CNISO SFM

CNISO SFM– Lecture Content

- Motivation
- Background and Related Work
- R&D Challenges
- Evaluation
- Summary and Future Work

Motivation

Structure From Motion

- Single image captures 3D world in a 2D array descriptor
  - Missing Information
- SFM is the process of 3D reconstruction and camera estimation out of multiple images
- SFM is composed of heavy computational CV algorithms
  - 1st phase – Scene Graph: feature extraction, feature matching, RANSAC
  - 2nd phase - Reconstruction: Bundle Adjustments and Triangulation

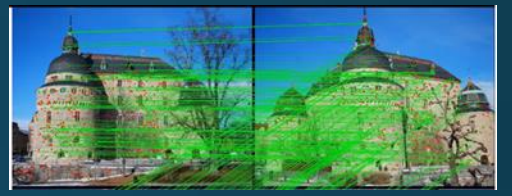
Project Goal

- Problem: SFM systems suffer runtime issues for large scale image datasets
- Goal: Improve efficiency and quality of SFM
- Technique: Using innovative CNISO algorithm introduced by Talker & Moses & Shimshoni
  - Based on spatial order preservation

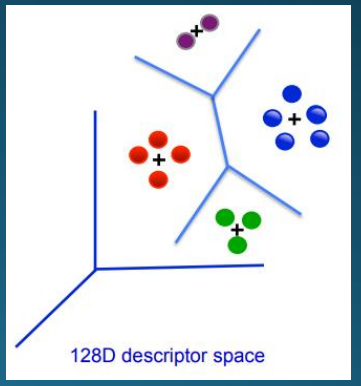
Background and Related Work

Feature Matching

- Given features extracted from 2 images → Finding Correspondence
  - SIFT/SURF for scale and rotation invariance



- Nearest Neighbor(NN)
  - Distance function
  - Lowe's Ratio test → ratio between 1st and 2nd best matches distances
- Vocabulary tree
  - Visual Words built offline (K-means)
  - Two descriptors match if belong to same cell
- RANSAC for outliers removal



RANSAC in Feature Matching

- The process (repeat:)
  - Select Randomly a minimum number of points to determine model parameters
    - 8 points (matching pairs) required for Fundamental Matrix.
  - Solve parameters of the model
    - 8 points algorithm.
  - Determine number of inliers by how many points fit the model
    - predefined tolerance epsilon.
    - point projected closely to its match is considered an inlier.

RANSAC in Feature Matching (2)

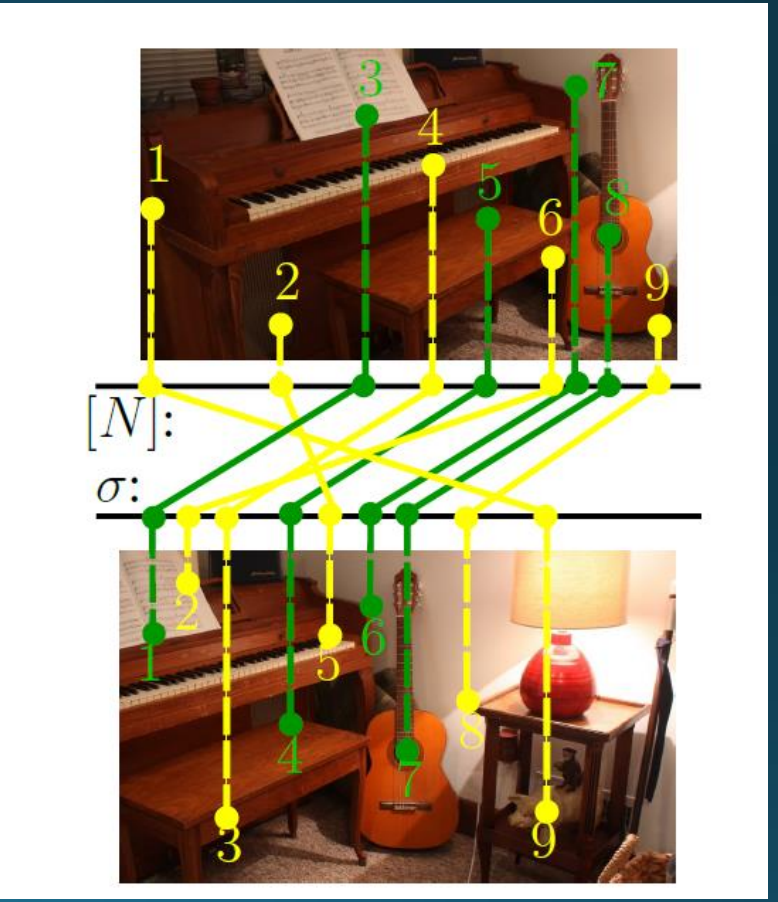
- Process Termination
  - If **#Inliers/#Total Matches > threshold**
  - If reached **maximum of N times**.
- High quality matching enhances RANSAC process
  - Motivation for NN Lowe's Ratio
- Post Termination
  - Re-estimate the model parameters using all identified inliers

