Simultaneous Localization and Mapping summary

J.M.M. Montiel (University of Zaragoza), J. Ramiro Martinez-de Dios (University of Seville), I. Pitas (Aristotle University of Thessaloniki)
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 3.0

Artificial Intelligence & Information Analysis Lab

Simultaneous Localization and Mapping

3D scene mapping and vehicle/sensor (primarily camera) localization:

- ***Mapping***: create or get 2D and/or 3D maps.
- ***Localization***: find the 3D location based on sensors.
- ***Simultaneous Localization and Mapping (SLAM)***.
- Information fusion in localization and mapping.

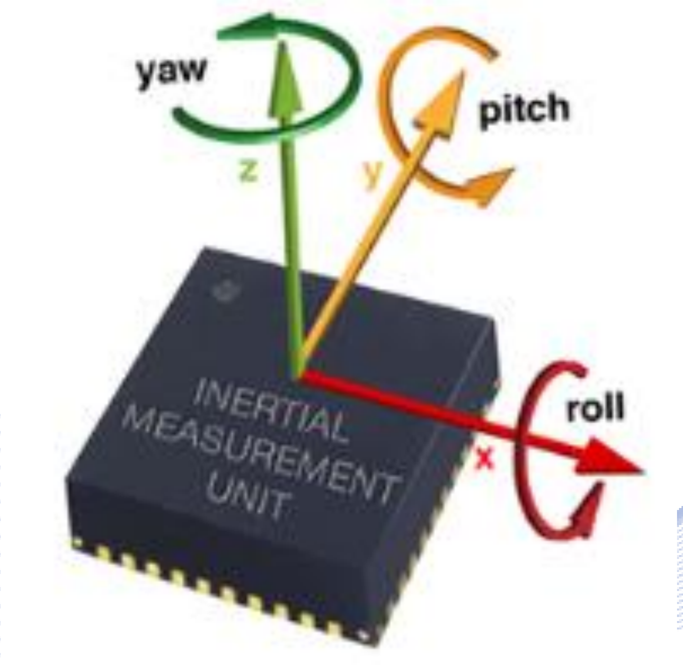
Simultaneous Localization and Mapping

- **Sensors**
- Mapping
- Localization
- SLAM
- Data fusion in drone localization.

Sensors

On-vehicle Sensors:

- Lidar
- Monocular camera
- IMU
- laser altimeter
- RTK D-GPS.



3D Localization and Mapping

- Sensors
- **Mapping**
- Localization
- SLAM
- Data fusion in drone localization.

Mapping techniques

- To obtain a map, robots/drones need to take measurements that allow them to perceive their environment.
- Relationship between the inner state of the robot (i.e., 6-DoF pose) and the state of the map (e.g., the position of features in a map).
- Two main approaches:
  - ***Odometry-based methods.***
  - ***Simultaneous Localization and Mapping (SLAM).***

Visual odometry

Structure from Motion (SfM) recovers:

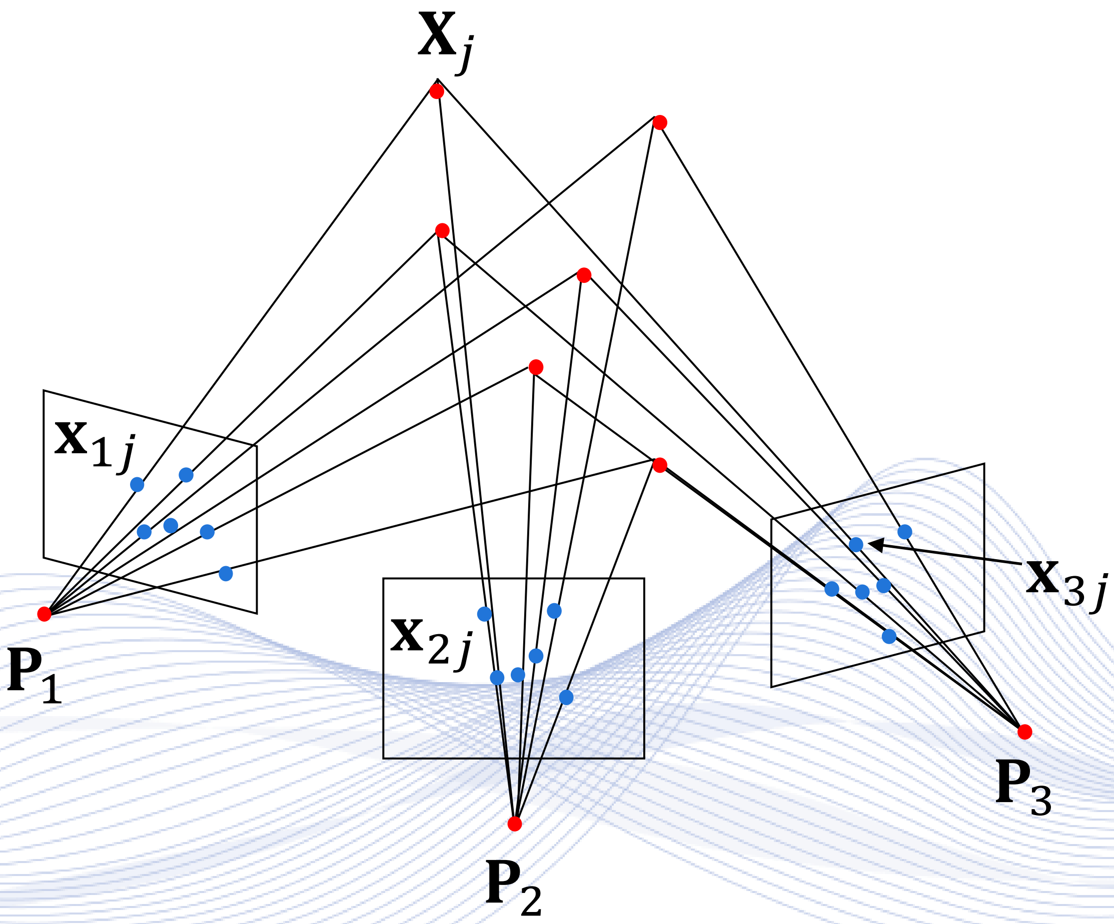
a) the relative camera poses and

b) the three-dimensional (3D) scene structure,

from a set of camera images (calibrated or noncalibrated).

- Visual Odometry is a particular case of SfM.
- Focuses on estimating the 3D motion of the camera sequentially, as a new frame arrives, in real time.

Visual odometry

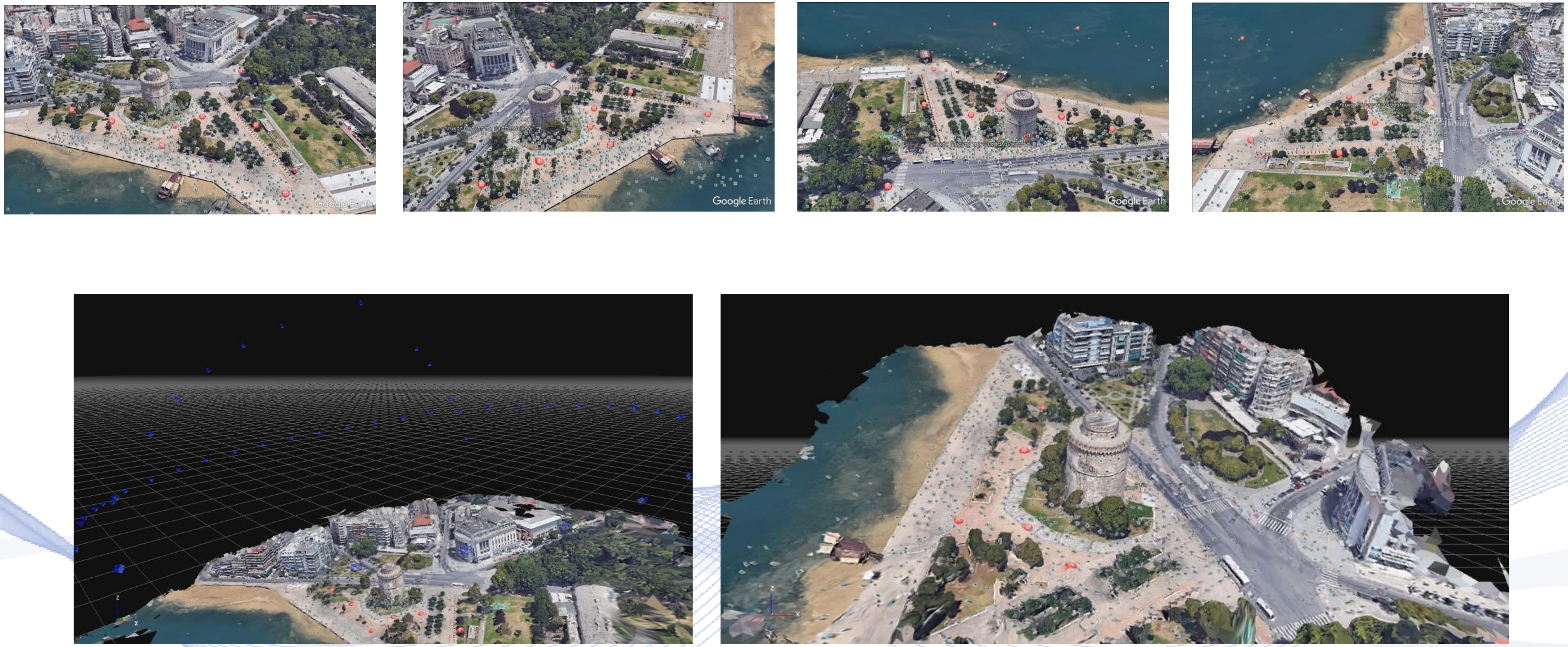


Structure from Motion

Structure from Motion (***SfM***):

- Unknown camera location/orientation.
- Cameras can be fully, partially or non-calibrated.
- Unknown feature correspondences across views.
- Computation up to scale factor:
  - Camera location.
  - 3D location of the matched feature points.

Structure from Motion



3D models reconstructed using 50 images from Google Earth.

Visual odometry based methods for mapping

- Usually fast, efficient, simple to implement methods.
- Accurate enough, if the proposed sensor does not induce drifts, due to noise or non-linearities.

