NRI: Collaborative Research: Autonomous Quadrotors for 3D Modeling and Inspection of Outdoor Infrastructure

PI: Junaed Sattar

junaed@umn.edu

Interactive Robotics and Vision Lab, Minnesota Robotics Institute

Department of Computer Science and Engineering

University of Minnesota

Driven to Discover℠

Goals

- To develop technologies to collect visual and inertial data necessary for constructing, offline, high-accuracy 3D maps of the structure for civil and industrial infrastructure
- to introduce algorithms for online processing including localization, path planning and obstacle avoidance.

Partners

1) Junaed Sattar (PI, U Minnesota)
2) Stergios Roumeliotis (Former PI, U Minnesota)
3) Philippos Mordohai (co PI, Stevens Institute of Technology)
4) Peter Seiler (co PI, University of Michigan)

junaed@umn.edu

Interactive Robotics and Vision Lab, Minnesota Robotics Institute

Department of Computer Science and Engineering

In 2020-2021

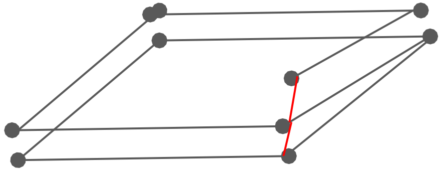
- Minnesota:
  - A Fast and Robust Place Recognition Approach for Stereo Visual Odometry Using LiDAR Descriptors (Mo, Sattar)
  - Learning Rolling Shutter Correction from Real Data without Camera Motion Assumption (Mo, Islam, Sattar)
  - Saliency-guided Visual Attention Modeling (Islam, Wang, De Langis, Sattar)
- Stevens:
  - Multi-view Surface Reconstruction (Batsos, Joyce, Mordohai)
  - Fast stereo 3D reconstruction (Batsos, Mordohai)

Place Recognition

Robots recognizing places which they have previously visited

Benefit to SLAM

- Relocalization
- Loop Closure



http://www.cvlibs.net/datasets/kitti/

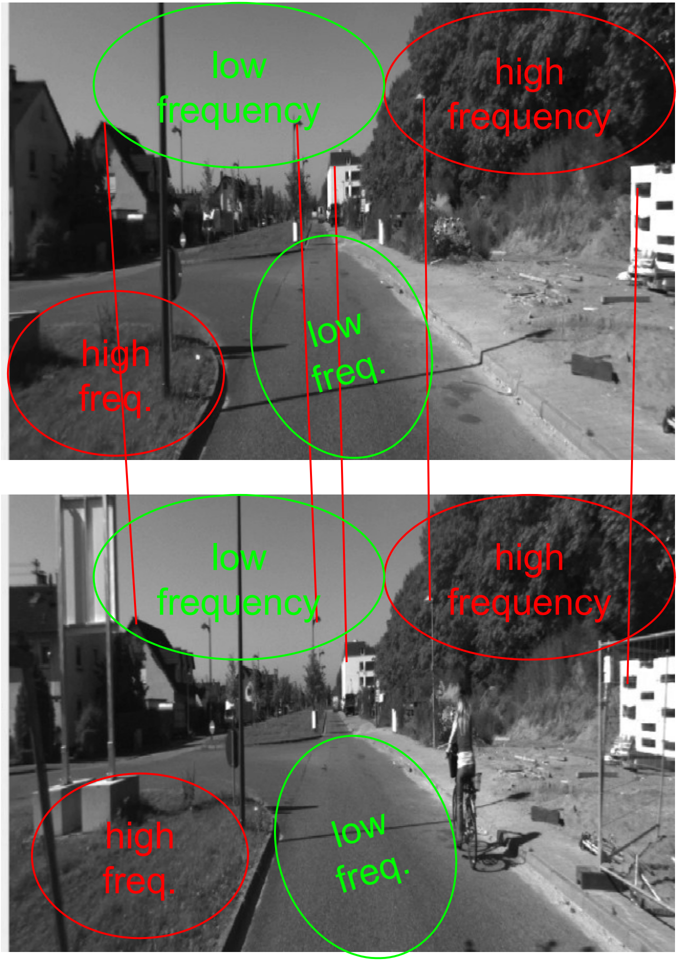
Image-based Approach

Idea

- Use an image to represent a place
- Check similarity between images

Similarity

- Feature correlation
  - BoW[1]
- Spatial layout
  - GIST[2]
- Learning-based
  - NetVLAD[3]



http://www.cvlibs.net/datasets/kitti/

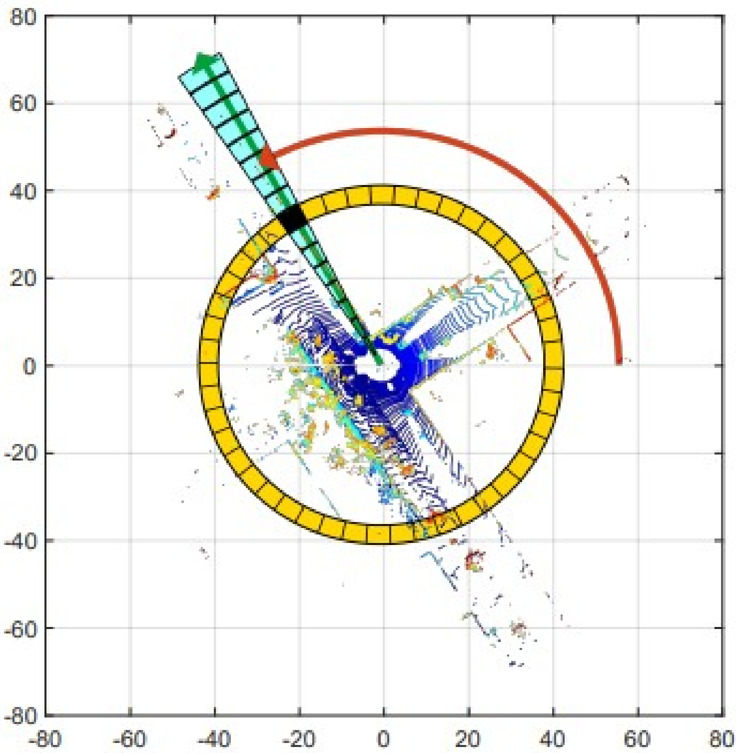
LiDAR Approach

Idea

- Use a point cloud to represent a place
- Check similarity between point clouds

Similarity

- Point cloud alignment
  - ICP[4]
- Point cloud features
  - SHOT[5]
- Global LiDAR descriptor
  - Scan Context[6]