CNISO

- Estimating Correct Number of Inliers using only Spatial Order preservation of feature matches.
- Input: matching between pair of images.
  - outliers contaminated
- Output: **number of inliers**, sections of overlap.

CNISO (2)

- Spatial Order preservation of feature matches.
- Under assumptions A1-A3 problem reduced to
  - Counting inversions (Kendall distance)
  - Finding roots of a quadratic equation
  - Performing optimal overlap section search
- Tested on MATLAB



CNISO (3)

Assumptions:

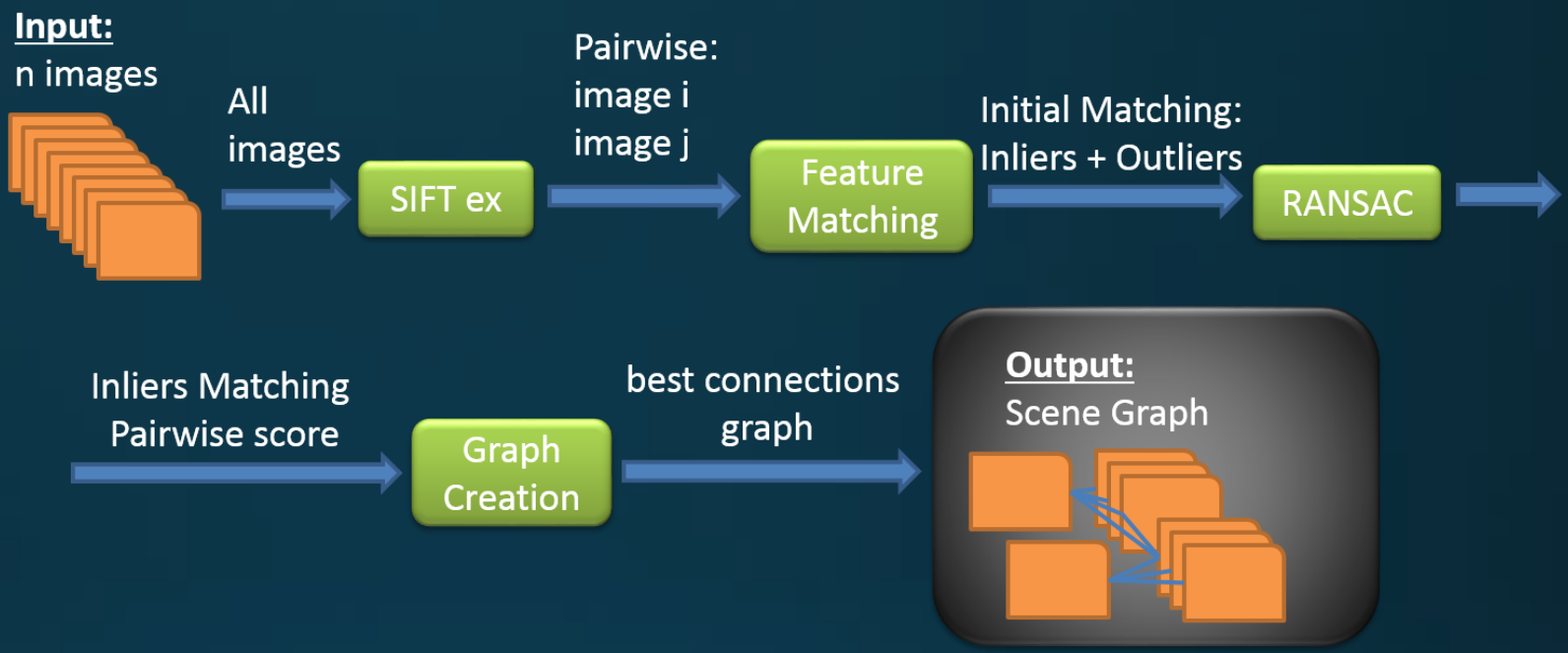
- **A1:** The spatial order of the correctly matched features is preserved
- **A2:** The spatial order of incorrectly matched features is random
- **A3:** The ranks of correctly matched features are distributed uniformly in [N] and in σ
  - **Relaxed (section search)**

