From Mapping to Localization: A Complete Framework to Visually Estimate Position and Attitude for Autonomous Vehicles

Guoyu Lu
Chester Carlson Center for Imaging Science
Rochester Institute of Technology

Xue-iuan Wong, James McBride
Ford Motor Company

Outline

Framework instruction

Map Generation

Feature Correspondence Establishment

Camera Pose Estimation

Experiment

Conclusion

Outline

Framework instruction

Map Generation

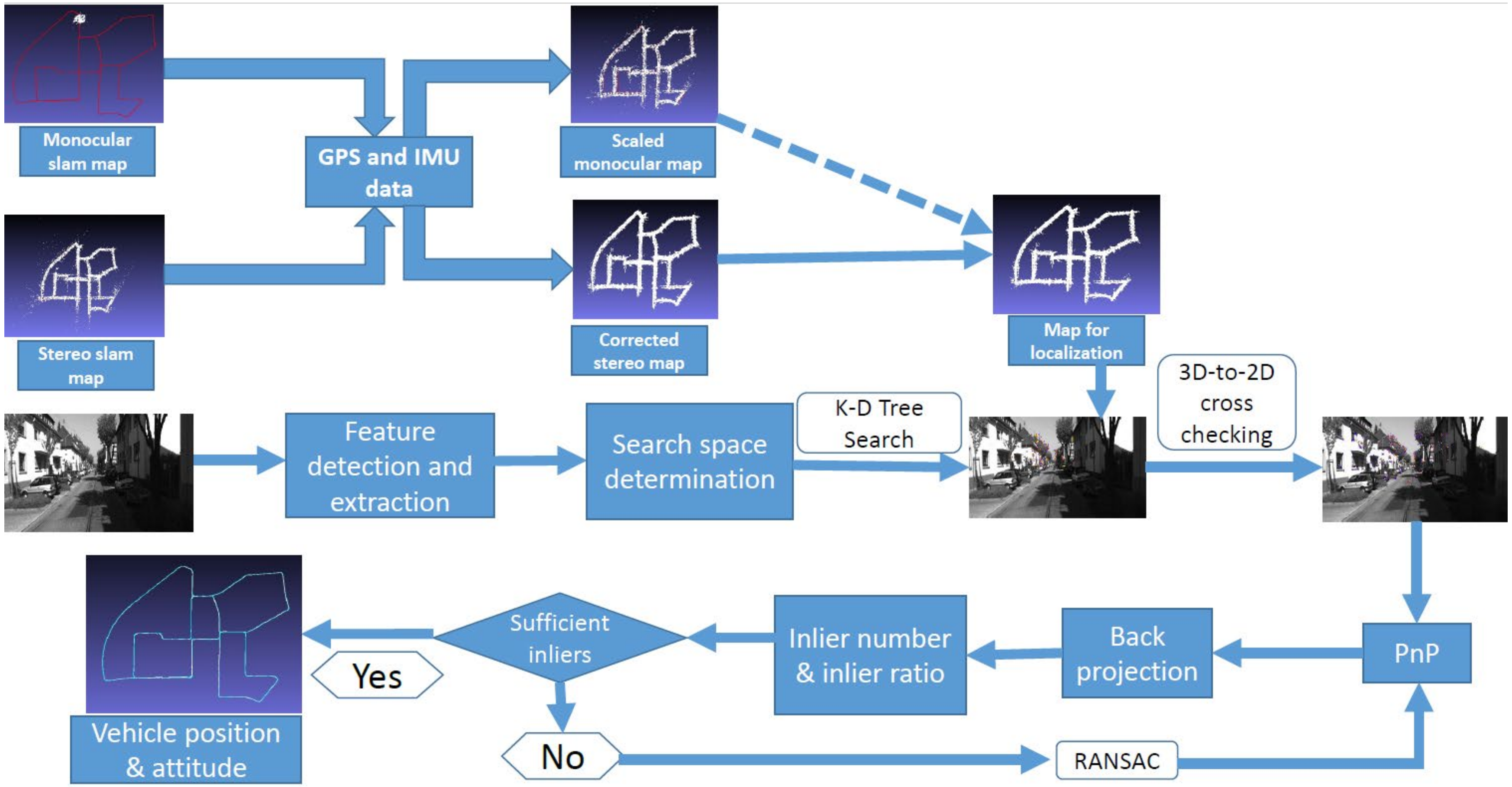
Feature Correspondence Establishment

Camera Pose Estimation

Experiment

Conclusion

Framework



Monocular slam map
GPS and IMU data
Scaled monocular map
Stereo slam map
Corrected stereo map
Map for localization
3D-to-2D cross checking
K-D Tree Search
Feature detection and extraction
Search space determination
Sufficient inliers
Yes
No
Vehicle position & attitude
Inlier number & inlier ratio
Back projection
PnP
RANSAC

Outline