Maximal height in each bin

https://medium.com/@BabakShah/lidar-in-self-driving-cars-cee29db94af7
https://cas-assignment.readthedocs.io/en/latest/assignment.registration.html
https://ieeexplore.ieee.org/document/8593953

LiDAR Approach for Stereo VO

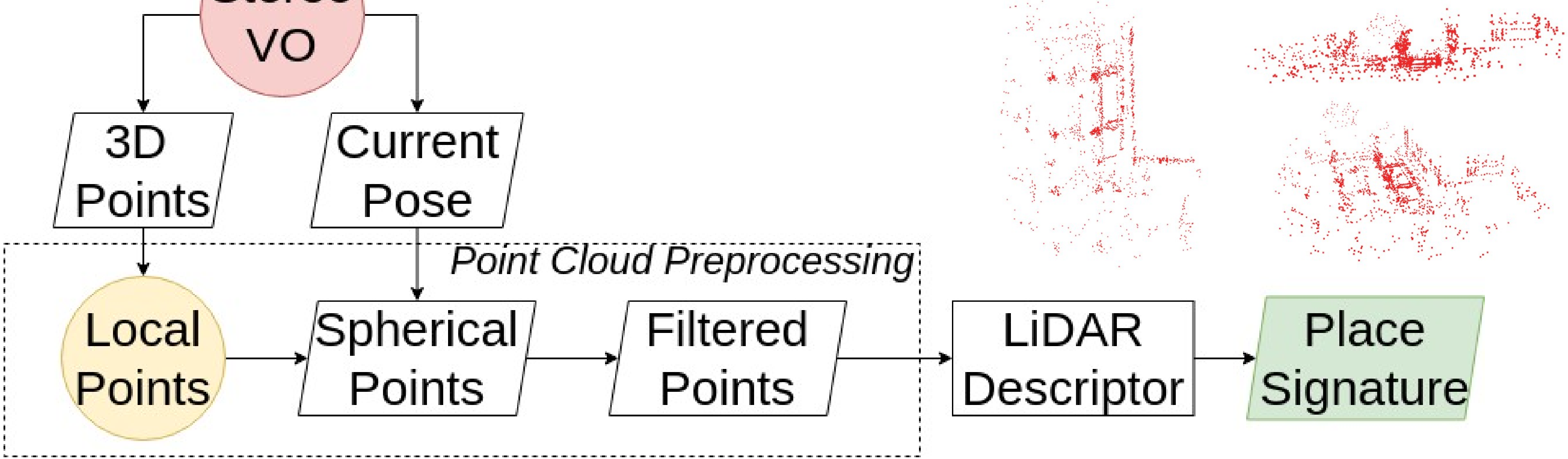
Motivations

- Stereo VO generates 3D points
  - absolute scale
  - not used by image similarity approaches
- 3D points can be potentially more stable than image similarity for place recognition
- Global LiDAR descriptors are computationally efficient

Challenges

1. 3D points are distributed in a frustum
2. Not as consistent as LiDAR scans

Imitating LiDAR Scan



Requirement: the camera motion is predominantly in the forward direction to accumulate points

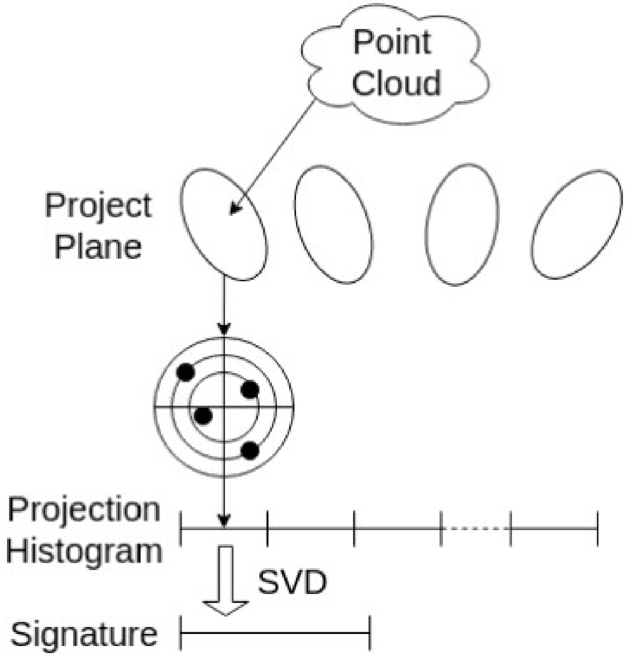
LiDAR Descriptors for Inconsistent Points

M2DP[He et al.], Scan Context[Kim et al.], DELIGHT[Cop et al.]

