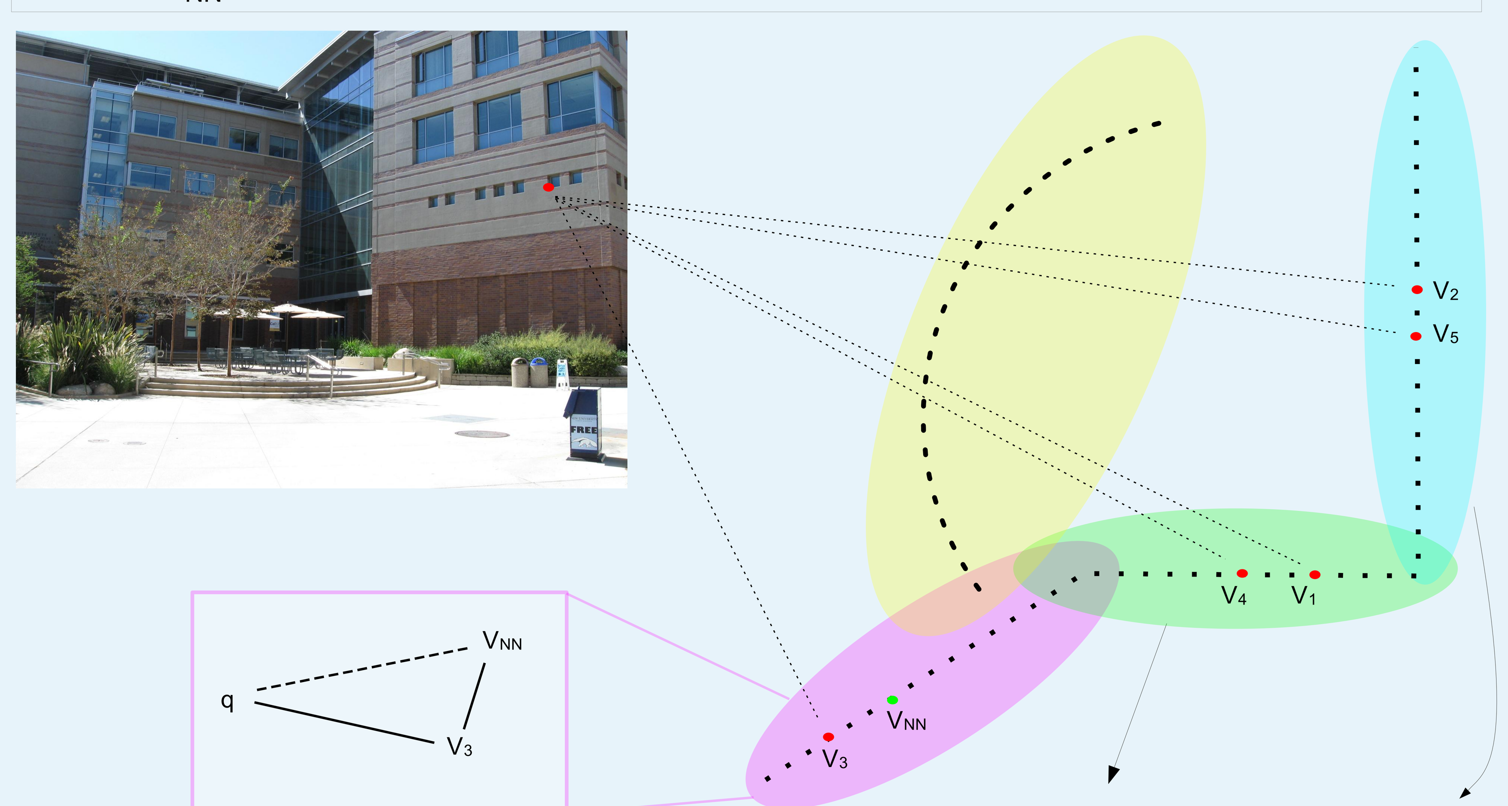
We propose to perform global k-NN matching of query features against all model observations using a soft k-ratio test. We present two approximations of the ratio-test that retrieve local discriminative correspondences that quickly indicate candidate pose clusters of the query image.

Global k-ratio test:
Compare 1st and kth+1 NN and add k correspondences. $\frac{\|q-v_1\|}{\|q-v_{k+1}\|}$

Local 1-ratio test:
Compare 1st and 2nd NN that fall in the same cluster. $\frac{\|q-v_{c1}\|}{\|q-v_{c2}\|}$

Local t-ratio test:
Use triangle inequalites to define an upper bound. $\frac{\|q-v_{c1}\|}{\|q-v_{c1}\|+\|v_{c1}-v_{NN}\|}$
Use v_{NN} as the nearest neighbor of v within the cluster.