Map Generation

SLAM to generate the initial 3D map due to its real- time and relatively accurate reconstruction performance.

Monocular-SLAM suffers from scale problem.

Both mono-SLAM and stereo-SLAM have the accumulation error issue

Those issue introduces the map error.

Map Generation

Apply GPS and IMU to correct the map.

Most autonomous driving companies (e.g., Google, Baidu, Ford) and map generation companies (e.g., TomTom, HERE maps) have GPS and IMU on their vehicle.

For monocular SLAM, we correct the 3D map scale to be the same as GPS and IMU.

Map Generation

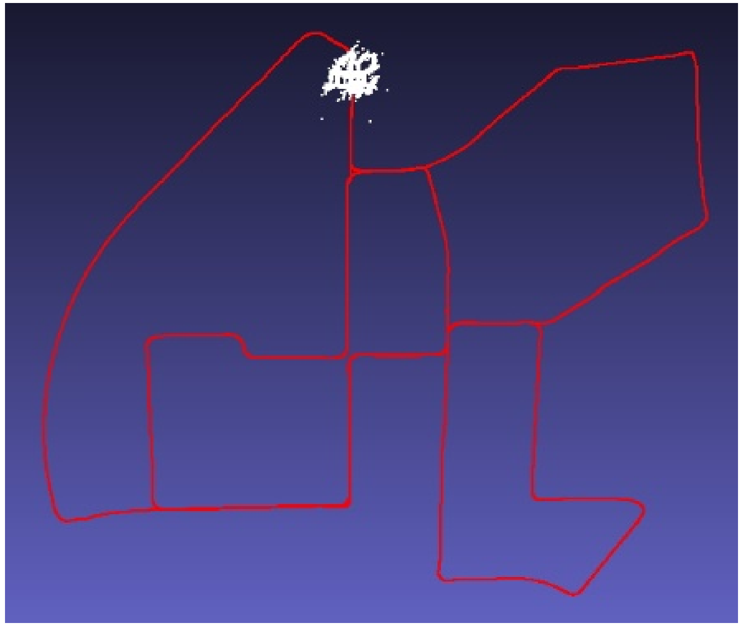
For both stereo and monocular SLAM, we back-project 3D points from world coordinate to camera coordinate based on estimated camera pose

Project 3D points from camera coordinate to world coordinate using GPS/IMU

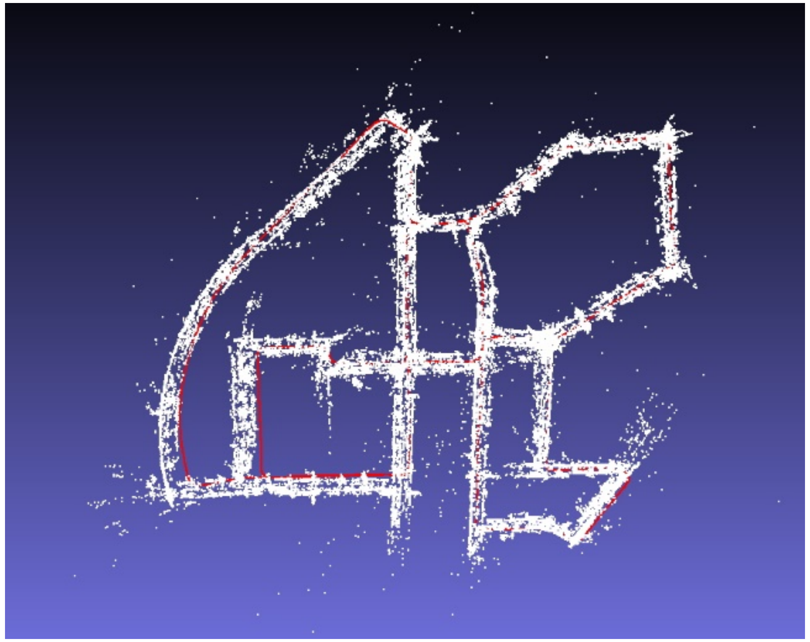
Each 3D point is associated with a descriptor

