Visual Odometry/SLAM (without inertial)

Kostas Daniilidis

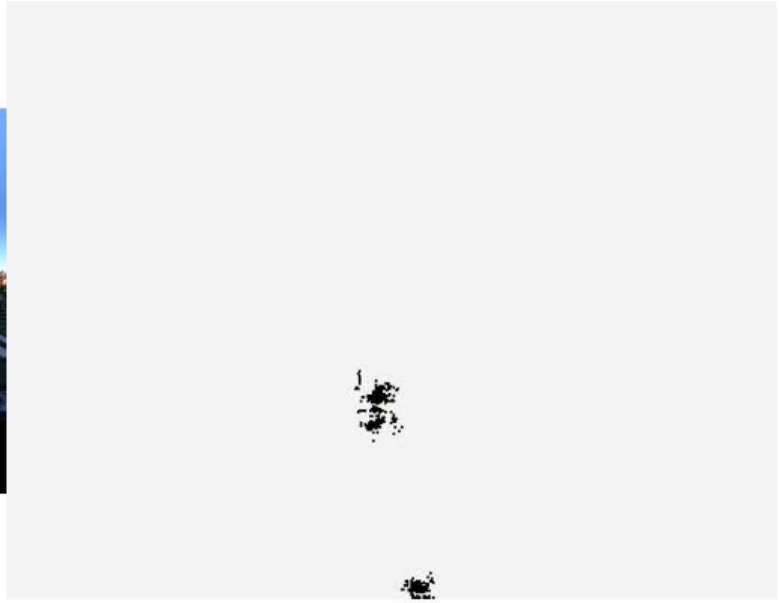
With figures and text from mathworks.com

https://www.mathworks.com/help/vision/ug/monocular-visual-simultaneous-localization-and-mapping.html

Extract camera trajectory from video

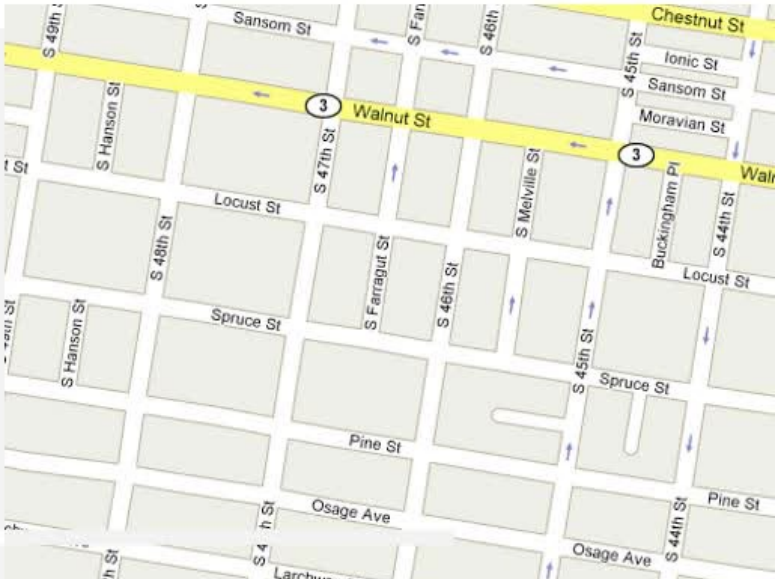
Panoramic image (from 6 cameras)

Reconstruction (global view)



Reconstruction (close-up)