Visual odometry based methods for mapping

Lidar Odometry and Mapping (LOAM):

- It is an odometry estimation and mapping method that calculates the trajectory of the laser, using high-level features based on the properties of rotatory lasers.

- It identifies as data features both corner and surface points.
- It generates a map that contains both of them separately.

Geometrical mapping

3D LIDAR:

- SLAM-like algorithm based on Prediction-Update Recursions.
- Extract from the LIDAR measurements: corner and surface points.
- ***Prediction***: Estimate LIDAR-based odometry from different scans using the ICP algorithm.
- ***Update***: Matching of the LIDAR scan with the estimated map.
- Good estimate of robot 6 DoF pose and geometrical map.

Geometrical mapping

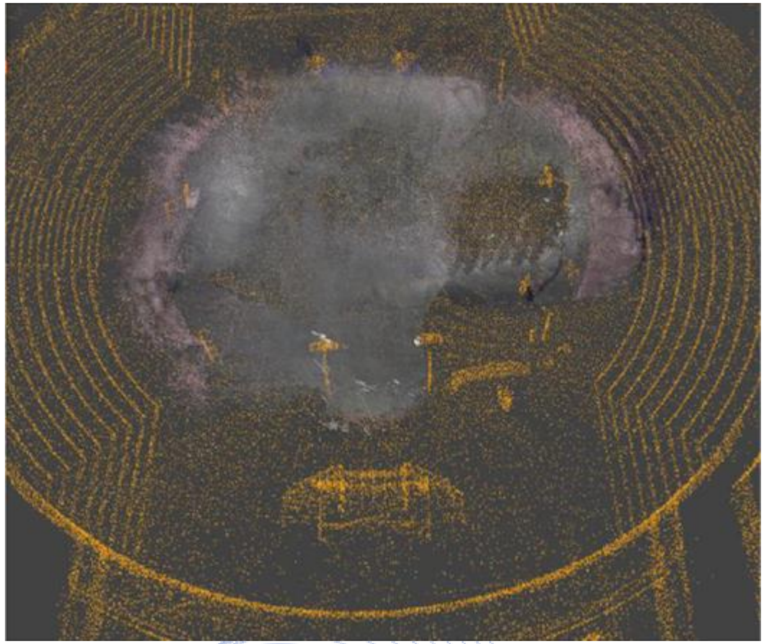
Visual camera:

- Extraction of features using detectors such as SURF, SIFT or ORB.
- Estimation of visual odometry.

- Robot odometry is a combination of:
  - LIDAR-based odometry.
  - Visual odometry.
  - IMU.

Geometrical mapping

Experiments

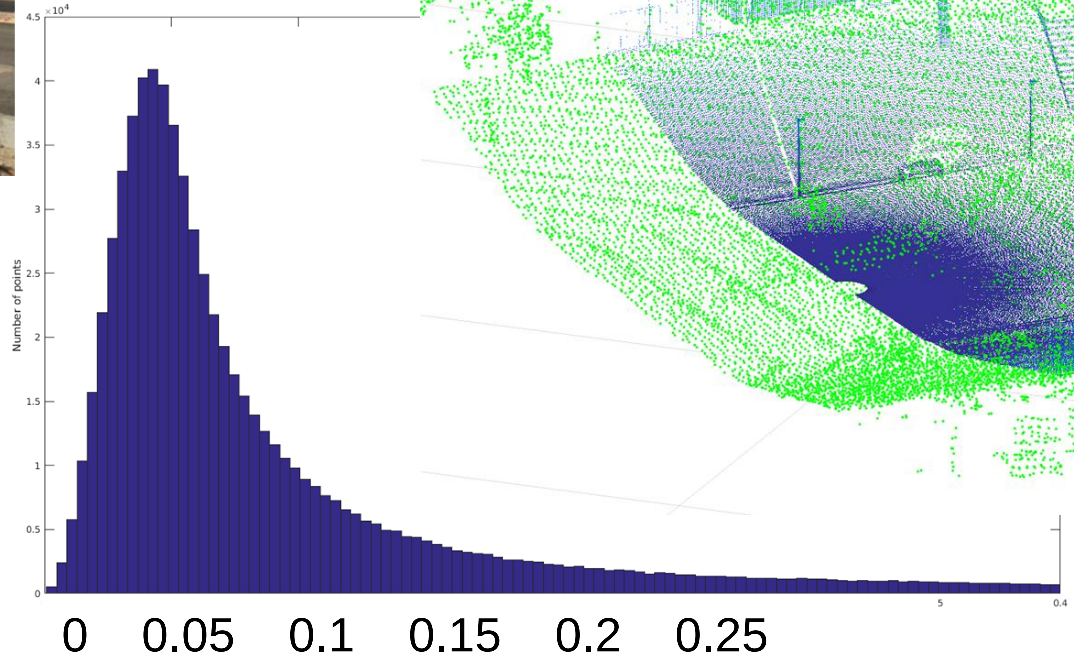


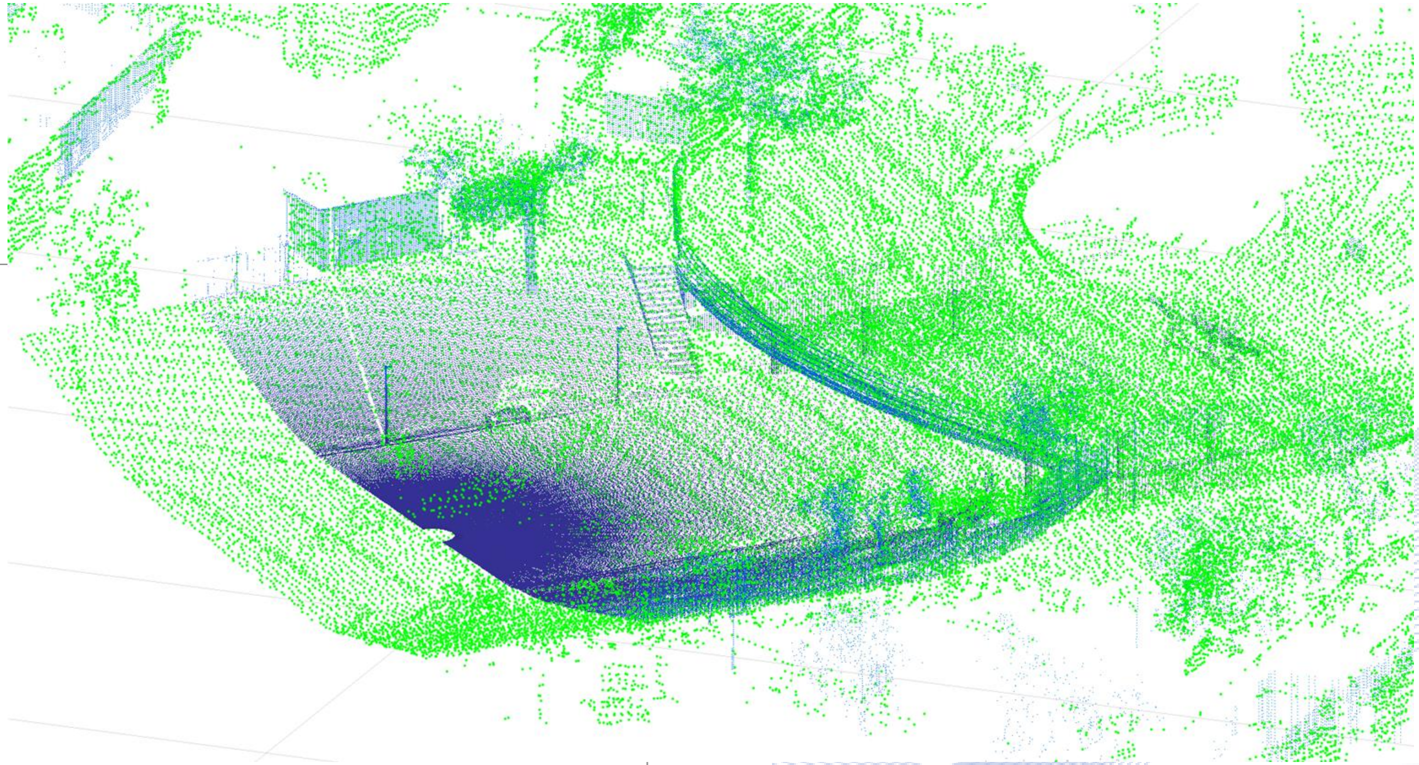
Repeatibility

Dataset	Mean Error (m)	Median Error (m)	Min Error (m)
1	0,1377	0,1073	0,00098
2	0,1053	0,0769	0,00045
3	0,0847	0,0578	0,00083
4	0,1074	0,0792	0,00078
5	0,1722	0,1560	0,00130

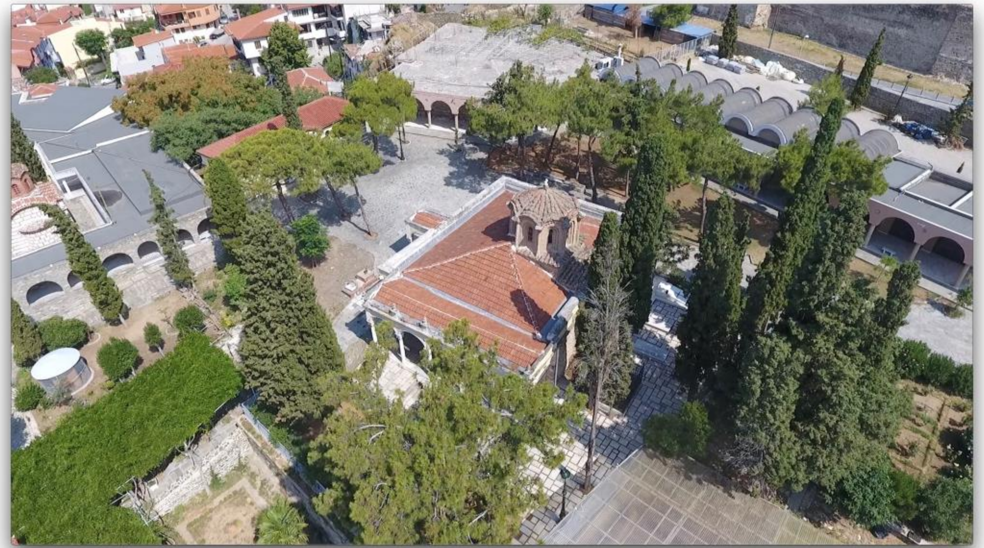
Geometrical mapping

Validation with a TOTAL STATION





3D Scene monument from multiple uncalibrated images



3D Scene mapping from multiple uncalibrated images: Vlatadon Monastery, Thessaloniki, Greece

Artificial Intelligence & Information Analysis Lab

3D Localization and Mapping

- Sensors
- Mapping
- **Localization**
- SLAM
- Data fusion in drone localization.