SLAM/GPS/IMU are only used for map generation, mainly for mapping companies. In online localization stage for consumer vehicles, only the generated map and query image are used to estimate camera pose.

Map from Monocular SLAM

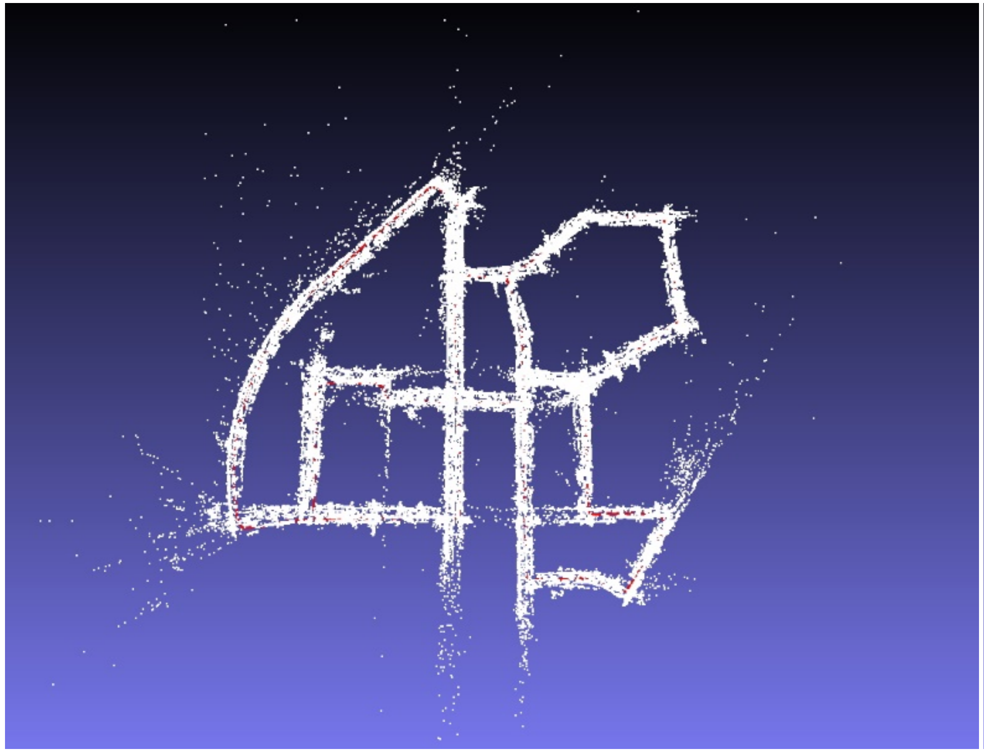


Before correction



After correction

Stereo



Before correction

After correction

Outline

Framework instruction

Map Generation

Feature Correspondence Establishment

Camera Pose Estimation

Experiment

Conclusion

Correspondence Establishment

Extract ORB features in the query image and search the correspondences in the map for each ORB feature.

Searching through the entire map with millions of points will consume too much time.

Restrict the search region to accelerate the matching process

Map Indexing

We use a clustering structure to group the 3D map points

Clustering is based on images where features are extracted. All map points from the same image belong to a cluster

Clusters are indexed in a KD-tree based on their camera poses

Searching Scope Determination

During the localization, we define the search scope as the number of clusters of map points that are used to match the query image points

The feature search scope defining the map point clusters is approximated by the vehicle last known camera pose.