Google Map



UPenn,
Tardif et al. 2008

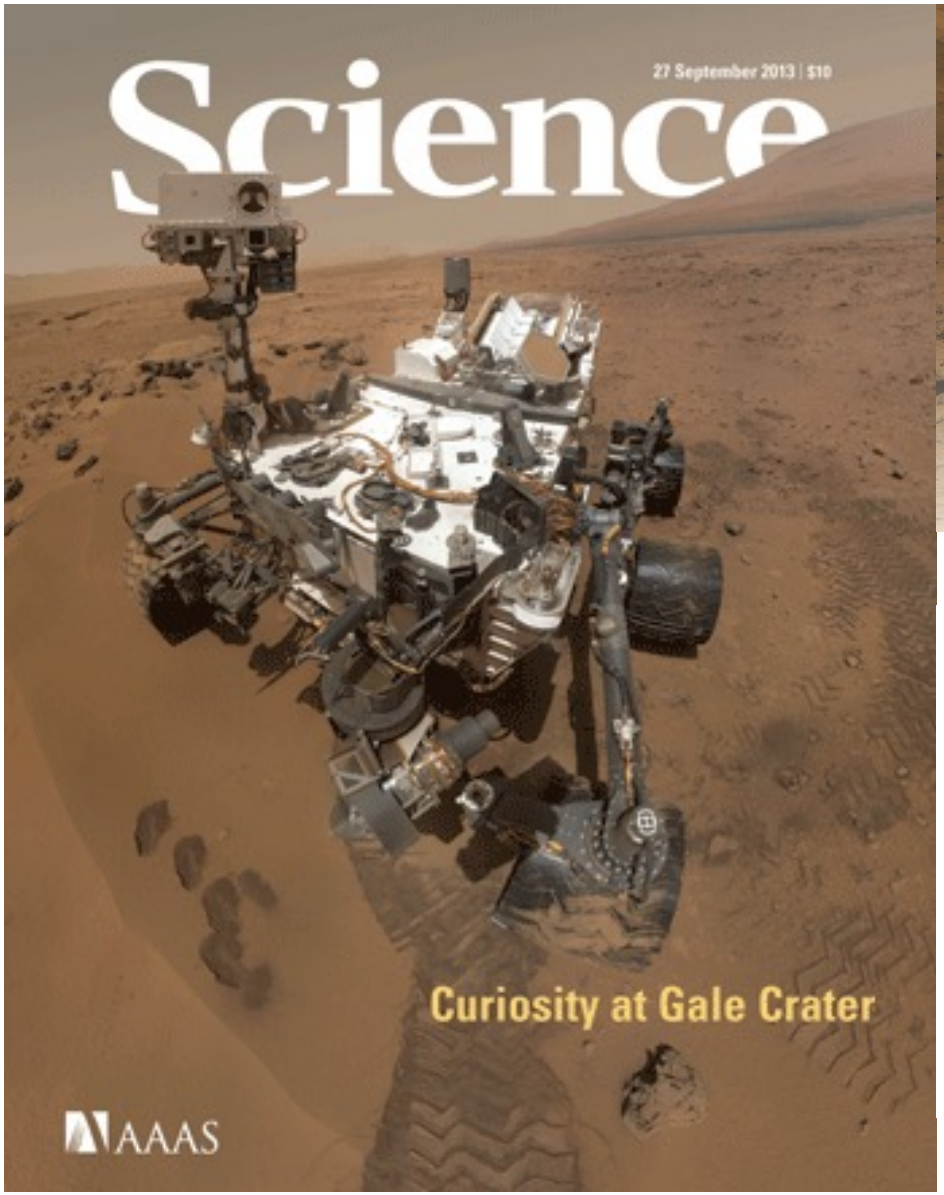
What is Odometry ?

- Measuring how far you go by counting wheel rotations or steps.
- Known as “path integration” in biological perception.
- More general, integration of velocity or acceleration measurements: inertial odometry.

What is Visual Odometry ?

The process of incrementally estimating your position and orientation with respect to an initial reference frames by tracking visual features.