Localization

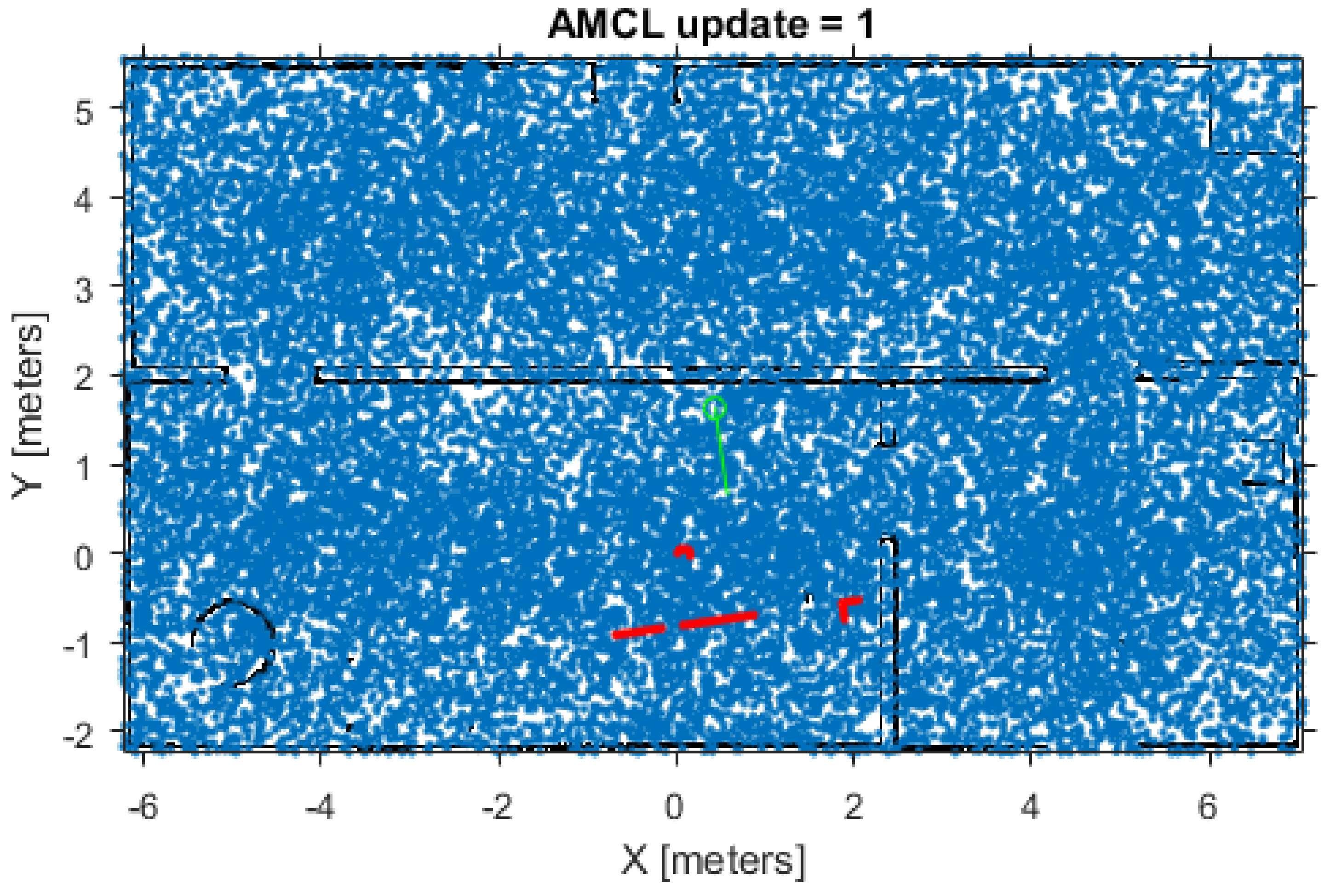
- Not an easy task.
- ***Unconstrained nature of robot/drone movements*** → use of high-fidelity algorithms and sensors that do not reply on them.
- Many methods used for Mapping, are also used for Localization.
- Localization methods can be used as an alternative, in case of GPS failure.

The “kidnapped” problem

Two different robot localization cases are possible:

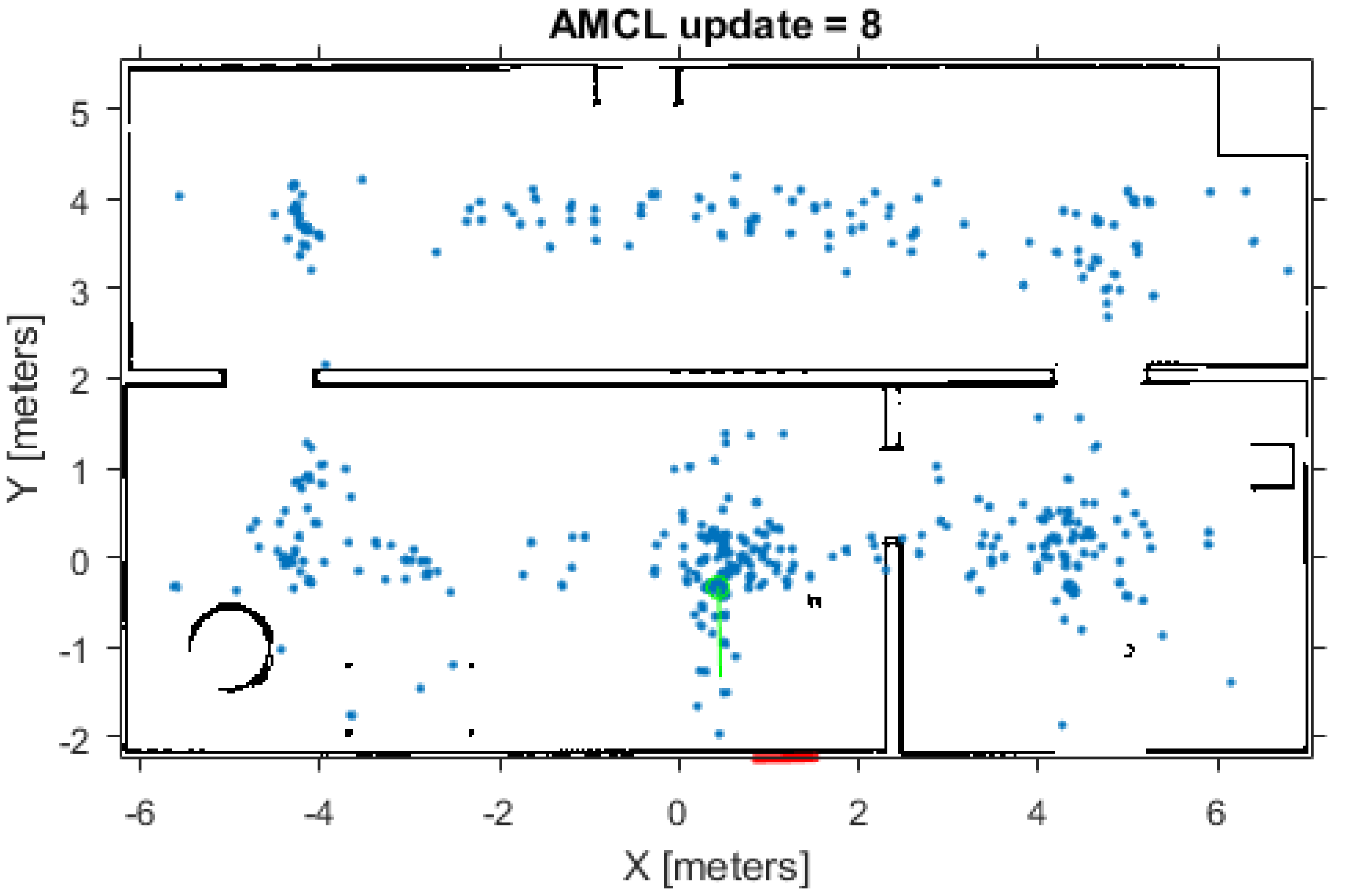
- Its initial position is known (with a respect to global map).
- The ***kidnapped robot problem*** (***KRP***):
  - An autonomous robot is carried (kidnapped) to an arbitrary location without being told.
  - KRP differs from global mobile robot localization as the mobile robot believes to be in a different position at kidnapping.
- Solution → ***Adaptive Monte Carlo localization (AMCL)*** algorithm.