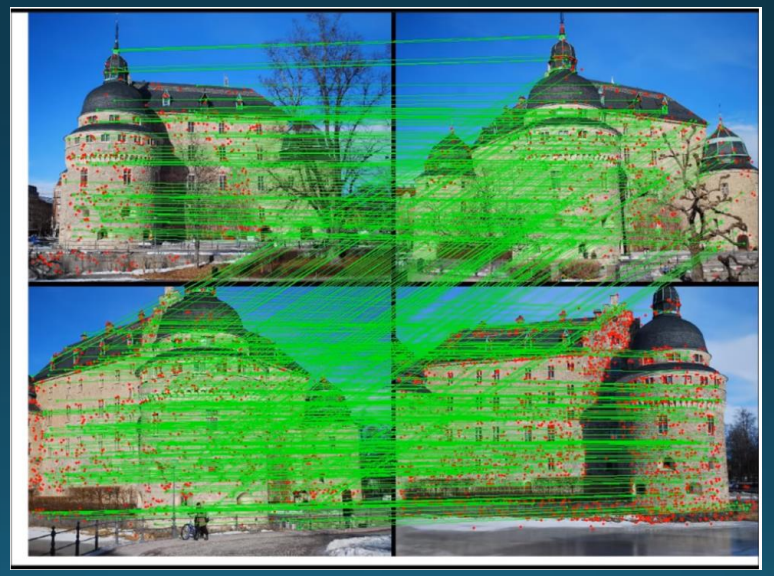
Popular SFM Pipeline

- 1st phase – Scene Graph
  - feature extraction
  - feature matching, RANSAC
- 2nd phase - Reconstruction and Camera Estimation
  - Bundle Adjustments and Triangulation
  - Incremental vs. Global approach