Visual odometry on the MARS



Visual Odometry on the Mars Exploration Rovers

A Tool to Ensure Accurate Driving and Science Imaging

BY YANG CHENG, MARK W. MAIMONE, AND LARRY MATTHIES

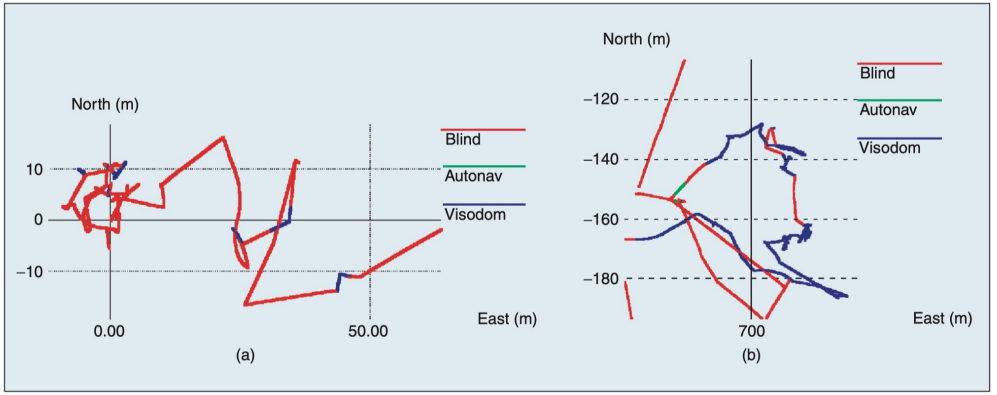
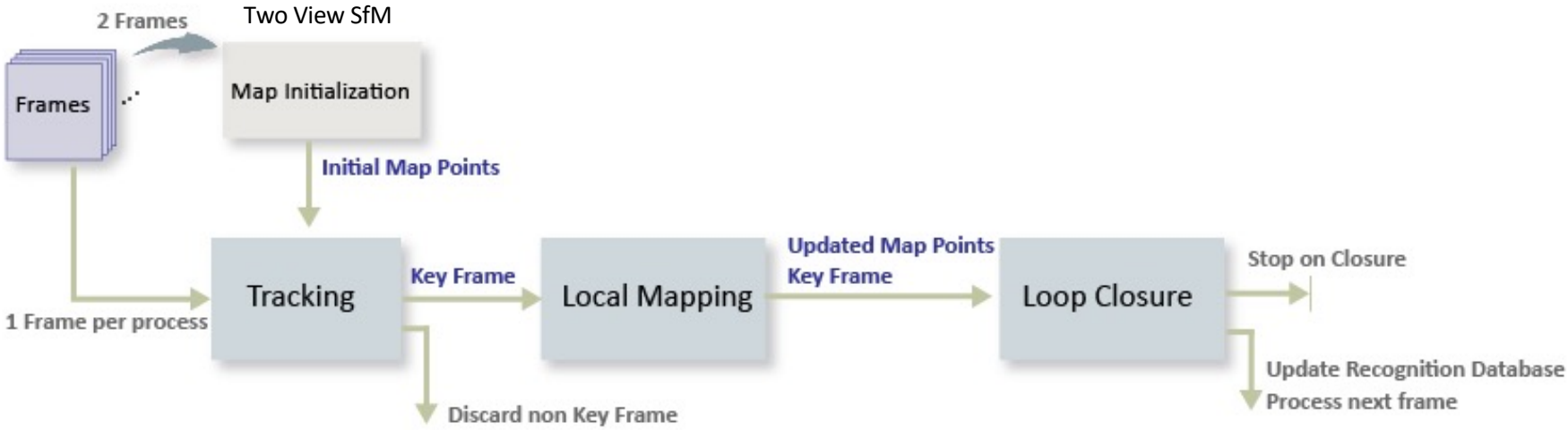


Figure 2. *Plot of* Opportunity*'s traverse history using visual odometry. (a) The drive in and around 20-m diameter Eagle Crater from Sols 1–70. (b) The drive in and around Endurance Crater from Sols 133–312. Units are in meters from the landing site origin, as measured on board the rovers. Red lines indicate directly commanded "blind" drives, green lines indicate autonomous hazard detection, and blue lines indicate visual odometry.*

Challenges

- Correspondence in tracking
  - Lack of texture
  - Lighting saturation
- Keeping a unique scale !
  - Reinitialization after a surface bump
- Drift (translation keeps being underestimated)
- Loop Closure