AMCL algorithm



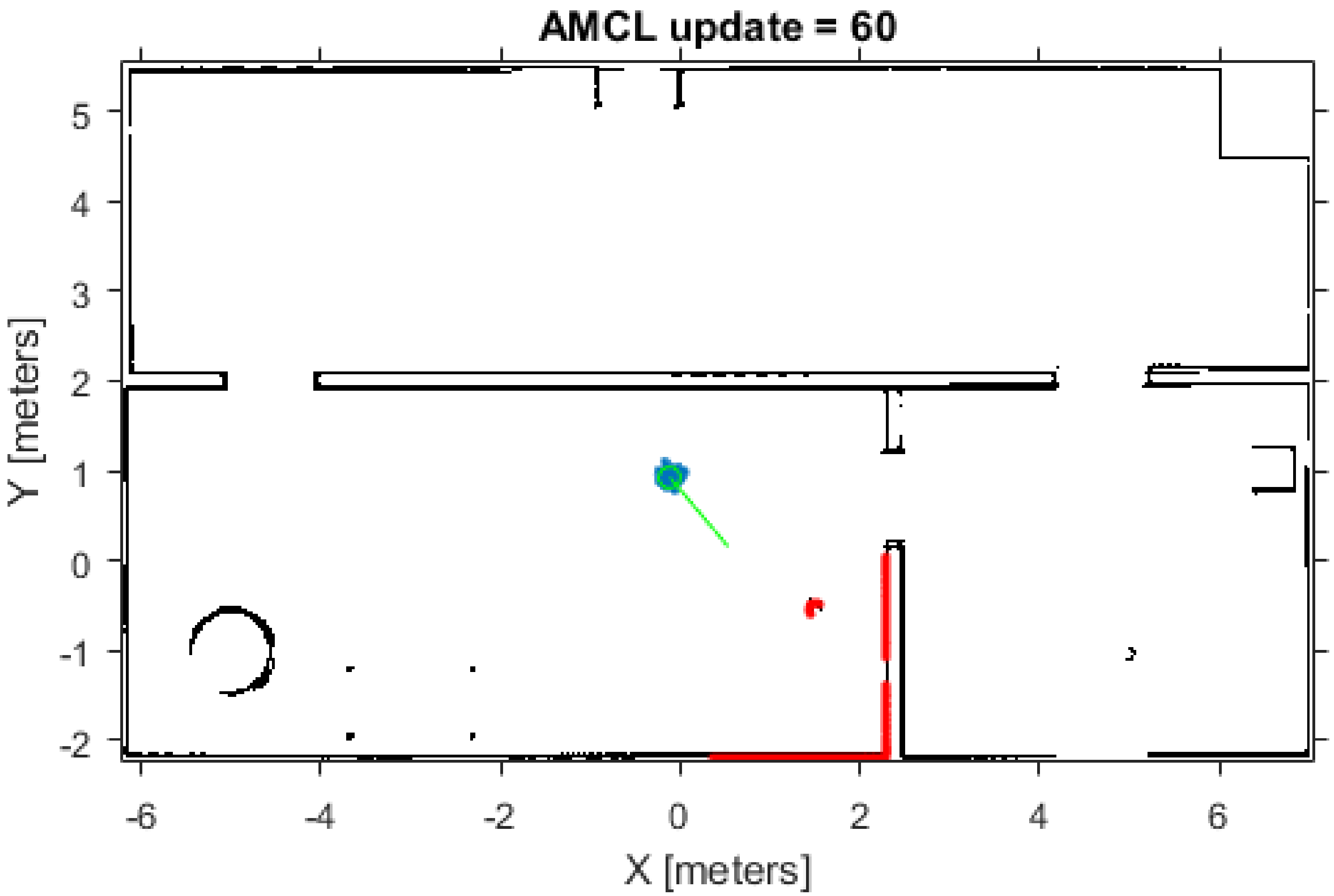
https://www.mathworks.com/help/search.html?qdoc=amcl&submitsearch=

AMCL algorithm



https://www.mathworks.com/help/search.html?qdoc=amcl&submitsearch=

AMCL algorithm



https://www.mathworks.com/help/search.html?qdoc=amcl&submitsearch=

3D Localization and Mapping

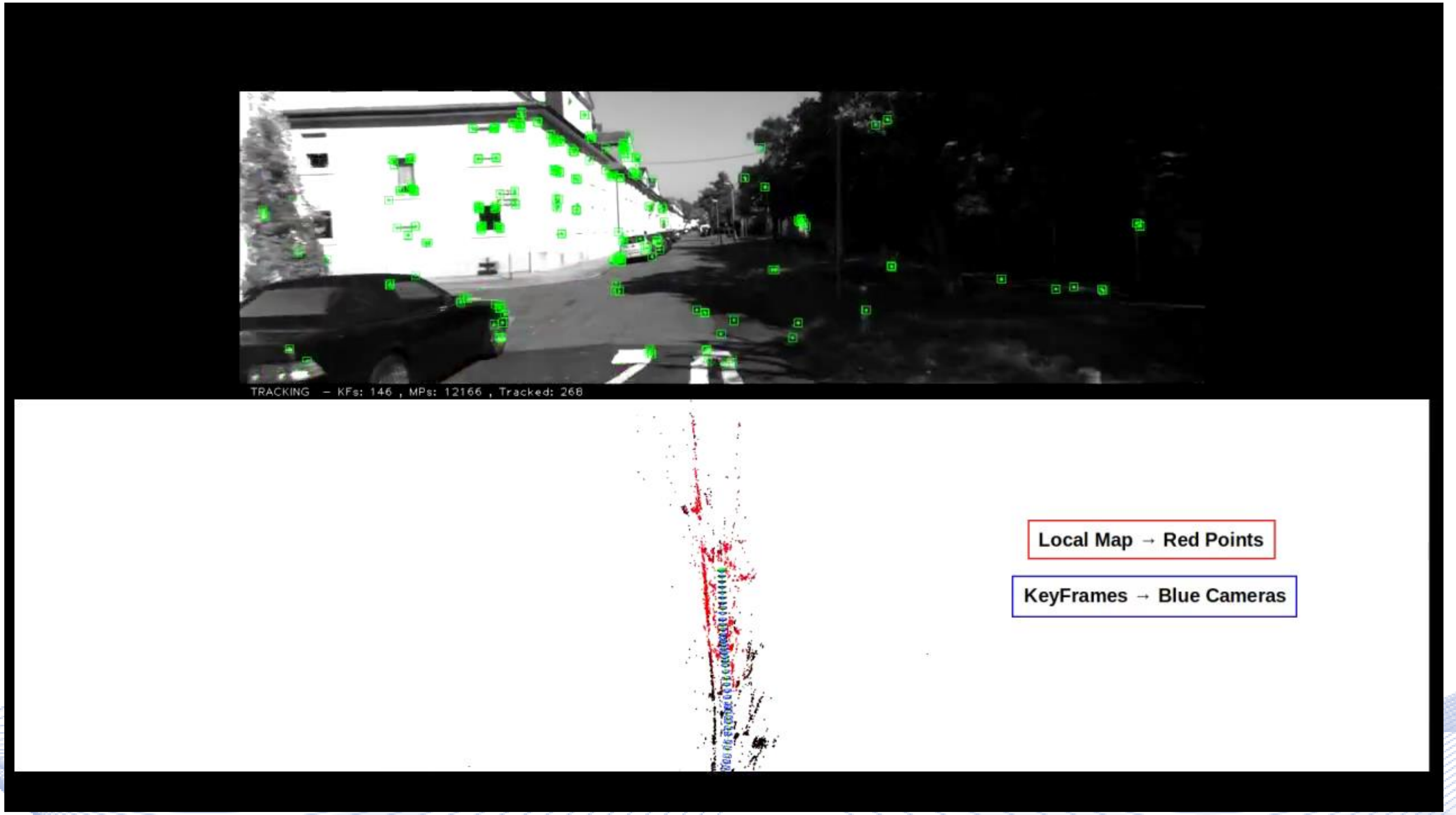
- Sensors
- Mapping
- Localization
- **SLAM**
- Data fusion in drone localization.

Methods for mapping based on SLAM

- Robot/drone uses the map, in order to update the robot inner state.

- Both the robot/drone and the map share information and are updated together.

Visual SLAM



https://youtu.be/sr9H3ZsZCzc

Visual SLAM

- From the sole input of the video stream:
  - ***Simultaneous estimation of the camera motion and the 3D scene***.
  - Real-time at frame rate.
  - Sequential processing.
  - ***The field of view of the camera*** ≪ ***than the map size***.

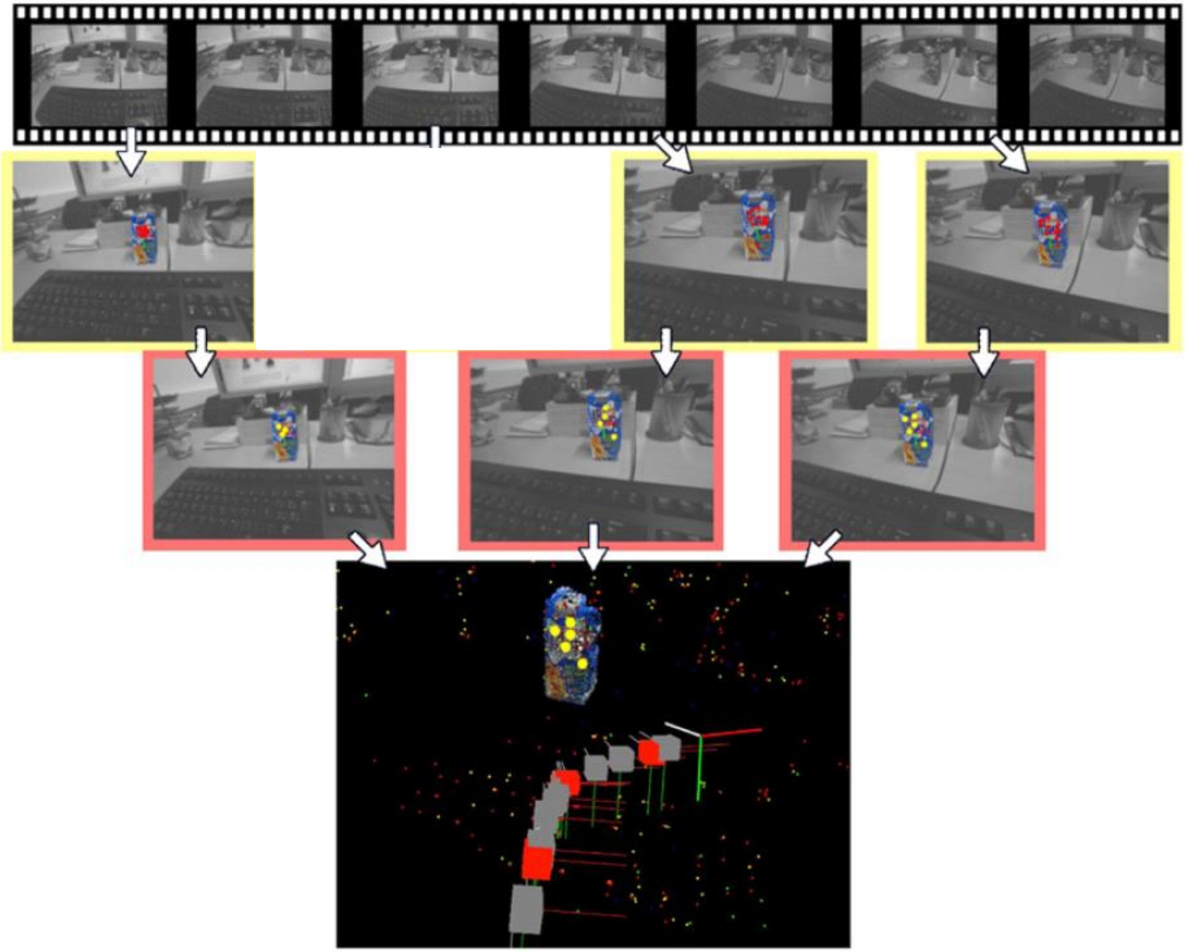
SLAM methods

- LSD SLAM:
  - Uses a ***randomly initialized scene depth*** from the first viewpoint that is later refined through measurements across subsequent frames. This method does not suffer from the degenerative cases of geometry methods.
- HECTOR SLAM:
  - A ***grid-based SLAM method*** that employs gradient-based optimization algorithms to watch each scan with the overall map. It is widely used for 2D mapping.