SFM Pipeline

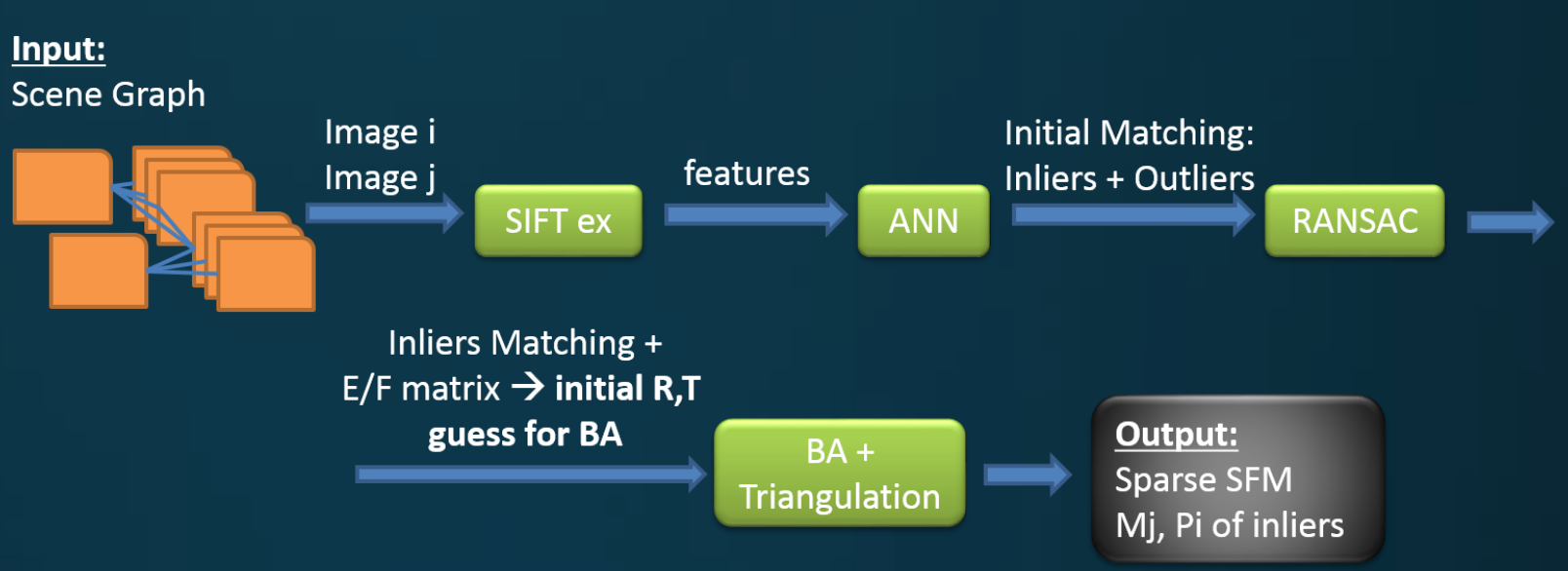
1st phase

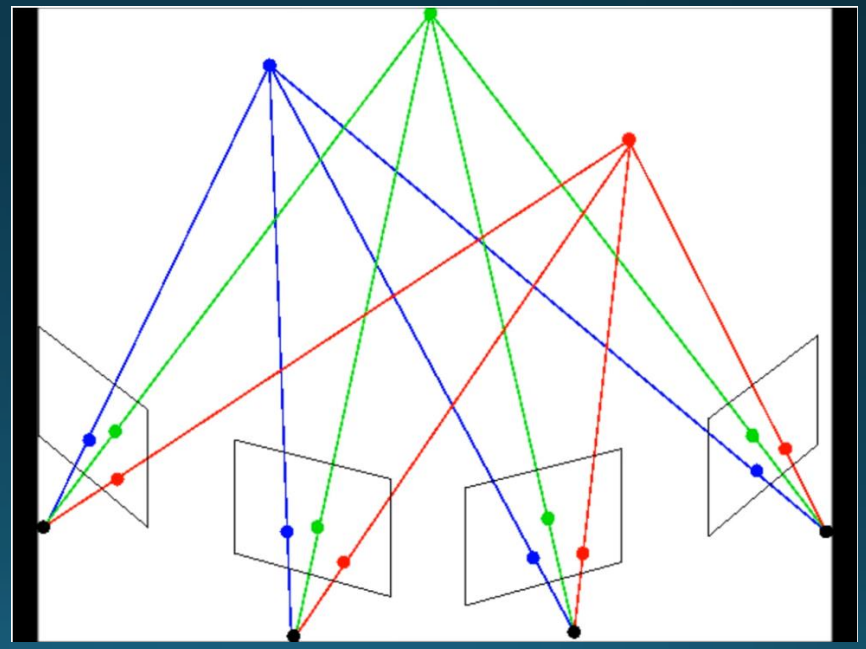




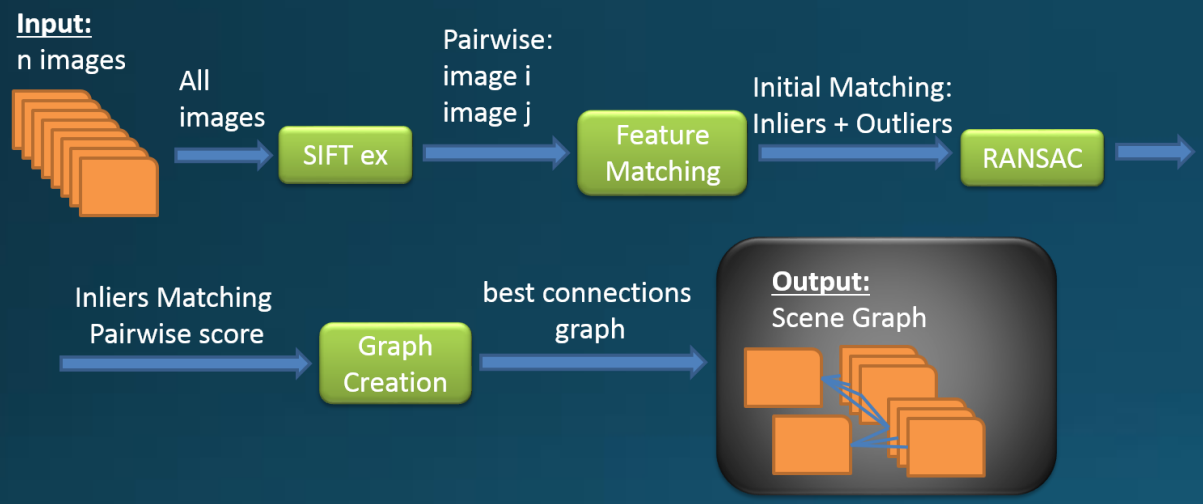
SFM Pipeline

2nd phase





SFM Pipeline – 1st phase



- **Focus on 1st phase**: Scene Graph Creation
  - Input: Collection of images
  - Feature Extraction for each image
    - SIFT/SURF
  - Feature Matching between image pairs
    - Done naïvely between all pairs, or selectively using some preprocessing filtering (image pairs removal)
    - Lowe's Ratio NN \ Approximate NN
    - RANSAC (Remove Outliers)
  - RANSAC Halting Condition
    - Use CNISO for Halting Condition
  - Image Pairs Removal (selective image pairs)
    - BOVW and TF-IDF histogram similarity
    - Vocabulary Tree + VAV / CNISO
    - Lowe's Ratio NN + CNISO