ORB-SLAM Pipeline

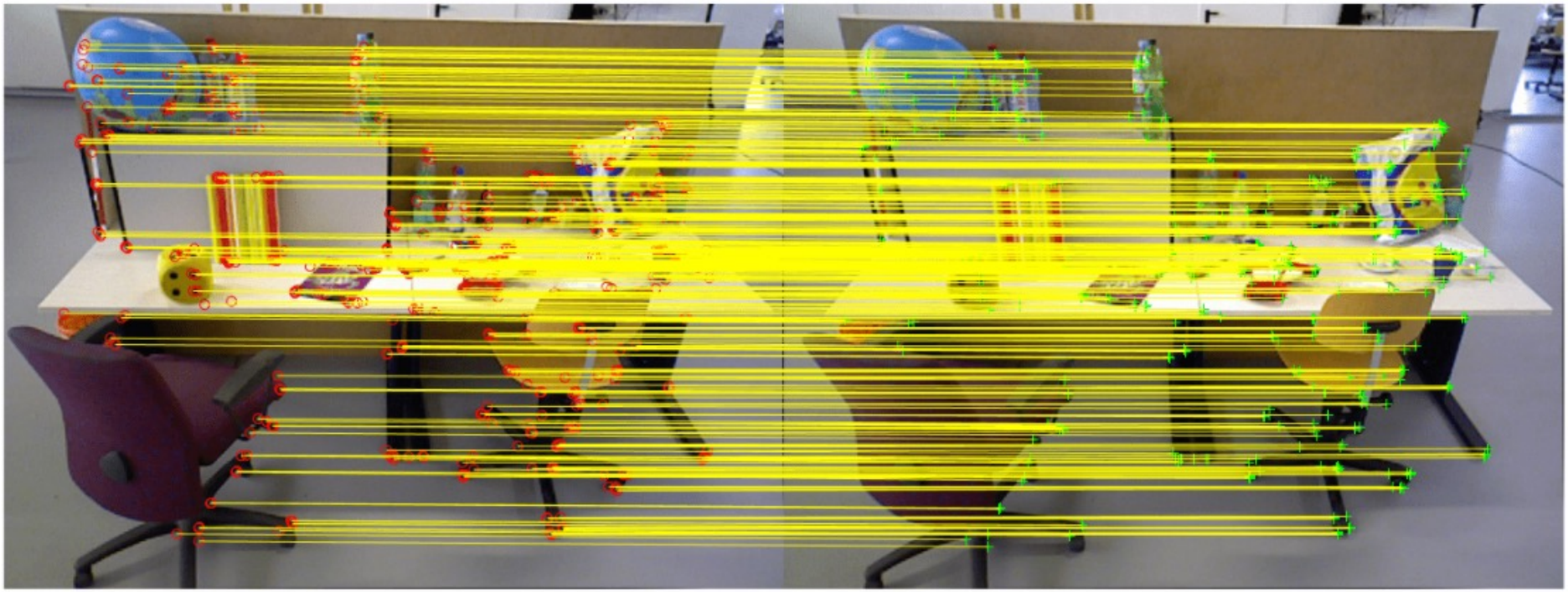


Key Frames: A subset of video frames that contain cues for localization and tracking. Two consecutive key frames usually involve sufficient visual change/long enough baseline.
Map Points: A list of 3-D points that represent the map of the environment reconstructed from the key frames.
Covisibility Graph: A graph consisting of key frame as nodes. Two key frames are connected by an edge if they share common map points. The weight of an edge is the number of shared map points.
Recognition Database: A database used to recognize whether a place has been visited in the past. The database stores the visual word-to-image mapping based on the input bag of features. It is used to search for an image that is visually similar to a query image.