Idea: Augment these descriptors with 3D structure information and grayscale intensity

Modifications

- Augment the descriptors with grayscale intensity information
  - For each bin:
    - Point count
    - Average grayscale intensity
  - Binarize average grayscale intensity to highlight bright bins
- Replace gravitational alignment with PCA alignment



M2DP

Experiments

Implementation

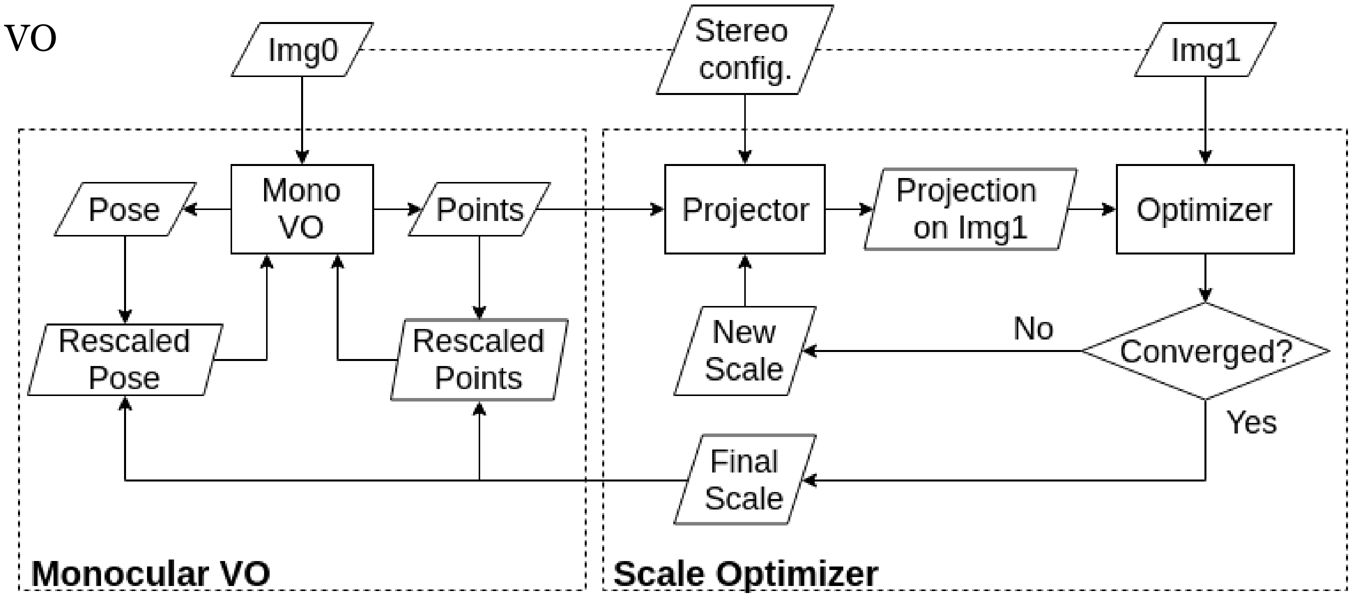
- Stereo VO: SO-DSO[Mo et al.]
- LiDAR range: 45.0m
- Point structure descriptor has twice as much weight than intensity descriptor

Evaluation

- KITTI [Geiger et al.] and RobotCar datasets[Maddern et al.]
- Accuracy metrics
  - the area under the precision-recall curve (AUC)
  - maximal recall at 100% precision

Scale Optimization

- Estimate pose and create 3D points using a monocular VO
- Project 3D points from one camera to the other camera in the stereo rig
- Find the optimal scale that minimizes the projection error
- Rescale the monocular VO