R&D Challenges

CNISO Implementation

- Choose programming language compatible with modern SFM systems
  - implementation in C++ for COLMAP, Theia, Bundler.
- Implement CNISO for wide verity of inputs due to large scale input
- Unit Testing CNISO by comparison to MATLAB output
  - Synthetic inputs created in MATLAB, automatically converted to C++ syntax
- Usability
  - Based solely on C++ STL
  - Simplifying integration to CMAKE build systems
  - Expand API - Adjust input for unordered matching by X positions

COLMAP SFM open source

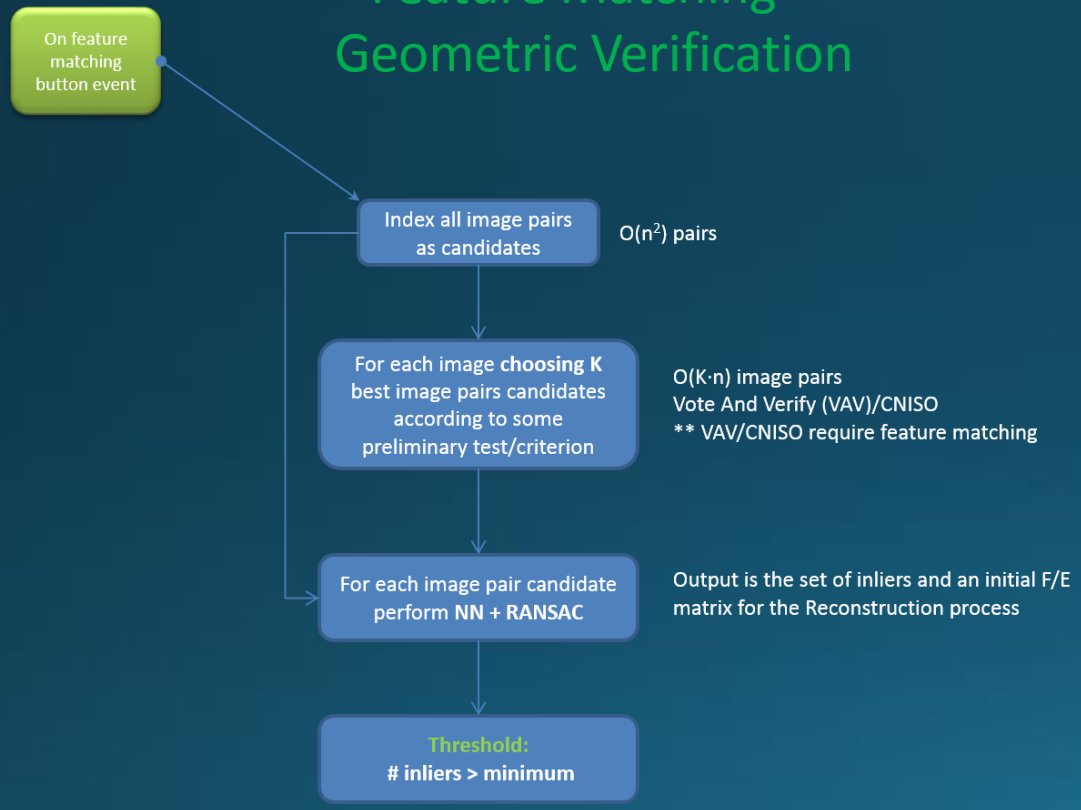
- COLMAP SFM open source was chosen for this project
- Holds *State-of-The-Art* algorithms and pipeline
- Considered as modern and emerging SFM system
- Cross-Platform open-source implemented in C/C++