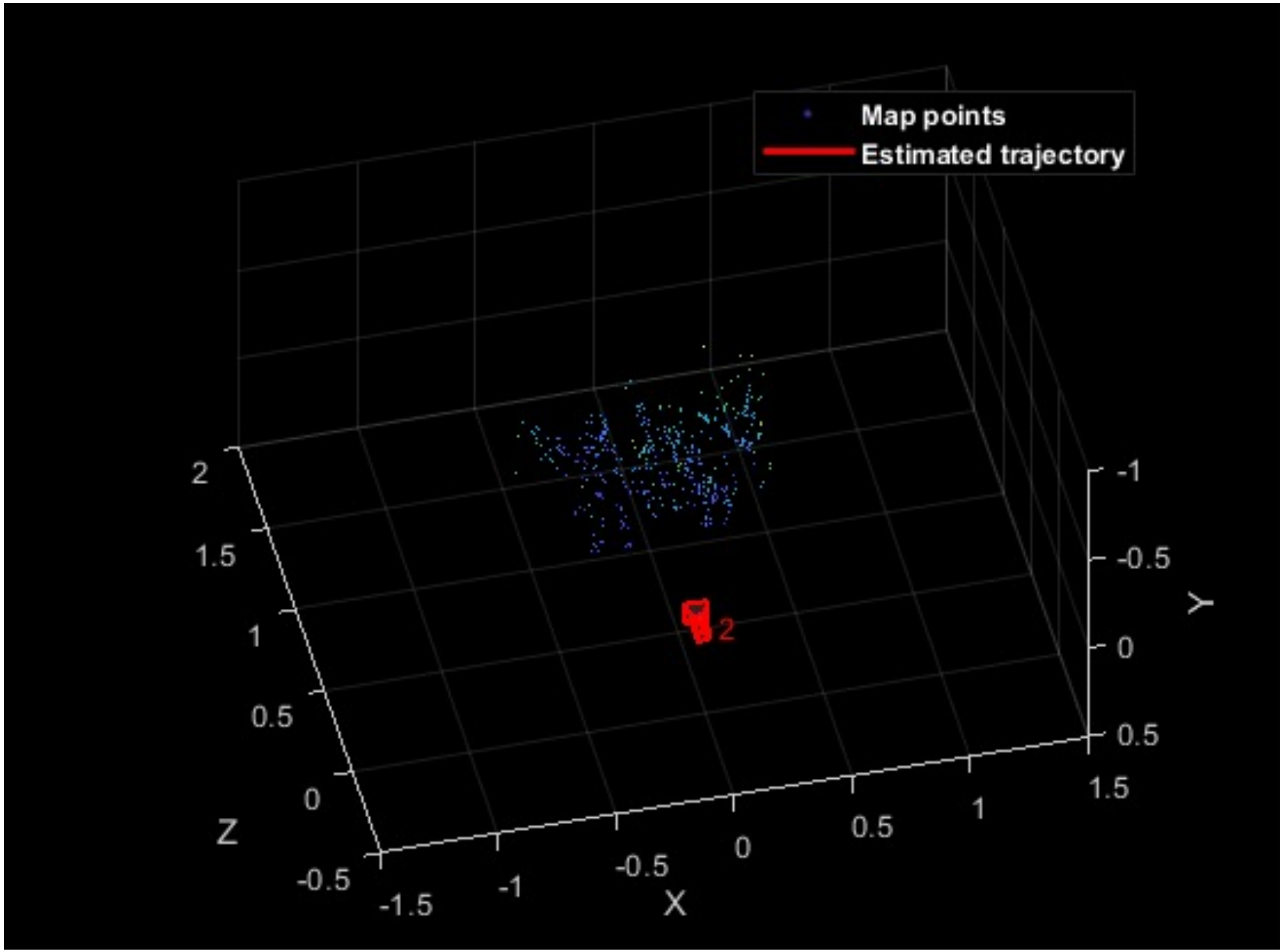
Map Initialization

- The devil is in the details
  - Deficient configurations for E-matrix
    - The scene is planar
    - There is no translation
  - Not sufficient inliers (90%)
- Keep comparing between 1st frame and k-th frame until none of the above
- Declare this last frame as 2nd keyframe
- Store feature and pose of keyframes
- Triangulate and store map-points.
- Refine initial reconstruction using 2-frame bundle adjustment