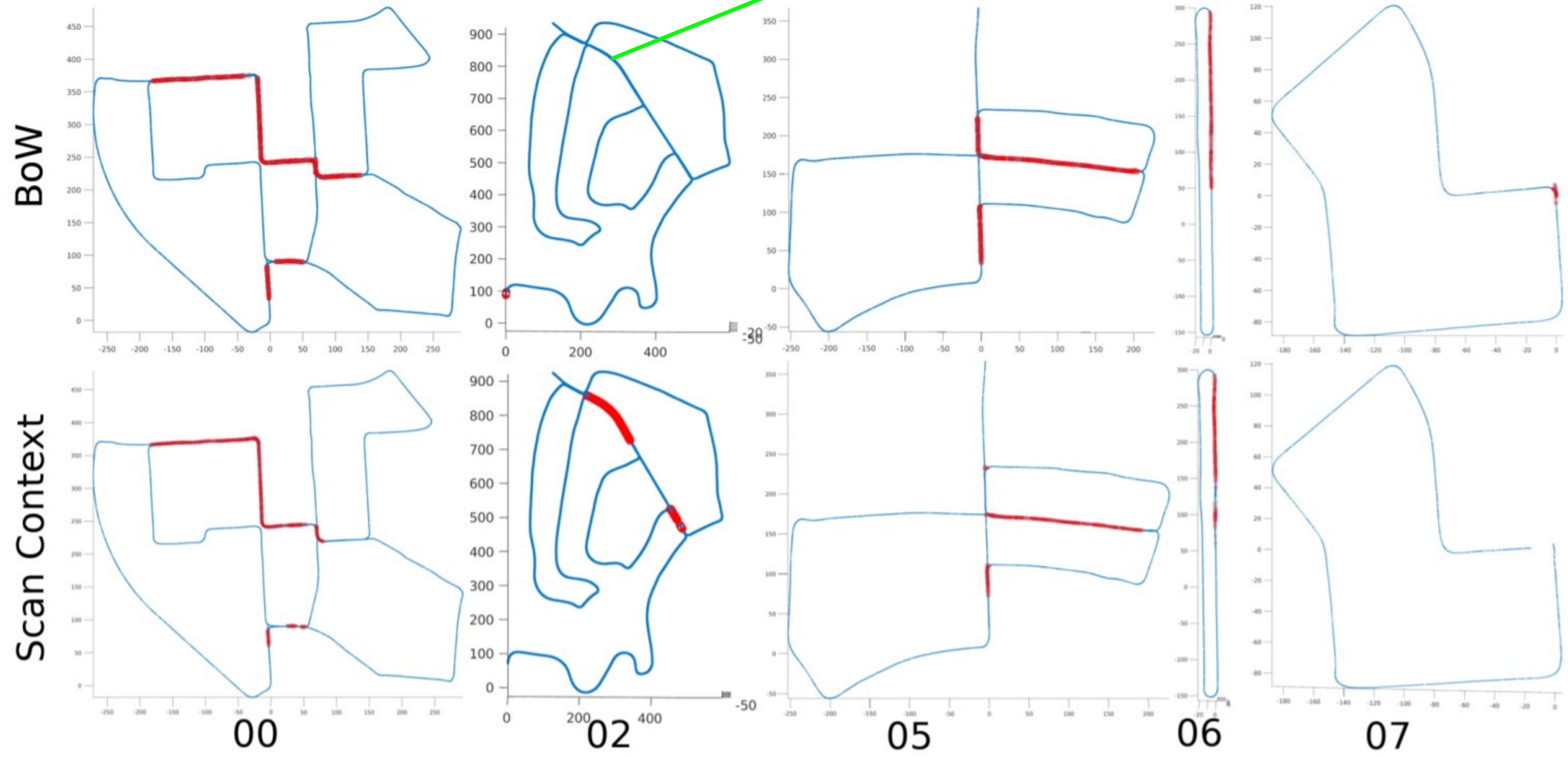


KITTI Accuracy

Method	DELI.	M2DP	S.C.	BoW	GIST
Seq. 00	0.754 0.616	0.639 0.191	0.733 0.599	**0.893 0.788**	0.841 0.774
Seq. 02	0.463 0.253	0.488 0.053	0.555 0.440	0.011 0.012	**0.613 0.597**
Seq. 05	0.622 0.483	0.522 0.062	0.653 0.566	**0.867 0.809**	0.756 0.659
Seq. 06	0.916 0.531	0.946 0.671	0.897 0.679	**0.968 0.963**	0.925 0.729
Seq. 07	0.000 0.000	0.000 0.000	0.000 0.000	**0.713 0.627**	0.350 0.149

TABLE I: AUC (first number) and maximal recall at 100% precision (second number) on KITTI dataset.

KITTI Plots



KITTI Efficiency

Method	DELI.	M2DP	S.C.	BoW	GIST
Imitate LiDAR Scan (c++)	1.151	1.204	**0.692**	-	-
Desc. extraction (c++)	**0.082**	46.10	0.123	37.41	160.0
Query descriptor (Matlab)	103.2	3.418	7.334	115.0	**1.106**
Total	104.4	50.72	**8.149**	152.4	161.1

TABLE II: Run time analysis in milliseconds.

RobotCar

- Challenging for place recognition
  - Recognize place across seasons

(a) Parks Road in spring.

(b) Parks Road in winter.

(c) Holywell Street in spring.

(d) Holywell Street in winter.

RobotCar Accuracy

Tests	Spr. Spr.	Spr. Sum.	Spr. Fall	Spr. Win.	Sum. Sum.	Sum. Fall	Sum. Win.	Fall Fall	Fall Win.	Win. Win.
[12]	0.774	0.736	0.589	0.419	0.764	0.557	0.489	0.599	0.443	0.597
NBLD	0.651	0.700	0.611	0.351	0.672	0.496	0.379	0.454	0.351	0.491
DELI.	0.869	0.677	0.445	0.040	0.836	0.612	0.008	0.498	0.003	0.014
M2DP	0.900	0.851	0.498	0.322	0.853	0.519	0.276	0.540	0.349	0.541
S.C.	**0.956**	**0.944**	**0.782**	0.729	**0.928**	**0.779**	0.618	**0.644**	0.491	**0.814**
BoW	0.558	0.342	0.208	0.300	0.305	0.418	0.371	0.002	0.293	0.001
GIST	0.932	0.918	0.679	**0.778**	0.914	0.694	**0.738**	0.003	**0.606**	0.000

(a) AUC.

Tests	Spr. Spr.	Spr. Sum.	Spr. Fall	Spr. Win.	Sum. Sum.	Sum. Fall	Sum. Win.	Fall Fall	Fall Win.	Win. Win.
DELI.	0.334	0.070	0.026	0.000	0.434	0.187	0.000	0.055	0.000	0.008
M2DP	0.302	0.232	0.001	0.010	0.032	0.011	0.058	0.117	0.039	0.013
S.C.	0.758	**0.558**	**0.408**	**0.322**	**0.685**	**0.415**	**0.325**	**0.346**	**0.247**	**0.519**
BoW	0.032	0.021	0.023	0.031	0.005	0.034	0.100	0.000	0.043	0.000
GIST	**0.794**	0.377	0.242	0.176	0.503	0.242	0.156	0.000	0.109	0.000

(b) Maximal recall at 100% precision.

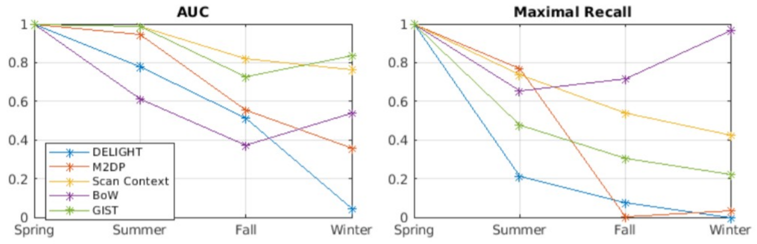


Fig. 9: Robustness against seasonal visual appearance change, using spring as query season. Values are normalized by Spring-Spring.