COLMAP Technological Overview

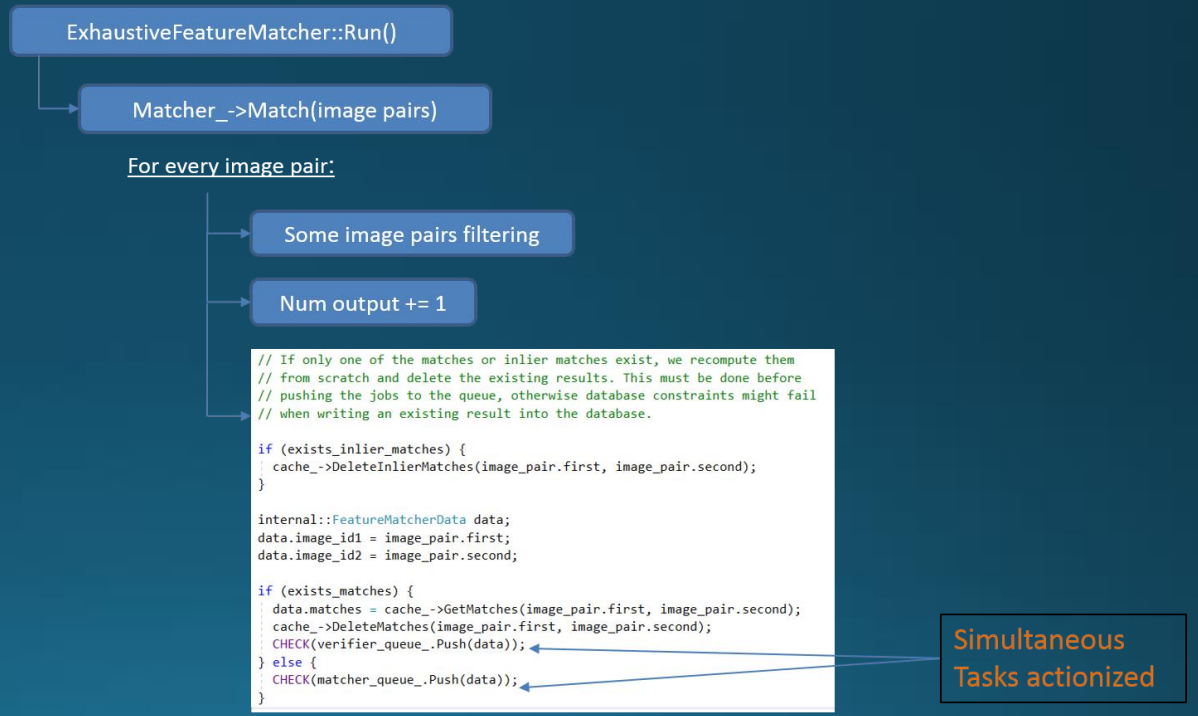
- Windows\Linux\MAC
- C\C++ CMAKE cross-platform build system
- Boost and Ceres-Solver (Google) math
- QT and OpenGL for 3D GUI
- SQLite DB
- Python Utilities (build process)
- Optional: Cloud Clusters & CUDA GPU capabilities