1st and 26th frame



Initial map

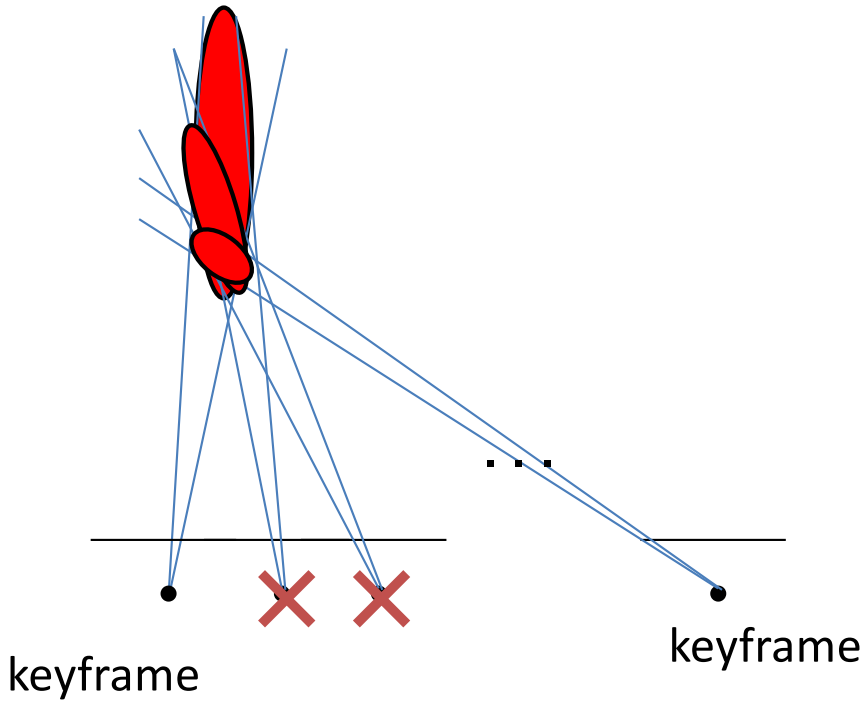

Map points
Estimated trajectory
2
1.5
1
0.5
0
-0.5
Z
-1.5
-1
-0.5
0
0.5
1
1.5
X
-1
-0.5
0
0.5
1.5
Y

Tracking

1. ORB features are extracted for each new frame and then matched with features in the last key frame that have known corresponding 3-D map points.
2. Estimate the camera pose wrt last keyframe with the Perspective-n-Point algorithm (RANSAC).
3. Given the camera pose, project the map points observed by the last key frame into the current frame and search for feature correspondences.
4. With 3-D to 2-D correspondence in the current frame, refine the camera pose by performing a pose-only bundle adjustment.
5. Project the local map points into the current frame to search for more feature correspondences and refine the camera pose again.
6. This frame is a key frame if both of the following conditions are satisfied:
    1. At least 20 frames have passed since the last key frame or the current frame tracks fewer than 100 map points
    2. The map points tracked by the current frame are fewer than 90% of points tracked by the reference key frame.

Alternative: Keyframe Selection

- Pose (translation) update depends on triangulated points whose error depends on baseline and distance.
- Wait until error in 3D triangulation decreases and then update pose: keyframe