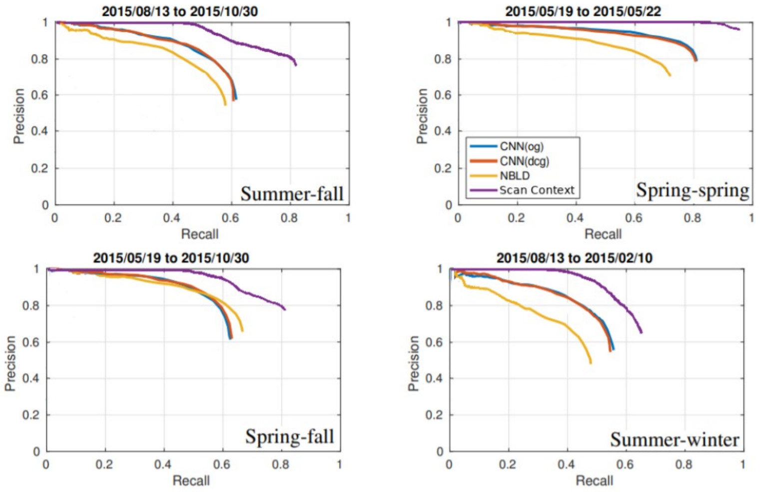


Fig. 10: Precision-recall curves of Scan Context compared with that of [12].

Intensity Contribution

Tests	Spr. Spr.	Spr. Sum.	Spr. Fall	Spr. Win.	Sum. Sum.	Sum. Fall	Sum. Win.	Fall Fall	Fall Win.	Win. Win.
Structure	0.955 0.270	0.940 0.216	0.762 0.390	0.699 0.154	**0.931** 0.279	0.753 0.066	0.610 0.105	**0.652** 0.147	**0.500** 0.049	0.778 0.134
Intensity	0.834 0.230	0.645 0.050	0.344 0.039	0.112 0.021	0.831 0.151	0.390 0.057	0.086 0.027	0.290 0.056	0.096 0.027	0.478 0.032
Fused	**0.956** **0.758**	**0.944** **0.558**	**0.782** **0.408**	**0.729** **0.322**	0.928 **0.685**	**0.779** **0.415**	**0.681** **0.325**	0.644 **0.346**	0.491 **0.247**	**0.814** **0.519**

TABLE V: AUC (top sub-rows) and maximal recall (bottom sub-rows) at 100% precision of Scan Context with structure and/or grayscale intensity.

Use Case Analysis

Proposed approach

- Requirements
  - Stereo cameras
  - Forward motion
- High accuracy and robustness in visually challenging environments
- High efficiency
- Robust to repetitive textures

BoW

- Higher accuracy when there is not much visual appearance change

Links

Code

https://github.com/IRVLab/so_dso_place_recognition

Project Page

http://irvlab.cs.umn.edu/robot-localization/fast-and-robust-place-recognition-approach-stereo-visual-odometry-using-lidar

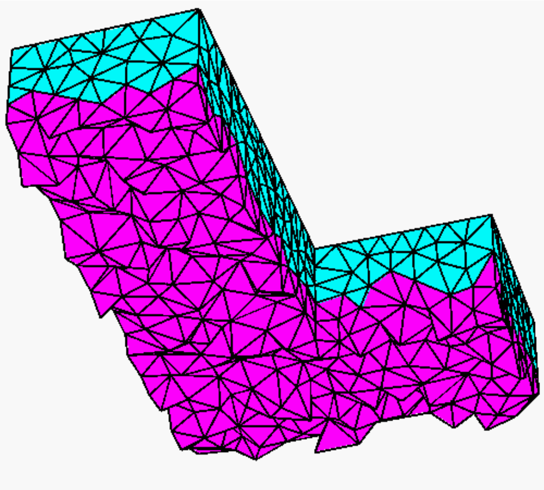
Multi-View Surface Reconstruction

Motivation

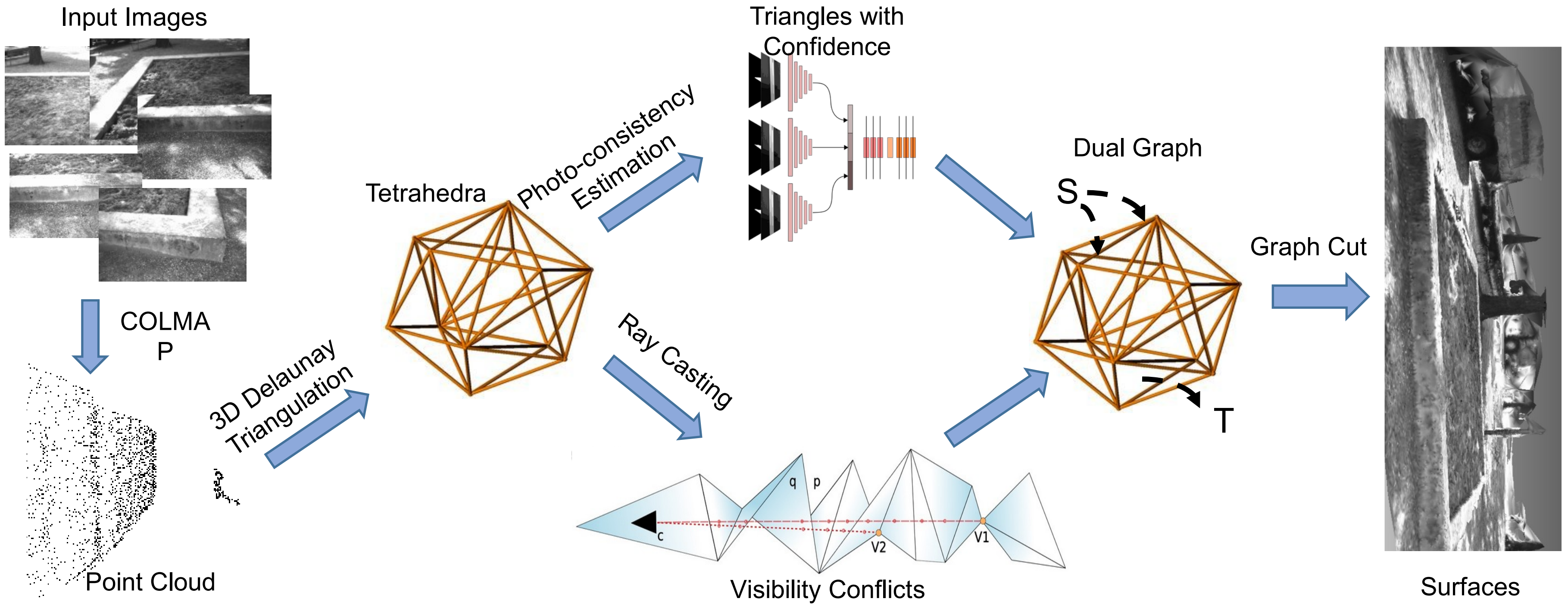
- Unlike binocular stereo, choice of representation is crucial for Multi-view Stereo
  - Volumetric
  - Point clouds
  - Depth map collections
  - Meshes
  - Implicit functions
- Depth map collections are currently the most popular
  - Straightforward to adapt deep cost volume processing to plane-sweeping stereo
  - Piecewise representation without global consistency guarantees