COLMAP Pipeline: Scene Graph

Feature Matching + Geometric Verification



ExhaustiveFeatureMatcher::Run()



COLMAP Configurations

- Lack of CUDA GPU
  - Leads to depth into details of COLMAP
  - Handling by
    - Decreasing image size and number of features
    - DB manipulations
- Parameters Tuning
  - Literature & Experiments
  - K, SIFT peak threshold, #minimum inliers threshold

Integrating CNISO in COLMAP

- **Image Pairs Removal**
  - Vocabulary Tree + CNISO
  - Lowe's Ratio NN + CNISO
- **RANSAC Halting Condition**
  - Use CNISO for Halting Condition
- **Combine** CNISO in both RANSAC halting condition and Image Pairs Removal

Evaluation

Planned Evaluation

- Data Set preparations
  - Manual Rotation of images for CNISO's algorithm
  - Choosing random images from collection (maintain fairness)
- COLMAP Baseline
  - Exhaustive
  - Vocabulary Tree + VAV
- CNISO's integrated capabilities

Measurements Overview

peak threshold 0.02667: num features ~2000 N=250 (image dataset size)					
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

Comparisons Preview

- Vocabulary Tree + VAV / CNISO **vs.** Exhaustive
- Exhaustive **vs.** Exhaustive + CNISO Pair Removal
- Exhaustive **vs.** Exhaustive + CNISO in RANSAC
- Exhaustive + CNISO Pair Removal **vs.** Exhaustive + CNISO Pair Removal + CNISO in RANSAC
- (Exhaustive **vs.** Exhaustive + CNISO in RANSAC) **vs.** (Exhaustive + CNISO Pair Removal **vs.** Exhaustive + CNISO Pair Removal + CNISO in RANSAC)

Vocabulary Tree CNISO Pair Removal

	peak threshold 0.02667: num features ~2000 N=250 (image dataset size)				
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

Vocabulary Tree CNISO Pair Removal (2)

	Vocab, K=10 CNISO - 15 inliers RANSAC	Vocab, K=10 VAV - 15 inliers RANSAC
Cameras	24/27/23	27/27/27
Images	24/27/23	27/27/27
Registered images	24/27/23	27/27/27
Points	2032/2359/2180	2478/1619/2557
Observations	8022/8834/7880	10074/6763/9166
Mean track length	3.94783/3.74481/3.61468	4.06538/4.17727/3.58467
Mean observations per image	334.25/327.185/342.609	373.111/250.481/339.481
Mean reprojection error	0.538054/0.653009/0.588337	0.525754/0.524831/0.534898
Running Time Matching	1.816 [minutes]	1.863 [minutes]
Running Time Reconstruction	1.629/1.932/1.988 [minutes]	0.869/1.478/0.608 [minutes]
RANSAC failures	26 out of 162-324 at most	20 out of 162-324 at most

** Therefore, difference RANSAC failures at precentage is at least 100*(26-20)/324 = 1.8%, and at most 3.6% (2.5% approximately). Apperantly, this is enough to cause such effect in results.

Vocabulary tree: VAV vs. CNISO

- CNISO faster, VAV higher quality
- Low quality compared to exhaustive → rejecting this alternative

NN CNISO Pair Removal

- Exhaustive **vs.** Exhaustive + CNISO Pair Removal

		peak threshold 0.0 2667: num features ~2000 N=250 (im ge dataset size)			
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

NN CNISO Pair Removal (2)

Runtime:

- 80% time saving in RANSAC operations
  - Removing weak pairs from reaching RANSAC's phase
  - Weak pairs results in slow to none convergence in RANSAC at all
- 15-20% time saving in reconstruction phase
  - Thinner scene graph with strong pairs correspondence

Quality:

- Slightly better mean re-projection error
  - Stronger pairs for accurate and robust reconstruction
- Slightly less cameras registered and 3D points reconstructed
  - Less pairs