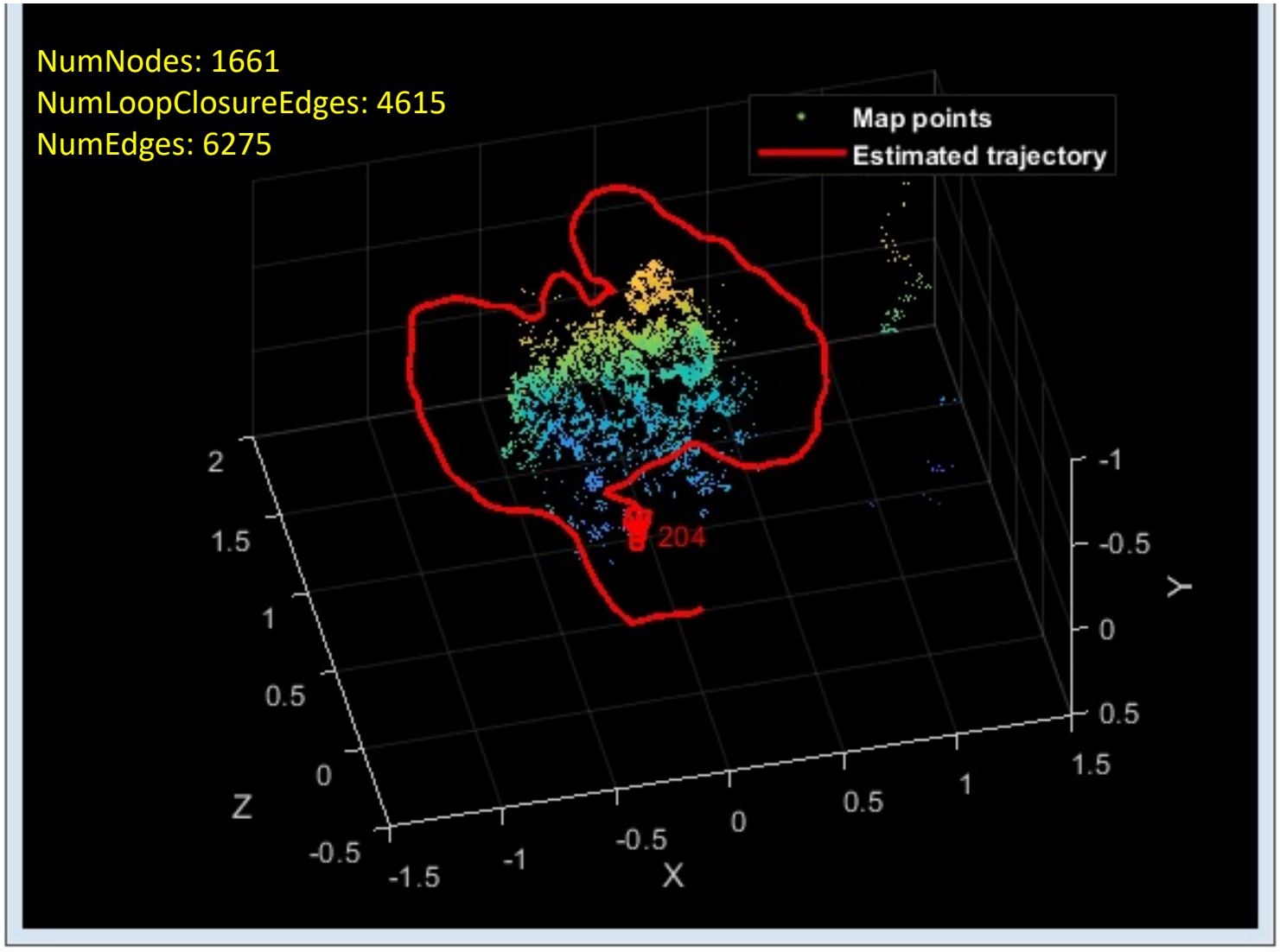
Local Mapping

- Upon keyframe is declared
  - Triangulate between this keyframe and last keyframe
  - Run bundle adjustment between them.