ORB-SLAM

- Among top performers in sparse features VSLAM.
- ***Robust, real-time, large scale operation***.
- Able to operate in general scenes.
- Prototype ORB-SLAM system ready to use.

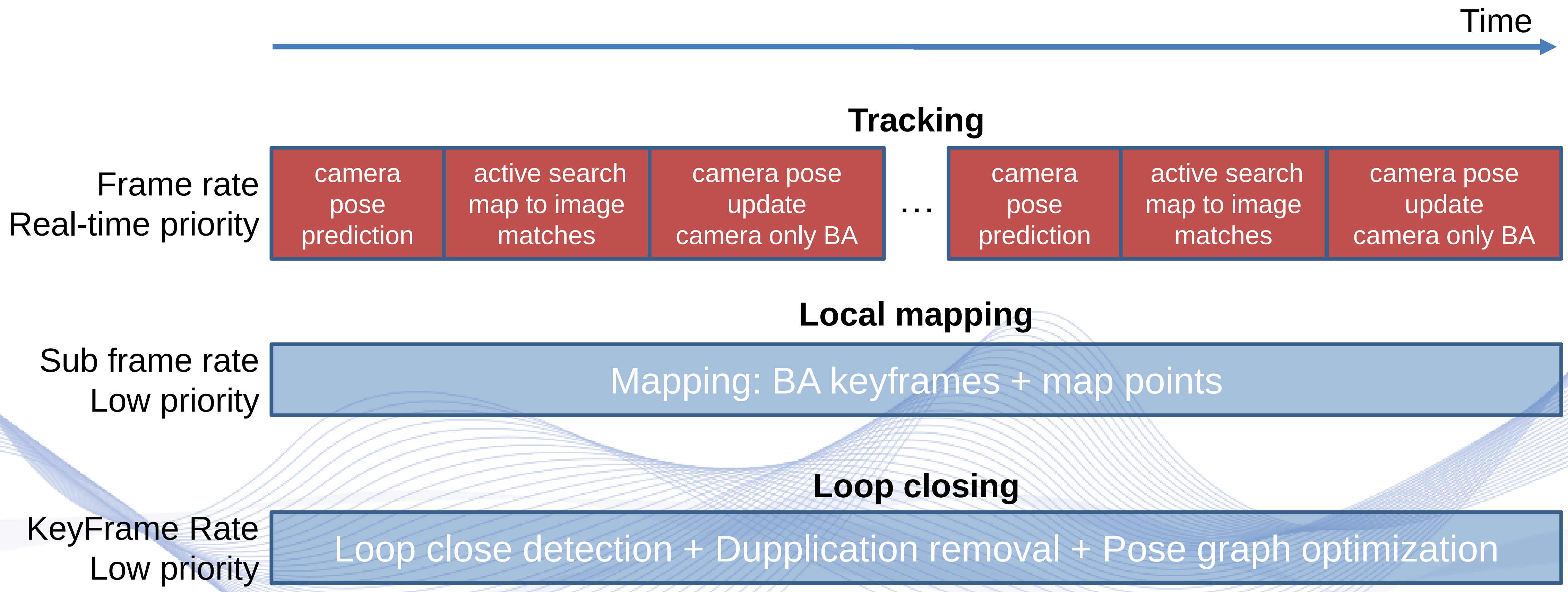
Frame / keyframe



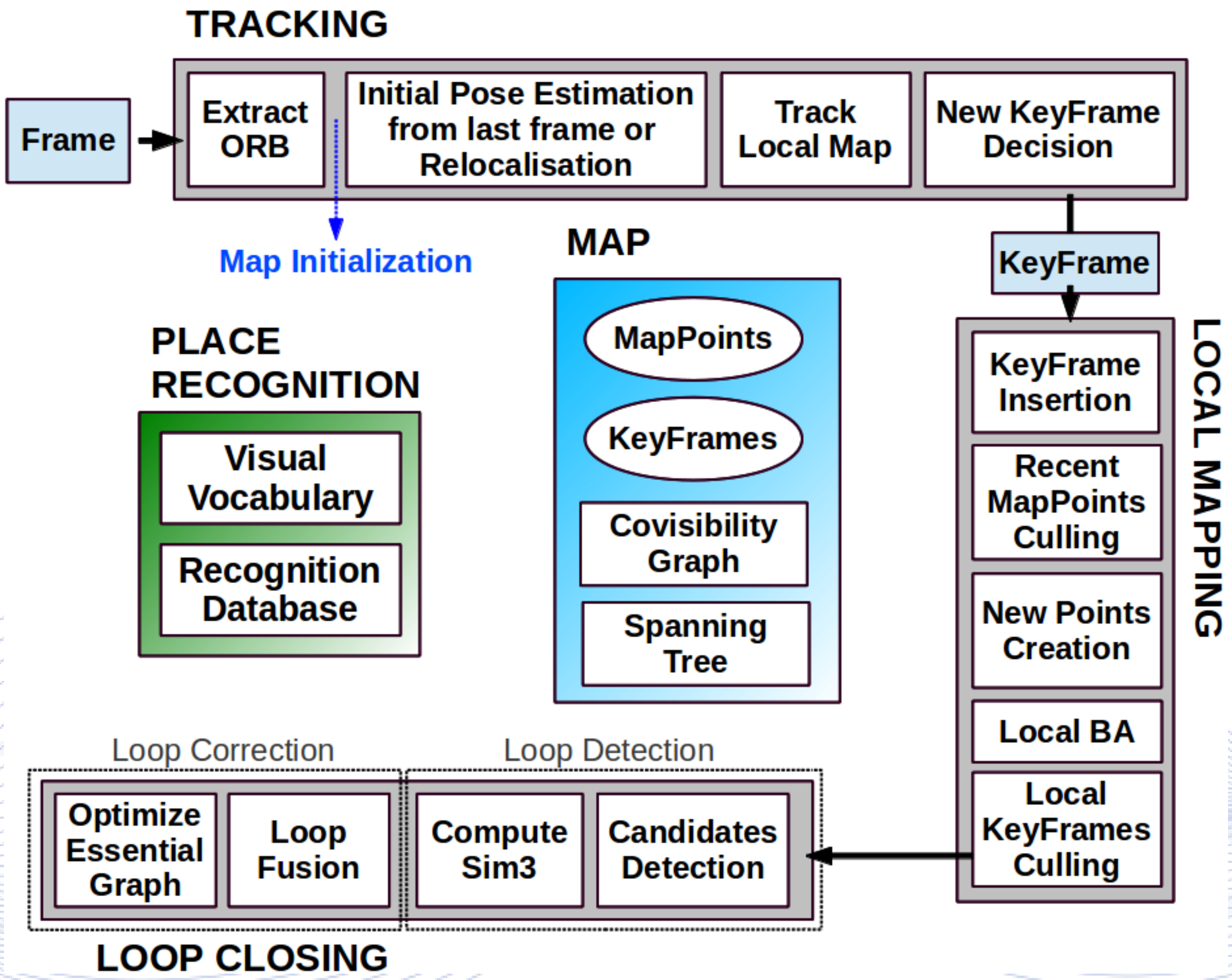
G. Klein and D. Murray . Parallel tracking and mapping for small AR workspaces. (ISMAR), November 2007

- Local Map.
- Keyframe-based systems:
- Localization and Mapping are separated into two steps:
- Full BundleAdjustment:
  - KeyFrames and map points.
- Regular video frames:
  - Only camera pose is computed.

Three threads running in parallel



ORB-SLAM system overview



ORB-SLAM system overview

- Full system including all stages of a typical VSLAM:
  - Tracking at frame rate.
  - Mapping, subframe rate.
  - ***Loop closing.***
  - Relocation.
  - FAST corner + ORB descriptor.
  - Binary descriptor.
  - Fast to compute and compare.