Mesh-based Representation

- Meshes are superior to volumetric and point-cloud representations for rendering, collision avoidance, etc.
- Given initial noisy point cloud, apply 3D Delaunay triangulation to obtain tetrahedra
  - Adaptive density, higher near likely surfaces
- Determine occupancy of each tetrahedron
  - Watertight, globally consistent surface can be obtained as boundary between free and occupied tetrahedra

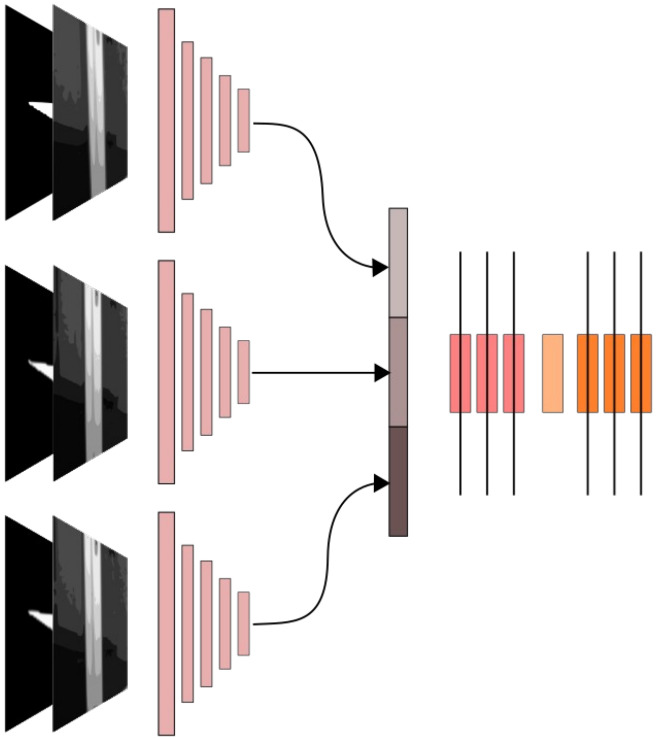


Overview



Input Images
COLMA
P
Point Cloud
3D Delaunay
Triangulation
Tetrahedra
Photo-consistency
Estimation
Triangles with
Confidence
Ray Casting
q
p
c
V2
V1
Visibility Conflicts
Dual Graph
S
T
Graph Cut
Surfaces

Triangle Confidence Estimation

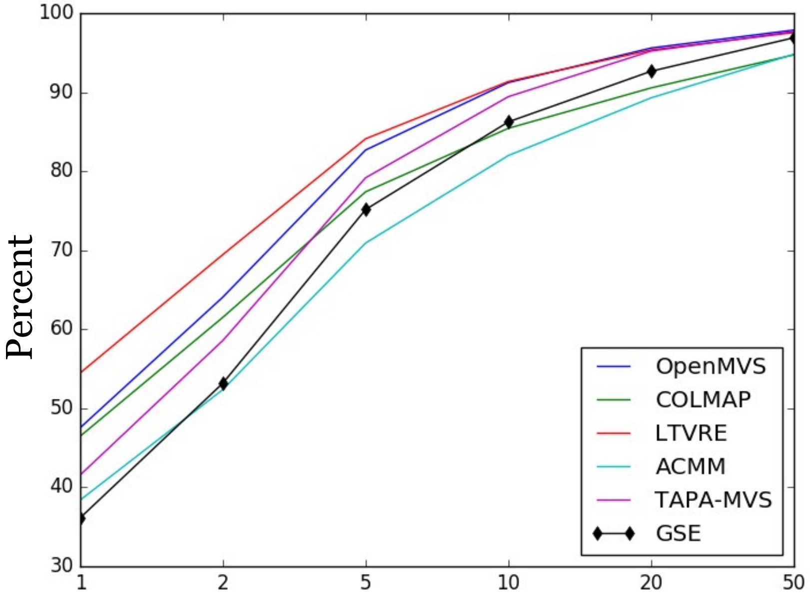


- A siamese network extracts features from the input patches
- The features are concatenated and passed through a number of fully-connected layers to estimate the photo-consistency of the triangle
- The **photo-consistency** estimate along with the **number of conflicts** and the **area of the triangle** are then passed through three fully-connected layers to compute the confidence of the triangle