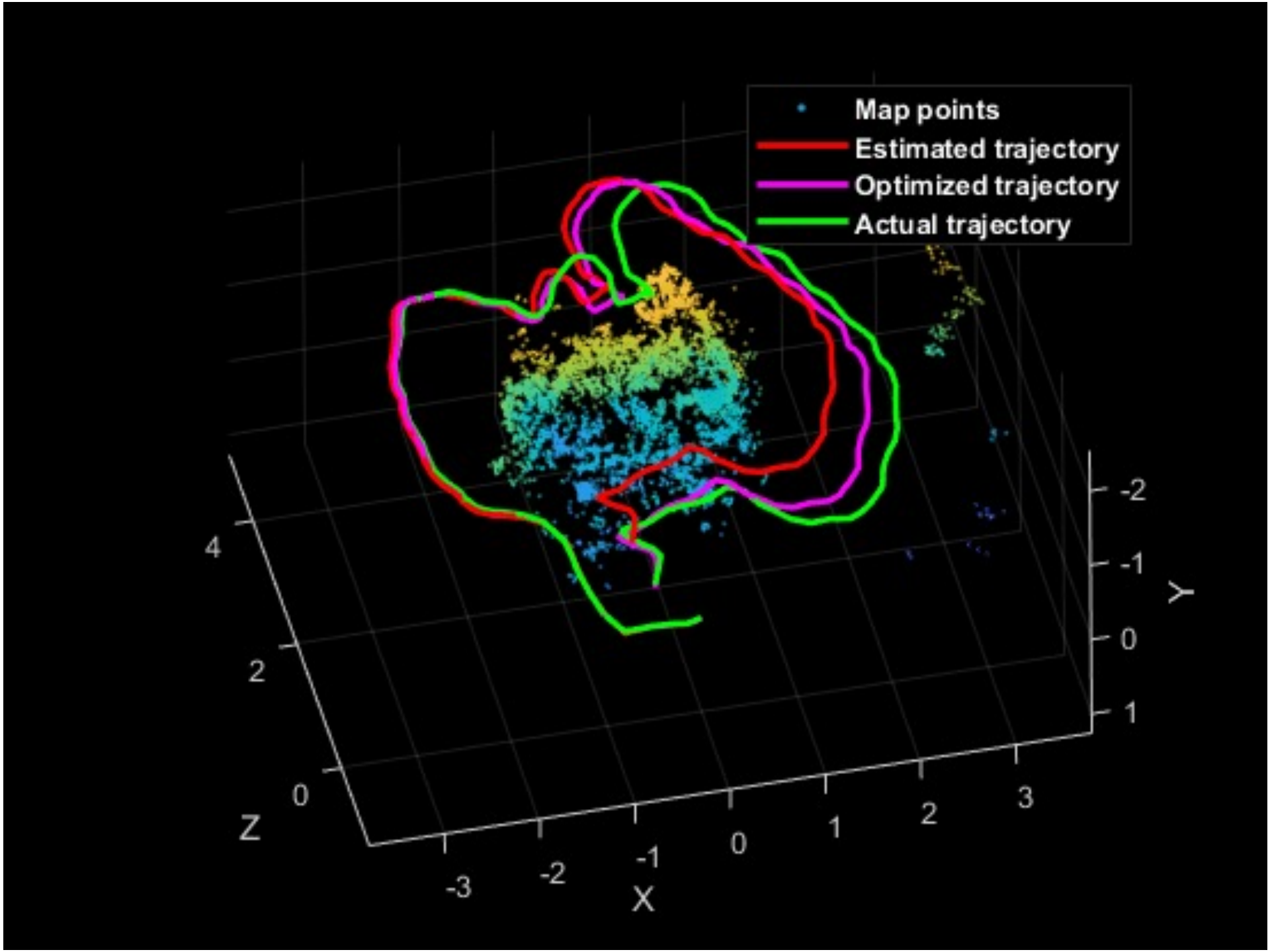
Loop Closure

- How do I know that I come back at a visited place?
  - Place recognition: search over all previous keyframes using “Bag of Visual Words”
  - If visual similarity is found we run a geometric consistency with RANSAC and 2-view SfM
- If loop closure declared
  - Run bundle adjustment (graph pose optimization g2o) to update all poses in essential graph (A subgraph of covisibility graph containing only edges with high weight, i.e. more shared map points).

Loop closure before optimization



Loop edge added between keyframe: 6 and 204


Map points
Estimated trajectory
Optimized trajectory
Actual trajectory
Z
4
2
0
X
-3
-2
-1
0
1
2
3
Y
-2
-1
0
1

Outliers in VO: without and with RANSAC