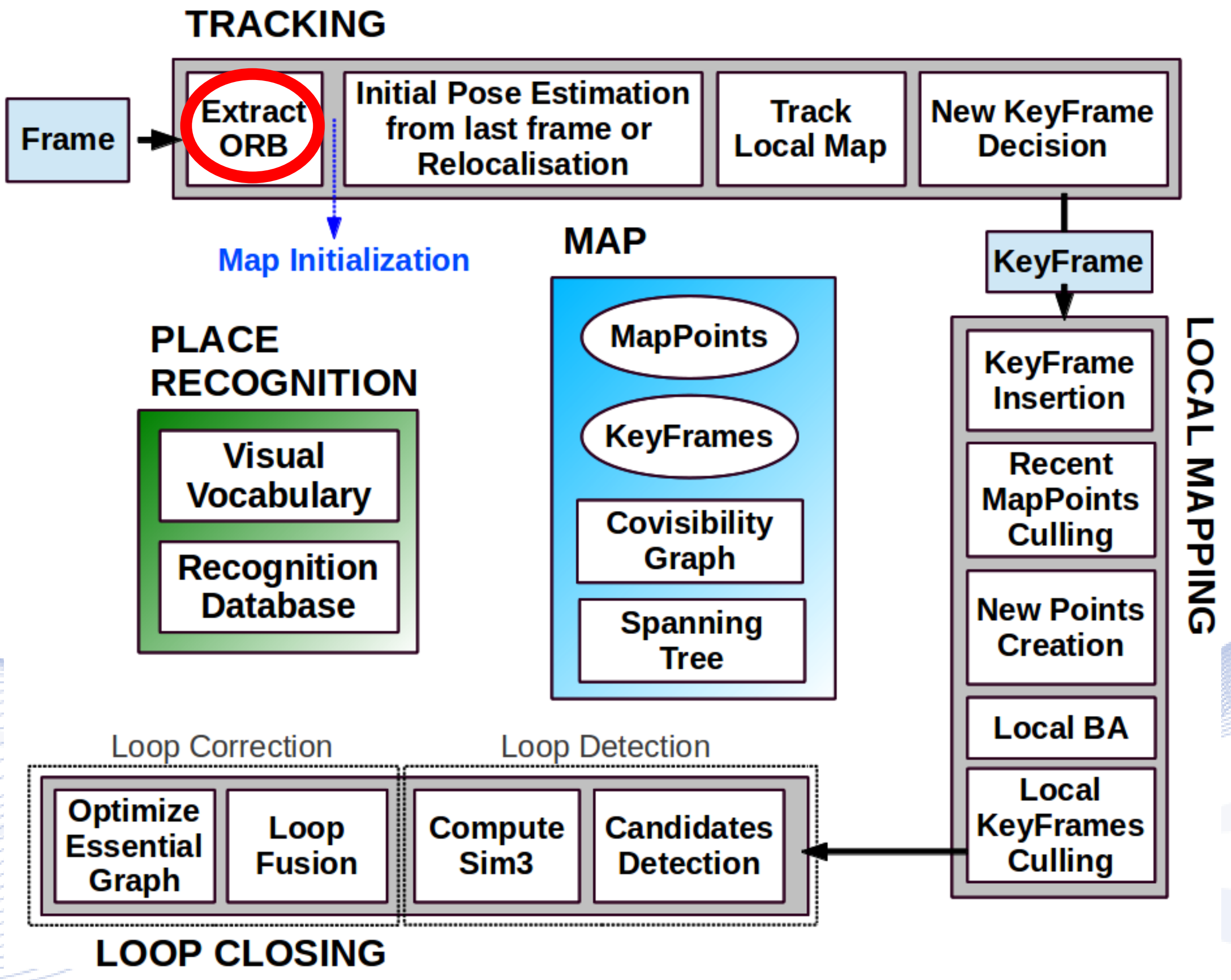
ORB-SLAM system overview

- Same feature all stages.
- ***Survival of the fittest policy for points and key-frames management***.
- Three thread architecture.
- All stages end up providing an accurate initial guess to non-linear re-projection error optimization:

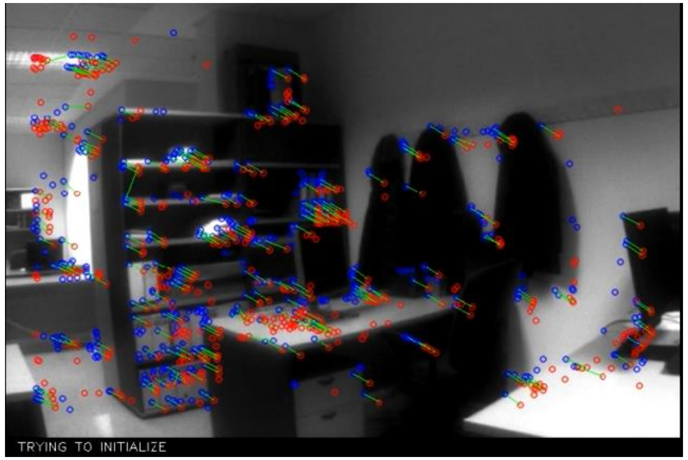
$$\underset{\mathbf{T}_{iw},\mathbf{X}_{wj}}{\operatorname{argmin}} \sum_{i,j} \rho\left(\left\|\mathbf{x}_{ij} - \pi_i(\mathbf{T}_{iw}, \mathbf{X}_{wj})\right\|^2\right).$$

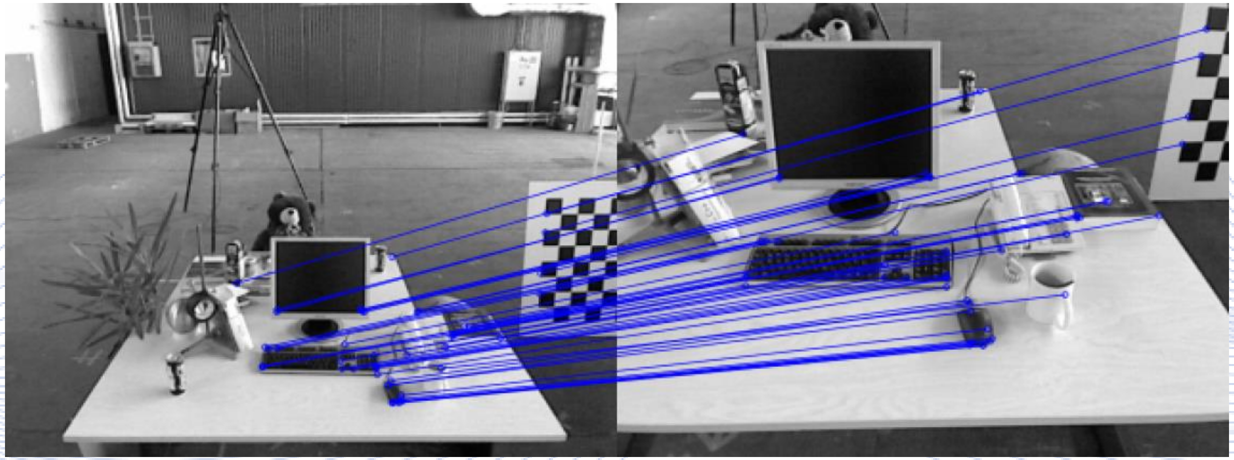
Features

- Repeatability.
- Accuracy.
- Invariance:
  - Illumination
  - Position
  - In-plane rotation
  - Viewpoint
  - Scale.
- Efficiency.



Features

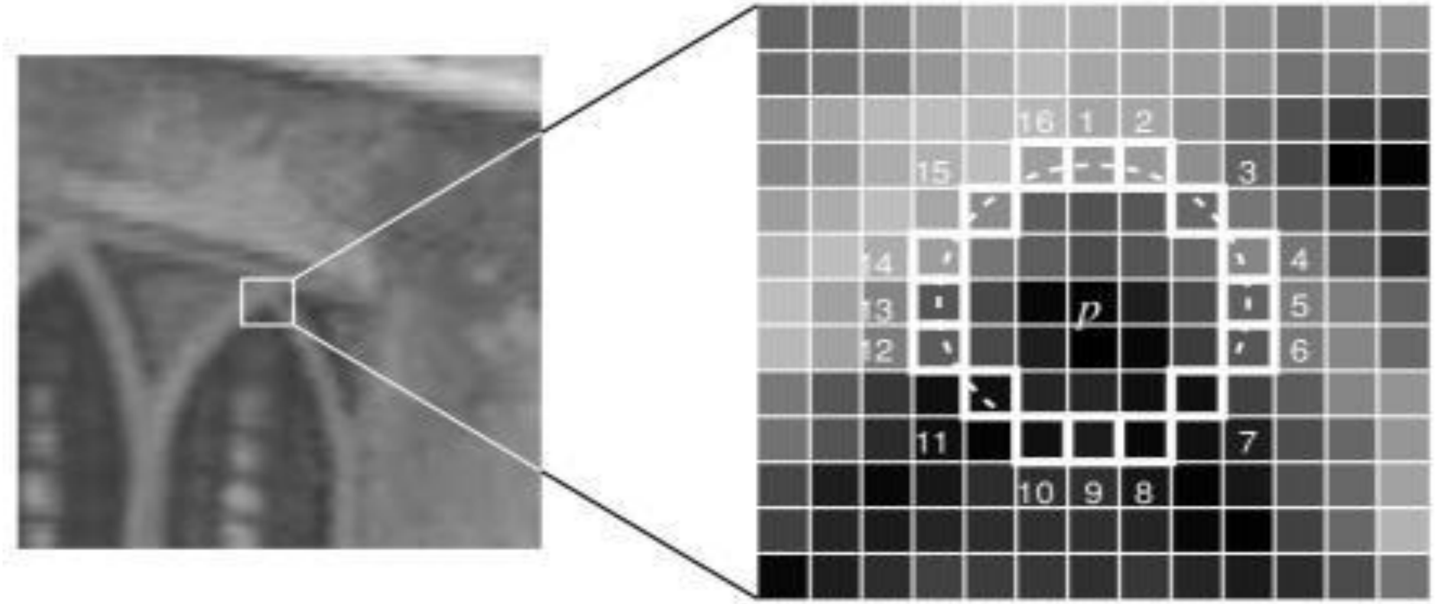




Popular features for visual SLAM

- **ORB: Oriented FAST and Rotated Brief.**
- 256-bit binary descriptor.
- Fast to extract and match (Hamming distance).
- Good for tracking, relocation and Loop detection.
- Multi-scale detection at same point appears on several scales.

FAST corner detector

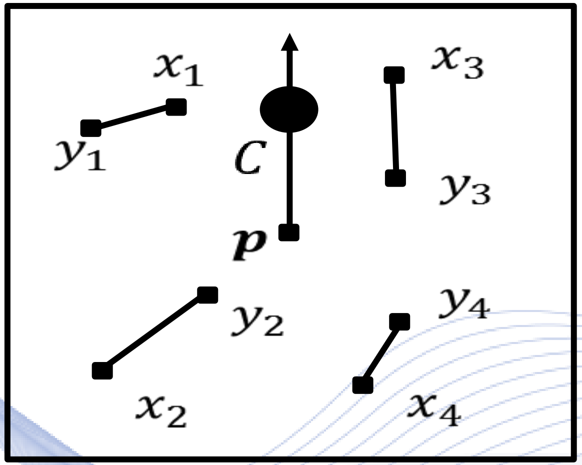


- ***Pixel* p *surrounded by consecutive pixels all brighter/darker than* p.**
- Much faster than other detectors.