Once the search region is defined, we extract the feature descriptors associated with the 3D points and store them in a temporary K-D Tree for fast searching correspondences.

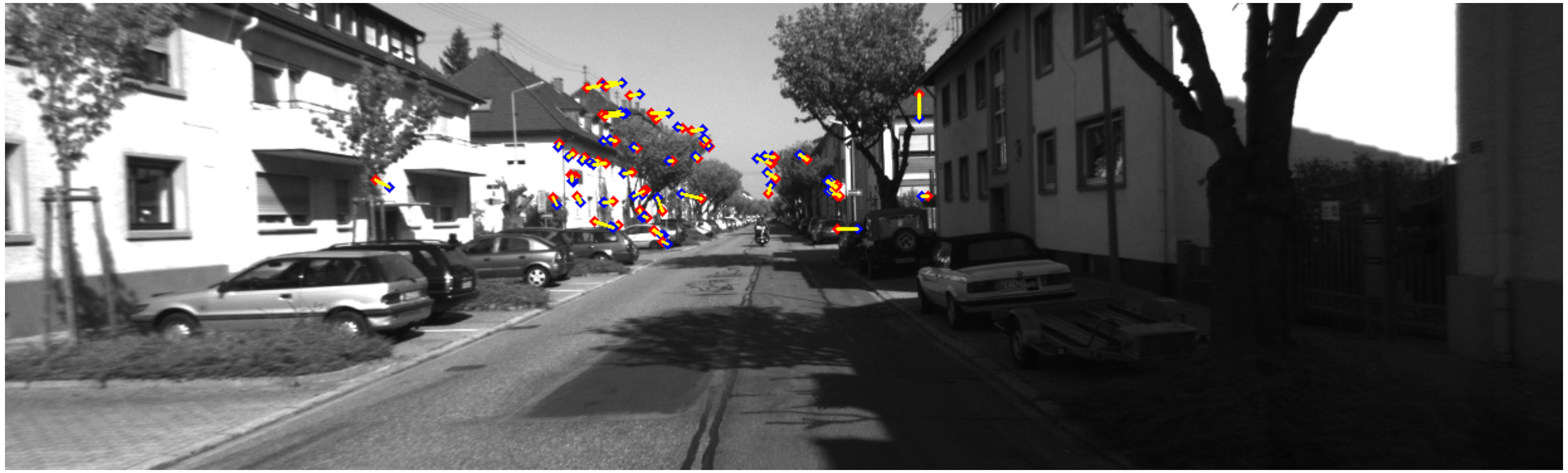
Feature Matching

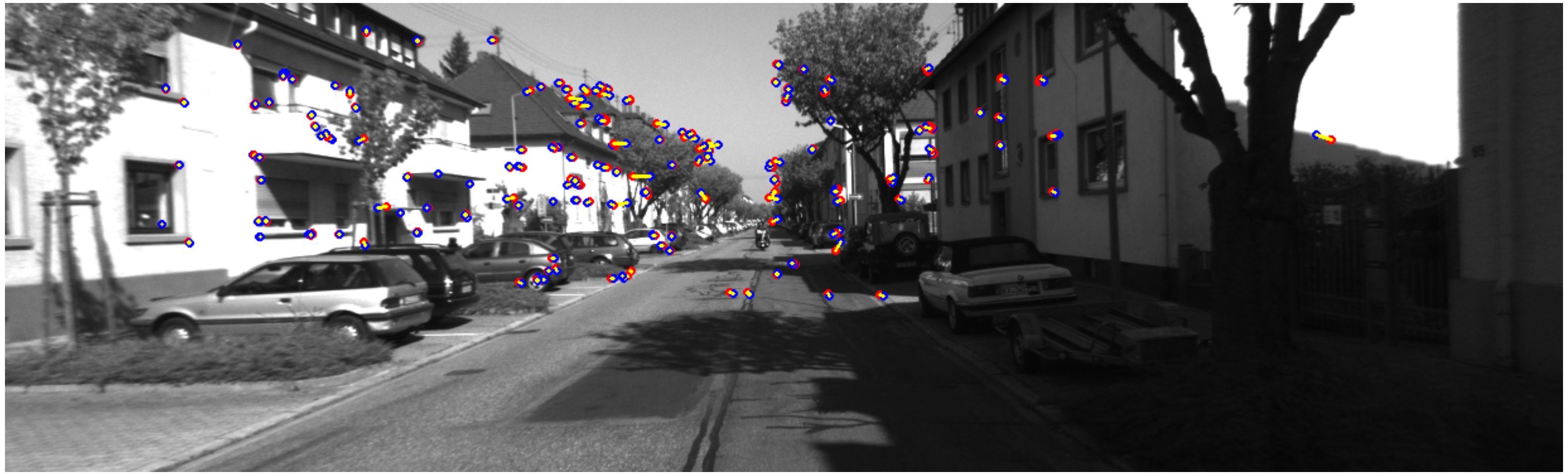
For ORB features, only the feature correspondences with a distance less than 55 are accepted