ETH3D Low-res Many-view Dataset

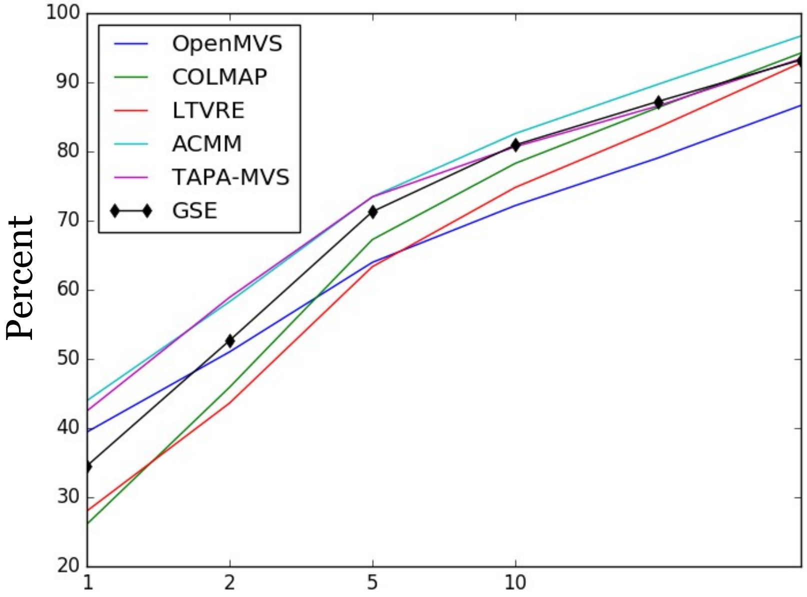
- 5 training and 5 testing scenes
- Each scene contains 400-500 views
- COLMAP point-clouds contain 10s of millions of points
- After simplification, point clouds average around 1 million points
- Delaunay triangulations range from 10s of millions to 100s of millions triangles
- Outputs of our method average less than 1 million points and slightly more than 1.5 million triangles

Quantitative Results: ETH3D Test Set



Threshold
(cm)

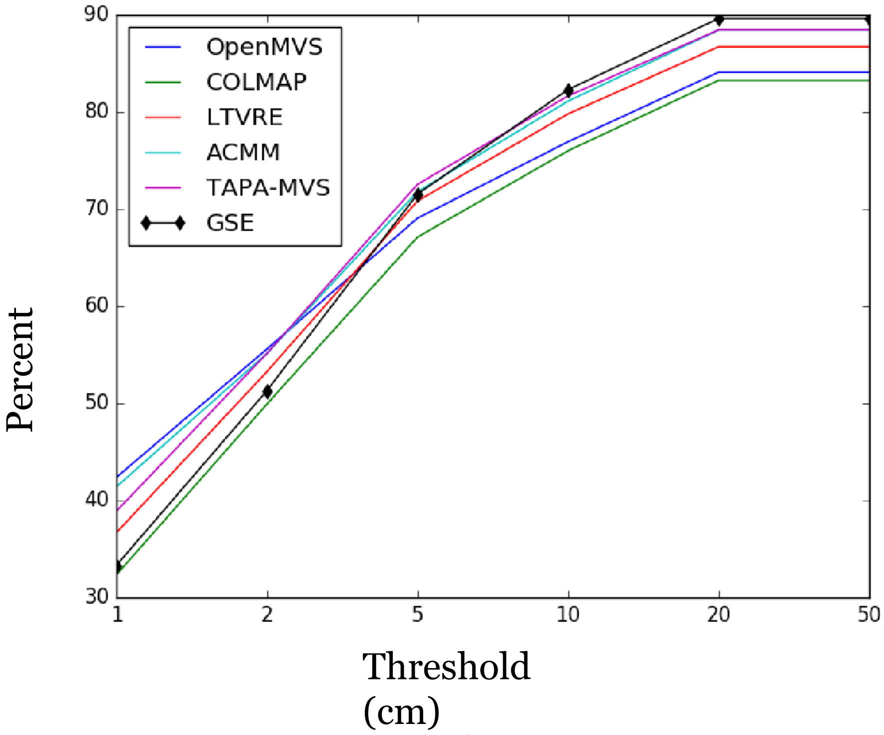
Accuracy



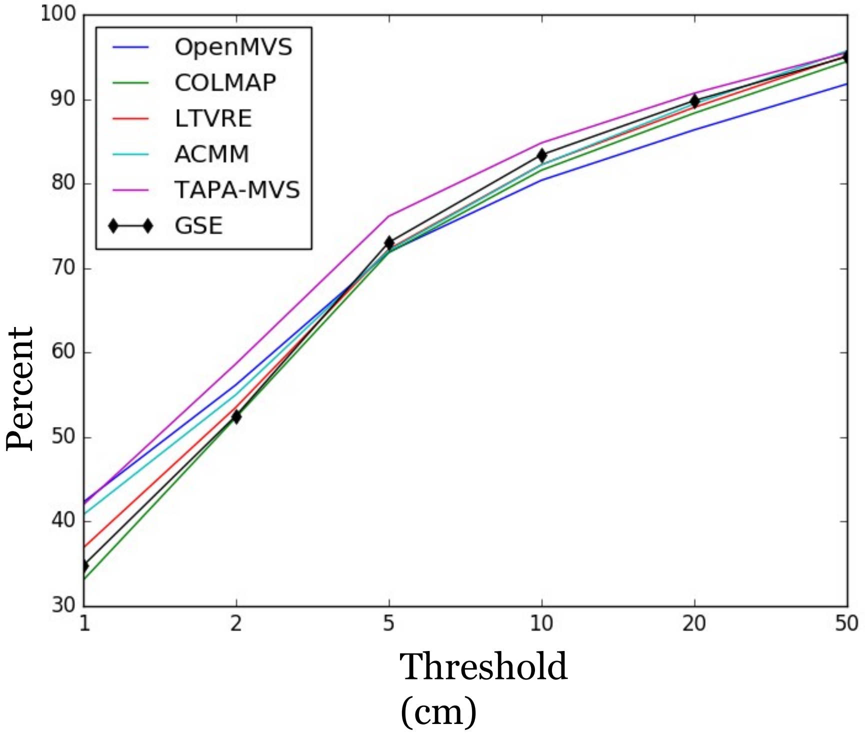
Threshold
(cm)