$\frac{\|q-v_3\|}{\|q-v_3\|+\|v_3-v_{NN}\|}$ $\frac{\|q-v_1\|}{\|q-v_4\|}$ $\frac{\|q-v_2\|}{\|q-v_5\|}$

Engineering Quad dataset:
- 5,129 training images
- 520 test images with ground truth
- 579,859 3D points
- 2,901,885 model observations

Cluster influence in localization (Eng-Quad):

#clusters	#images	#inliers	Median error [m]	Time [s]
Baseline	463	94	0.64	0.962
50 Exhaustive	512	66	0.45	56.822
50 CW-RT	477	127	0.66	0.915
500 CW-RT	480	133	0.61	0.934
5129 CW-RT	482	136	0.62	0.961

3. Pose Voting and Prioritized Back-Matching

By randomly subsampling N_F query features that pass a global k-ratio test, we quickly achieve high recall in location recognition, using each model image as a camera pose bin.

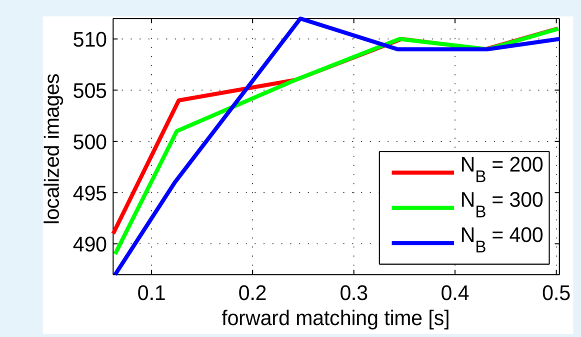
We prioritize back matching of the most voted model images:
- Match most voted model image observations against the query.
- If the model and query images overlap (>11 matches):
  - Accumulate matches for fine pose estimation.
  - Propagate votes to model images sharing the same tracks.
  - Stop if sufficient matches have been retrieved (N_B).

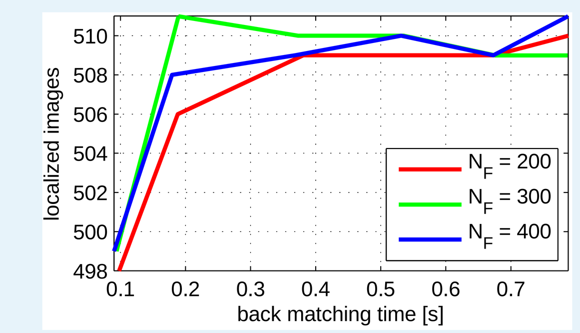
Recognition performance on the Eng-Quad dataset:

	Top-1	Top-5	Top-10	Time [s]
All	86.92%	91.15%	91.92%	0.833
N_F = 500	86.15%	90.96%	91.35%	0.502
N_F = 200	85.77%	90.38%	91.35%	0.242
N_F = 100	84.62%	89.62%	90.96%	0.125
N_F = 50	83.85%	88.46%	88.65%	0.064