When a 2D image feature to 3D point feature correspondence is established, the 3D point will also search through the 2D images features to verity this correspondence.

Only if the 2D image feature is also the nearest neighbor of the 3D point, this correspondence is accepted.

Cross matching





Outline

Framework instruction

Map Generation

Feature Correspondence Establishment

Camera Pose Estimation

Experiment

Conclusion

Camera Pose Estimation

Given the correspondences between the 3D points and their 2D projections in image plane, we apply perspective-n-point (PnP) to estimate the camera pose.

To optimize the perspective projection function for PnP, we first use Levenberg-Marquardt (L-M) method to solve the non-linear least square problem for the first frame, as the initial condition may be far from the solution

For later frames, we change to Gauss-Newton (G-N) algorithm to reduce computational cost

Previous localization result as the initial guess of current frame camera pose

Outline

Framework instruction

Map Generation

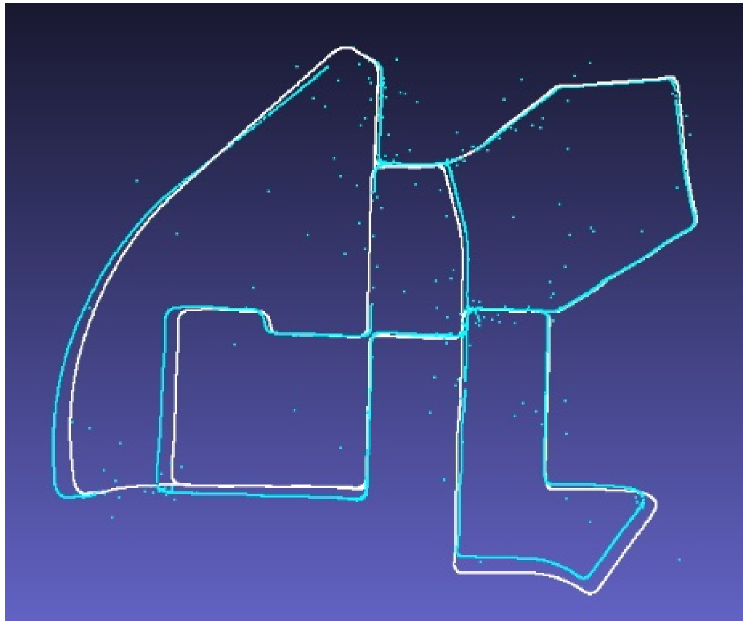
Feature Correspondence Establishment

Camera Pose Estimation

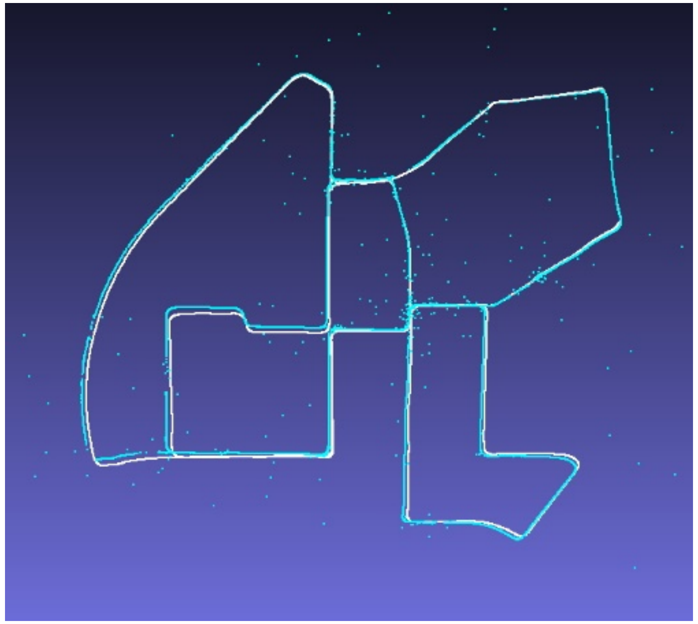
Experiment

Conclusion

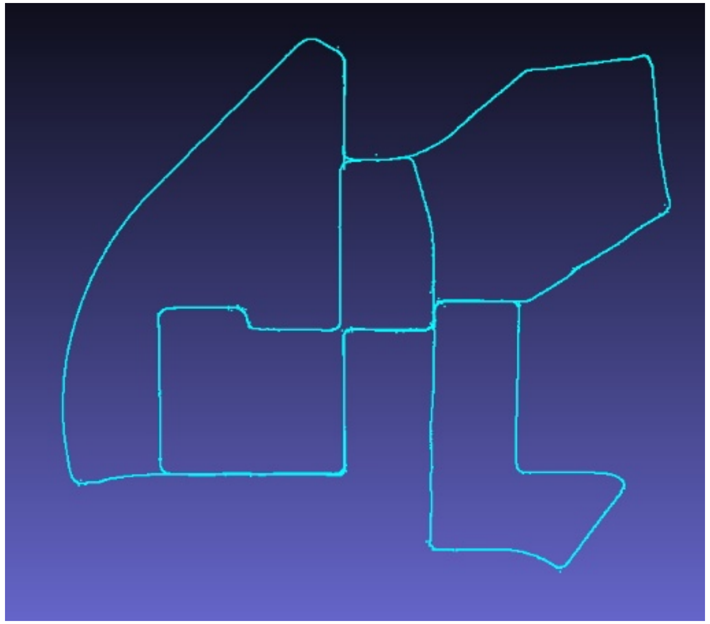
Localization based on different maps



(a) Monocular slam
after scale correction



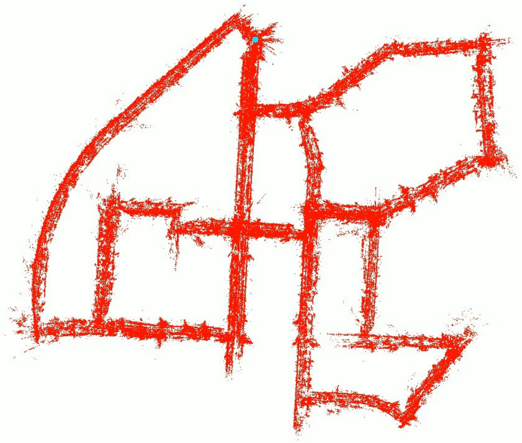
(b) Stereo slam



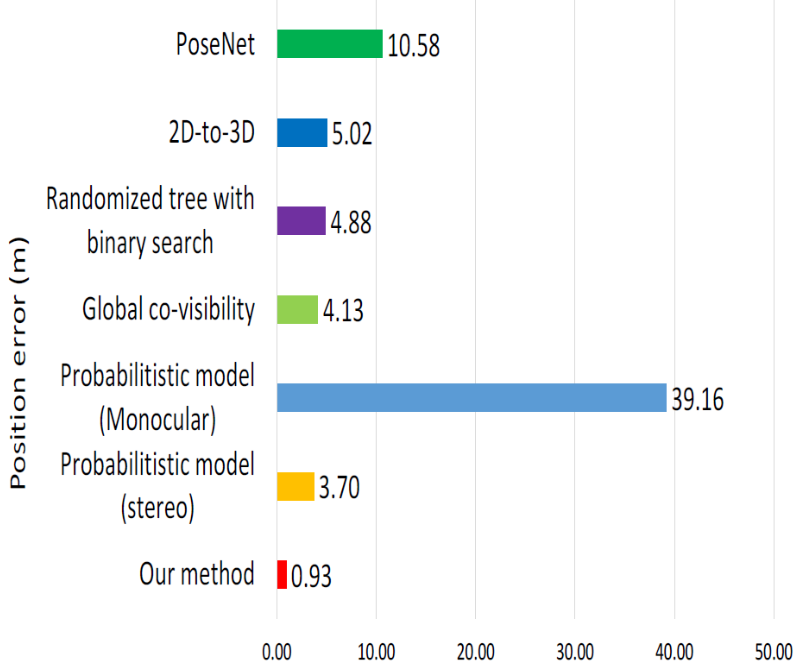
(c) Localization based on the map
generated from our method

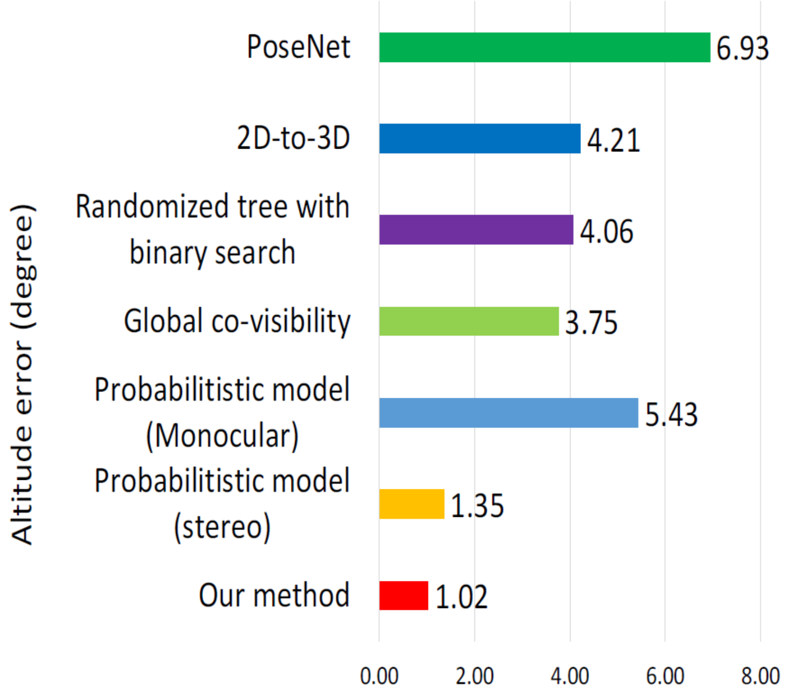
White point: ground truth vehicle path
Blue point: the vehicle path estimated through different maps

Demo for Visual localization for autonomous driving



Comparison with State-of-the-art Localization Methods





Probabilistic model (stereo and monocular) [Brubaker et al., 2016, TPAMI]
Global co-visibility [Liu et al., 2017, ICCV]
Randomized tree with binary search [Feng et al., 2016, TIP]
2D-to-3D [Sattler et al., 2017, TPAMI]
PoseNet [Kendall and Cipolla, 2017, CVPR]

Outline

Framework instruction

Map Generation

Feature Correspondence Establishment

Camera Pose Estimation

Experiment

Conclusion

Conclusion

We build a complete vehicle localization solution for autonomous driving purpose, which includes map generation, feature matching and position and attitude estimation.

We propose a map generation method that can provide sufficient density of map features and high precision.

We propose a feature search strategy dedicated to continuous localization task in autonomous driving, which can accurately and efficiently localize the vehicle.

Our localization method is simple to implement and robust to continuous vehicle pose estimation tasks.

Thank you.

Further questions: luguoyu@cis.rit.edu