Completeness

Quantitative Results: ETH3D F-1 Scores

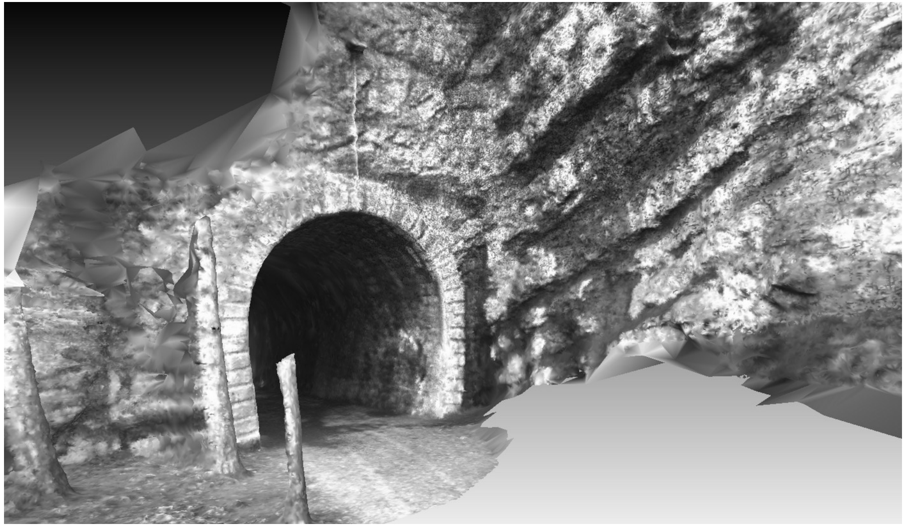


Training Set

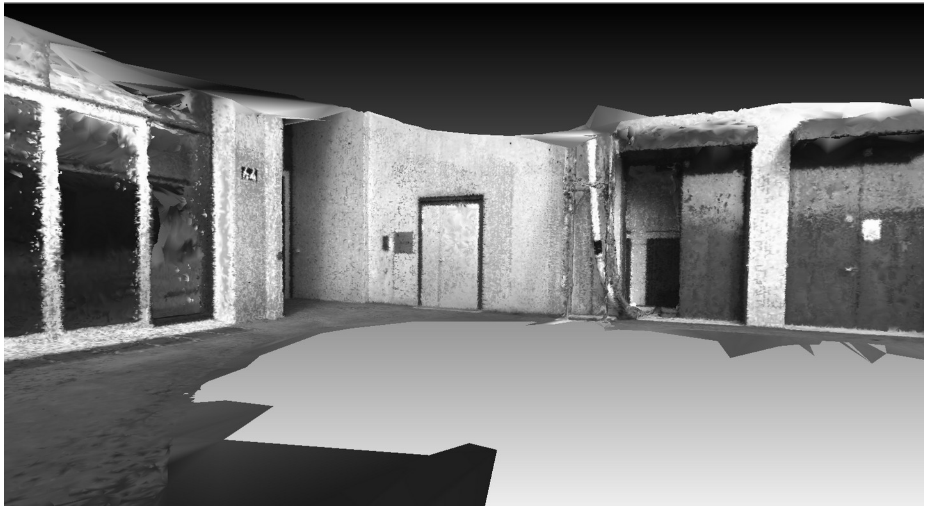


Test Set

Qualitative Results



Qualitative Results



Conclusions

- Our method employs more powerful representation than current SOTA methods
- Not end-to-end
  - Current graph networks limited in number of nodes

References

1. J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” in null, p. 1470, IEEE, 2003.
2. A. Oliva and A. Torralba, “Building the Gist of a Scene: The Role of Global Image Features in Recognition,” Progress in brain research, vol. 155, pp. 23–36, 2006.
3. R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5297–5307, 2016.
4. P. J. Besl and N. D. McKay, “Method for Registration of 3-D Shapes,” in Sensor Fusion IV: Control Paradigms and Data Structures, vol. 1611, pp. 586–607, International Society for Optics and Photonics, 1992.
5. F. Tombari, S. Salti, and L. Di Stefano, “A Combined Texture-Shape Descriptor for Enhanced 3D Feature Matching,” in 2011 18th IEEE international conference on image processing, pp. 809–812, IEEE, 2011.
6. G. Kim and A. Kim, “Scan Context: Egocentric Spatial Descriptor for Place Recognition within 3D Point Cloud Map,” in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4802–4809, IEEE, 2018.
7. K. P. Cop, P. V. Borges, and R. Dube, “DELIGHT: An Efficient Descriptor for Global Localisation using LiDAR Intensities,” in 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 3653–3660, IEEE, 2018.
8. L. He, X. Wang, and H. Zhang, “M2DP: A Novel 3D Point Cloud Descriptor and its Application in Loop Closure Detection,” in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 231–237, IEEE, 2016.
9. J. Mo and J. Sattar, “Extending Monocular Visual Odometry to Stereo Camera System by Scale Optimization,” in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). To appear, November 2019. arXiv preprint arXiv:1905.12723.
10. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The kitti dataset,” The International Journal of Robotics Research, vol. 32, no. 11, pp. 1231–1237, 2013.
11. W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000km: The Oxford RobotCar Dataset,” The International Journal of Robotics Research, vol. 36, no. 1, pp. 3–15, 2017.
12. Y. Ye, T. Cieslewski, A. Loquercio, and D. Scaramuzza, “Place Recognition in Semi-Dense Maps: Geometric and Learning-Based Approaches,” in Proc. Brit. Mach. Vis. Conf., pp. 72–1, 2017.