Query Image

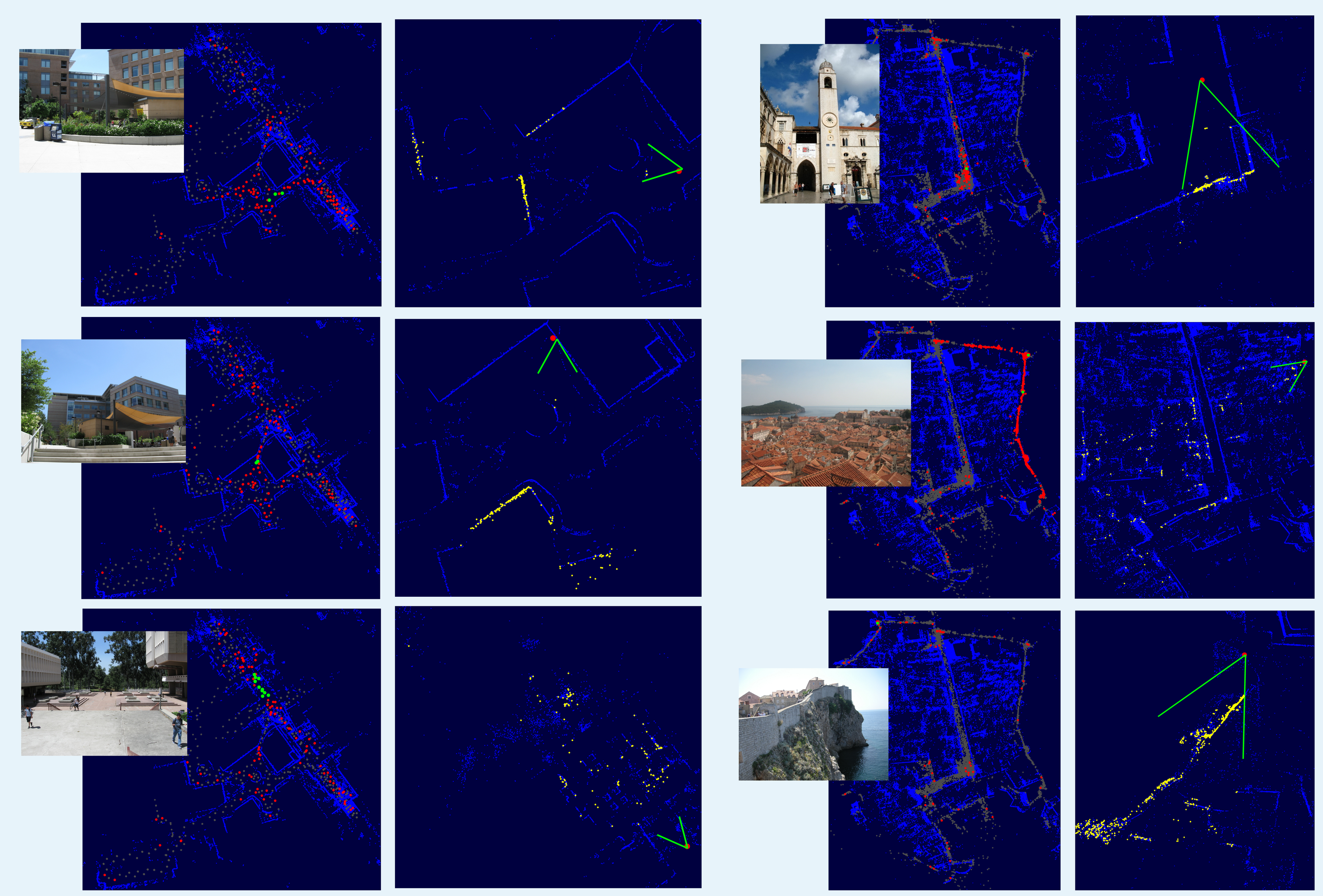
Anytime performance:





4. Experimental Results

- We retrieve global matches using a global k-ratio test from 5 nearest neighbors per query feature.
- Forward match of N_F=200 features using random sampling, and prioritize back matching up to N_B=200 features.
- We achieve state-of-the-art results on a handful of datasets with challenging repetitive structure.



Eng-Quad Dubrovnik

Localization Results

Eng-Quad 520	#images	#inliers	Median error [m]	Time [s]
Sattler et al 2011	402	43	2.01	1.52
Sattler et al 2012	457	43	1.93	0.32
50 Exhaustive	512	66	0.45	56.822
Ours (P3P)	509	112	0.67	0.69

Dubrovnik 777	#images	#inliers	Median error [m]	Time [s]
Sattler et al 2011	771	70	1.44	2.58
Sattler et al 2012	775	69	1.58	0.75
Ours (P3P)	777	591	0.66	0.48

Dubrovnik 800*	#images	#inliers	Median error [m]	Time [s]
Sattler et al 2012	795.5	<200	1.4	0.25
Zeisl 2015	796	-	0.56	3.78
Ours (P4Pf)	800	468	1.64	0.62

Rome 1000*	#images	#inliers	Time [s]
P2F 2010	924	-	0.87
Sattler et al 2012	991	<200	0.28
Ours (P4Pf)	1000	458	0.74

Selected References

[1] J. L. Schönberger and J.-M. Frahm. *Structure-from-motion revisited.* CVPR, 2016.
[2] T. Sattler, B. Leibe, and L. Kobbelt. *Improving image-based localization by active correspondence search.* ECCV, 2012.
[3] B. Zeisl, T. Sattler, and M. Pollefeys. *Camera pose voting for large-scale image-based localization.* ICCV, 2015.

Acknowledgements

This work is supported by NSF grants IIS-1618806 and IIS-1253538.
(*) Datasets with degenerate ground truth.