CNISO in RANSAC

- Exhaustive **vs.** Exhaustive + CNISO in RANSAC

		peak threshold 0.0[illegible]667: num features ~2000 N=250 (image dataset size)			
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

CNISO in RANSAC (2)

Runtime:

- 40% time saving in RANSAC operations
  - Some strong correspondence pairs stops earlier
  - Some of CNISO's true negative weak correspondence pairs stop earlier
    - Weak correspondence pairs are heavy on RANSAC computation

Quality:

- Practically same in all criterions
- Only difference caused due to "noisy" implementation anomaly
- Even though RANSAC's operation doesn't exhaust itself, the model computed and outliers removed are good enough for fast convergence of reconstruction phase

CNISO Pair Removal + CNISO in RANSAC

- Exhaustive + CNISO Pair Removal **vs.** Exhaustive + CNISO Pair Removal + CNISO in RANSAC

peak threshold 0.02667: num features ~2000 N=250 (image dataset size)					
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

CNISO Pair Removal + CNISO in RANSAC (2)

- Combine both CNISO pair removal and CNISO in RANSAC for further improvements.

Runtime:

- 15% time saving in RANSAC operations

Quality:

- Same. No deterioration and no noise

CNISO Pair Removal + CNISO in RANSAC (3)

		peak threshold 0.02667: num features ~2000 N=250 (image dataset size)			
	Original Vocab Tree Vocab, K=11 (i.e. 10) VAV - 15 inliers	Original Exahustive 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC 15 inliers RANSAC Dataset after manual rotation intervention	Exahustive + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention	Exahustive + CNISO in RANSAC + CNISO pairs removal 15 inliers RANSAC 15 inliers CNISO's NG thresh Dataset after manual rotation intervention
Cameras	39	76	75	72	72
Images	39	76	75	72	72
Registered images	39	76	75	72	72
Points	11300	13500	13200	13100	13100
Observations	42000	51000	49500	49000	49000
Mean track length	3.72	3.75	3.75	3.74	3.74
Mean observations per image	1070	670	660	680	680
Mean reprojection error	0.46	0.43	0.42	0.42	0.42
Running Time NN	72 [minutes] = 1.2 [hours]	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours	don't care - 12 hours
Running Time Matching		0.69 [minutes]	0.43 [minutes]	0.15 [minutes]	0.12 [minutes]
Running Time Reconstruction	7 [minutes]	~6.5 [minutes]	~6.5 [minutes]	~5.5 [minutes]	~5.5 [minutes]

CNISO Pair Removal + CNISO in RANSAC (4)

- The effect of CNISO pair removal on CNISO in RANSAC
- RANSAC's Curve improvement decreased (40% vs. 15%)
- At Exhaustive + CNISO in RANSAC (40%)
  - Both high and low correlated pairs are considered
  - Weak pairs responsible for massive portion of RANSAC operation
- At Exhaustive + CNISO Pair Removal + CNISO in RANSAC (15%)
  - Low correlated pairs are filtered before RANSAC operation → Improvement only relates to high correlated pairs

Summary and Future Work

Summary

- RANSAC Halting Condition
  - 40% time savings for RANSAC's operation
  - No quality deterioration for reconstruction (Highest quality Exhaustive)
- Pairs Removal - Three main approaches
  - 1st Exhaustive NN matching for all image pairs
    - High quality but slow
  - 2nd Vocabulary tree matching VAV
    - Fast but relatively poor quality
  - 3rd Exhaustive NN matching with CNISO for pair removal
    - Evaluated 11% faster compared to Exhaustive (4 minutes against 4.5 minutes)
    - Still very high quality
    - When combined with RANSAC's halting condition
      - Evaluated 13% faster compared to Exhaustive
      - High quality remains the same

Future Work

- Run COLMAP with CUDA
- Use CNISO's number of inliers as weights for scene graph and use overlapping sections to determine left and right images for incremental SFM order decisions.
  - In COLMAP no need for CUDA for this purpose
- Use CNISO's overlapping sections for image pairs removal in Scene Graph phase, by multiple pairs conflicts detection.

Questions

THANKS!