Binary descriptors: rBRIEF

- Computed around a FAST corner.
- Has orientation.

$$D_i(\boldsymbol{p}) = \begin{cases} 1 & \text{if } I(\boldsymbol{p}+\boldsymbol{x_i}) < I(\boldsymbol{p}+\boldsymbol{y_i}) \\ 0 & \text{otherwise} \end{cases}$$



$\boldsymbol{p} \triangleq$ interest point
$C \triangleq$ intensity centroid

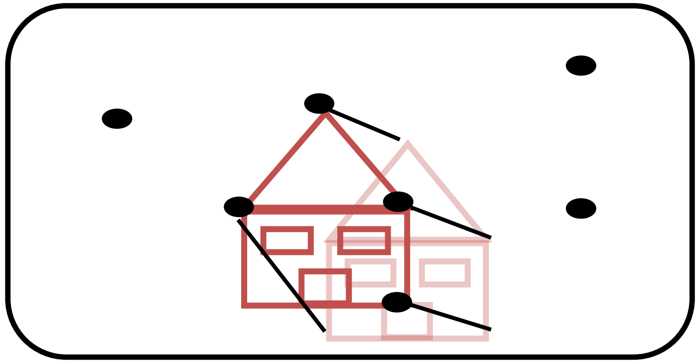
Hamming distance 5

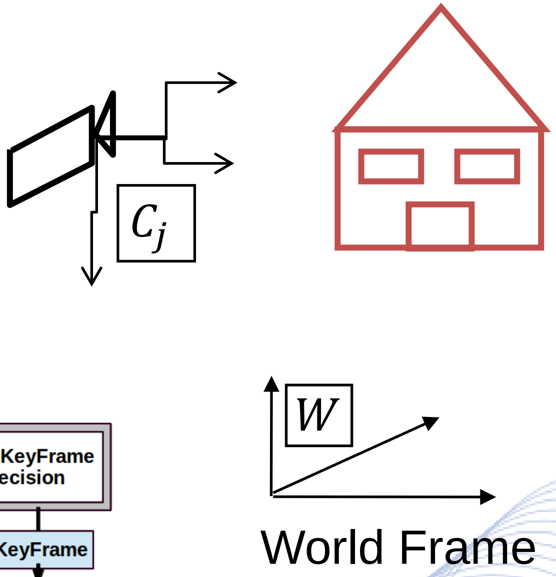
Hamming distance 51

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. ORB: an efficient alternative to SIFT or SURF, ICCV 2011

Artificial Intelligence & Information Analysis Lab

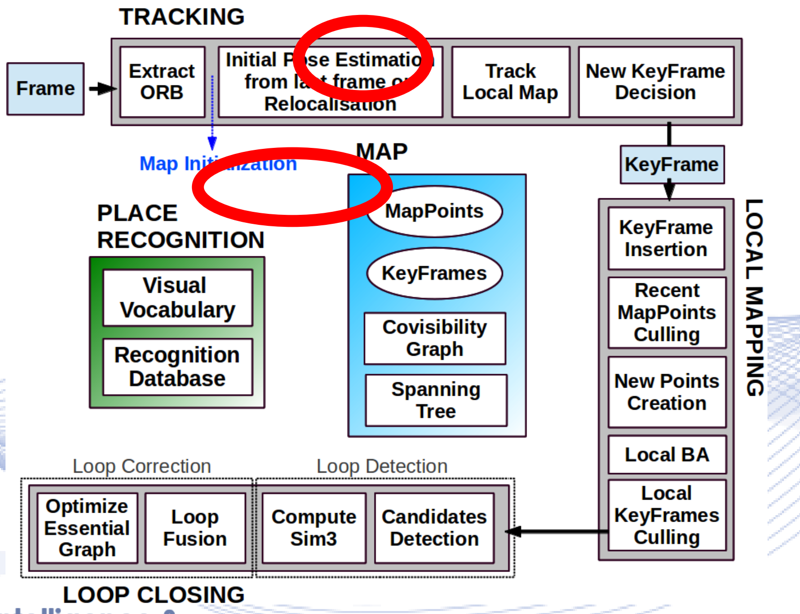
Camera tracking





- Map assumed perfectly estimated.
- ORB point detection in the image.
- Camera assumed close to its last pose, predicted by motion model.
- Map back-projected in the image.
- Putative matches ORB similarity.
- **Camera pose optimization $\mathbf{T}_{iw}$ fixing all map points $\mathbf{X}_{wj}$:**

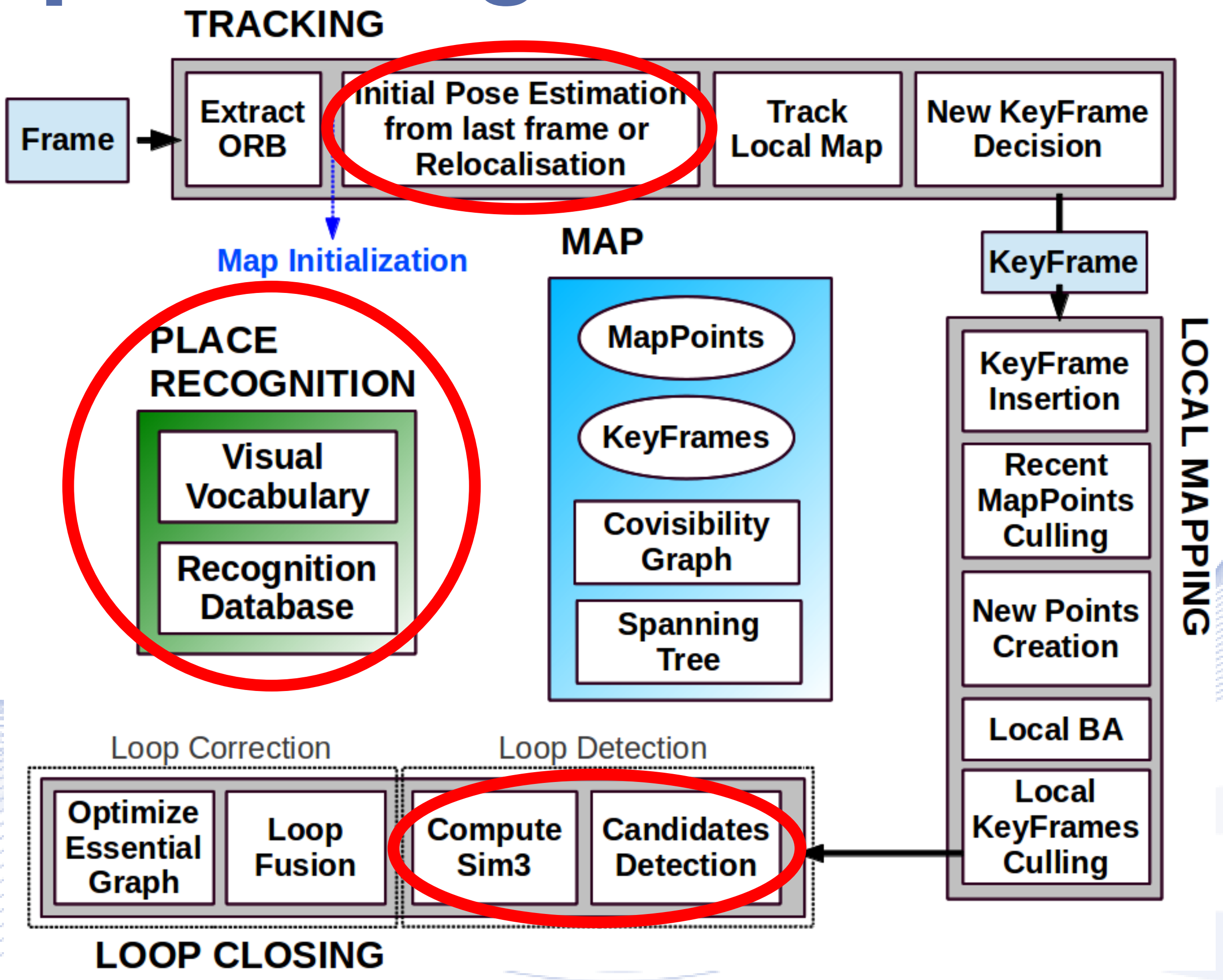
$$\underset{\mathbf{T}_{iw}}{\operatorname{argmin}} \sum_j \rho\left(\left\|\mathbf{x}_{ij} - \pi_i(\mathbf{T}_{iw}, \mathbf{X}_{wj})\right\|^2\right)$$



Place recognition: Relocation / Loop closing

Loop closing to avoid map duplication:

- Loop detection.
- Loop correction: correct accumulated drift.



Why is place recognition difficult

YES

Likely algorithm answer:

YES

TRUE POSITIVE

Artificial Intelligence & Information Analysis Lab

Why is place recognition difficult

Likely algorithm answer:

NO	**NO** ☐	**TRUE NEGATIVE** ☐
NO	**YES** ☐	**FALSE POSITIVE** ☐

Why is place recognition difficult

Likely algorithm answer:

NO **YES** **FALSE POSITIVE**

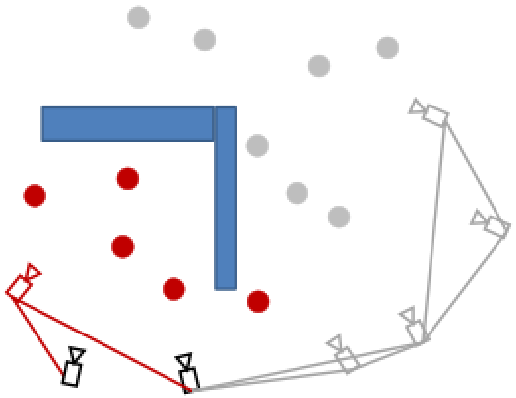
Perceptual aliasing is common in indoor scenarios

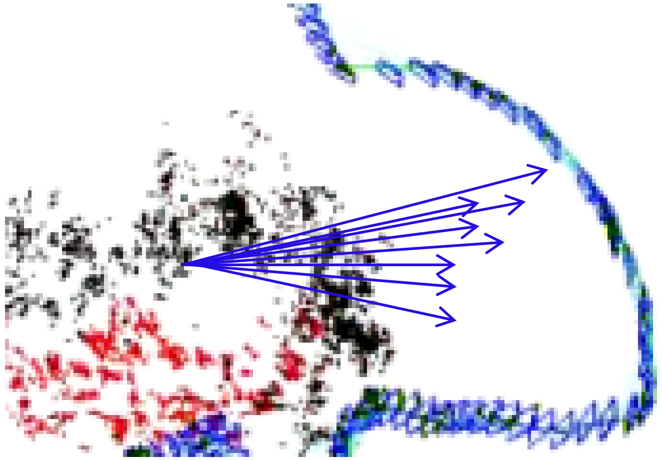
False positives

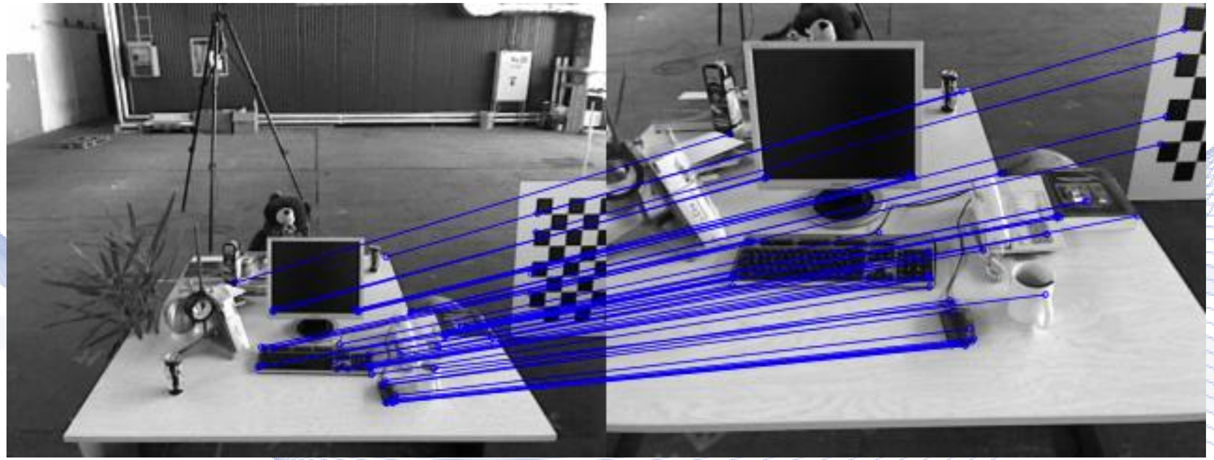
- False positives may ruin the map.

- You must add robustness in the SLAM back-end:
  - Rigidity.
  - Repeated detection.

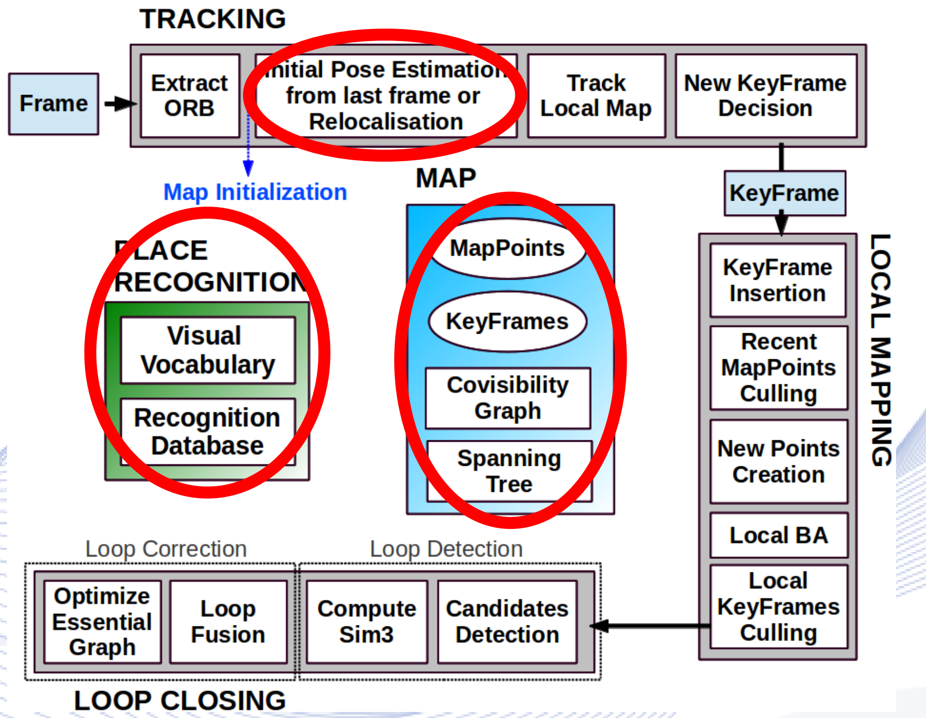
Relocation



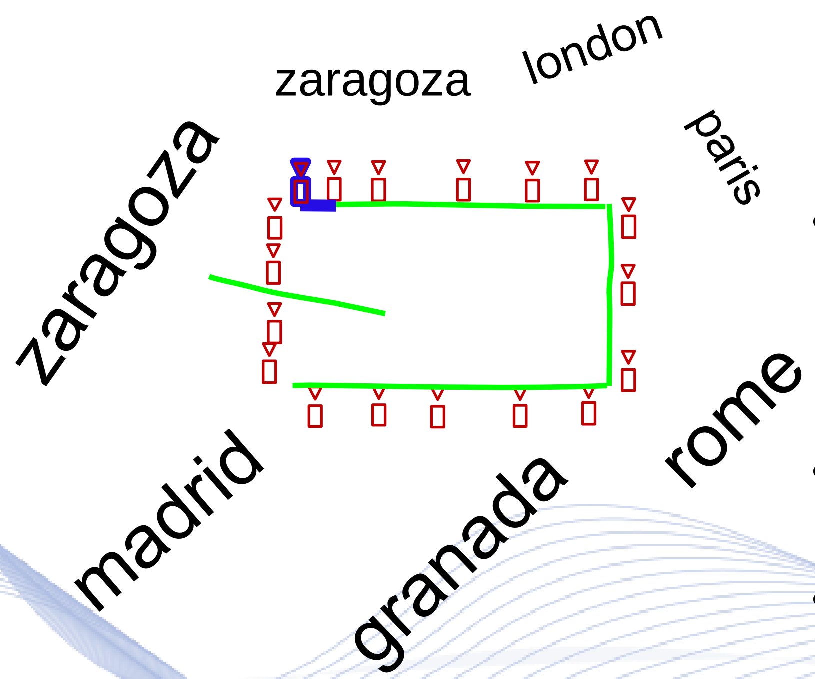




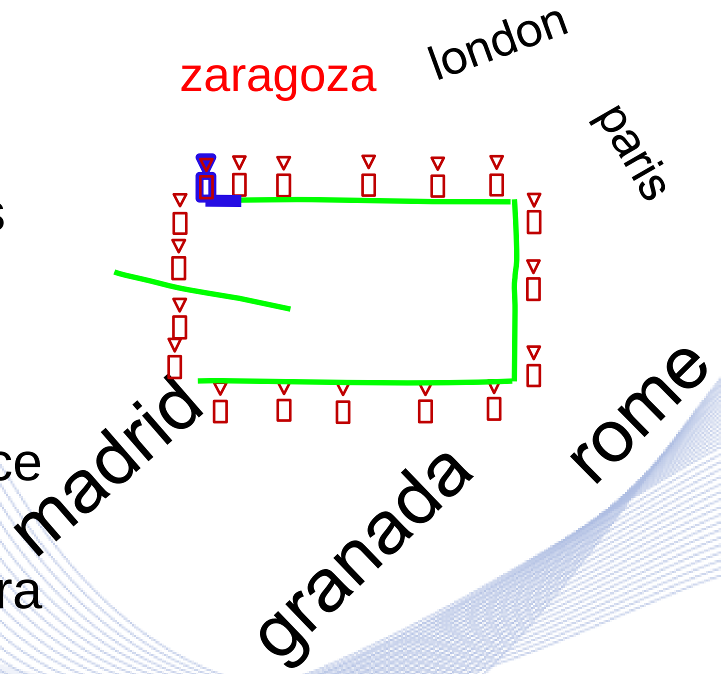
query most similar keyframe



Loop closure detection and correction



- Monocular VSLAM in exploratory trajectory drifts:
  - Translation
  - Orientation
  - Scale
- Predict-match-update loop fails
  - Duplicated map



- Loop detection by place recognition
- Computation relative to camera pose
- Pose graph cameras correction
- Map correction

H. Strasdat, J.M.M. Montiel and A.J. Davison, "*Scale Drift-Aware Large Scale Monocular SLAM",* RSS, 2010.

SLAM in dynamic scenes

3D Localization and Mapping

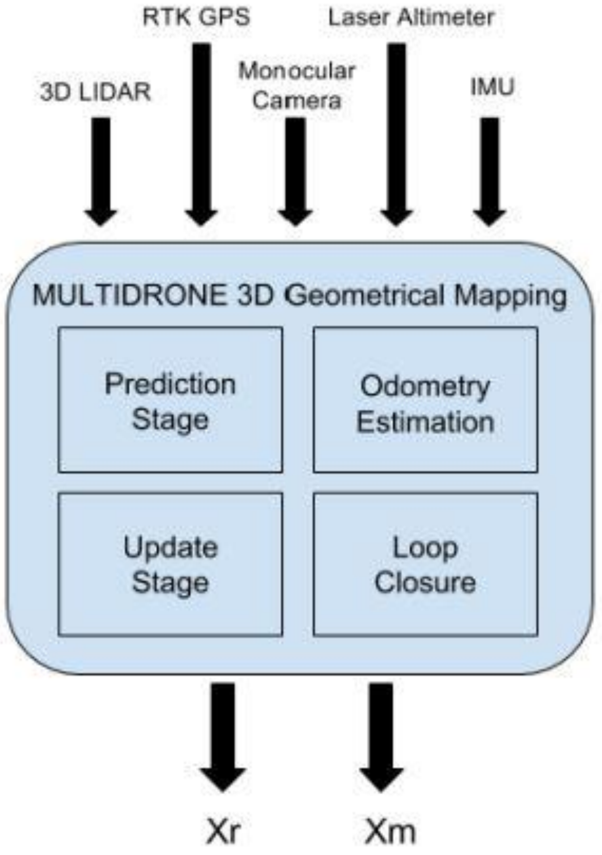
- Sensors
- Mapping
- Localization
- SLAM
- **Data fusion in drone localization.**

Data fusion in drone localization

- The main idea of improving accuracy on localization and mapping in Multidrone is to exploit the synergies between different sensors such as:
  - RTK GPS.
  - 3D LIDAR.
  - Monocular camera pointing downwards.
  - Laser altimeter.
  - Inertial Measurements Units (IMU).

Data fusion in drone localization

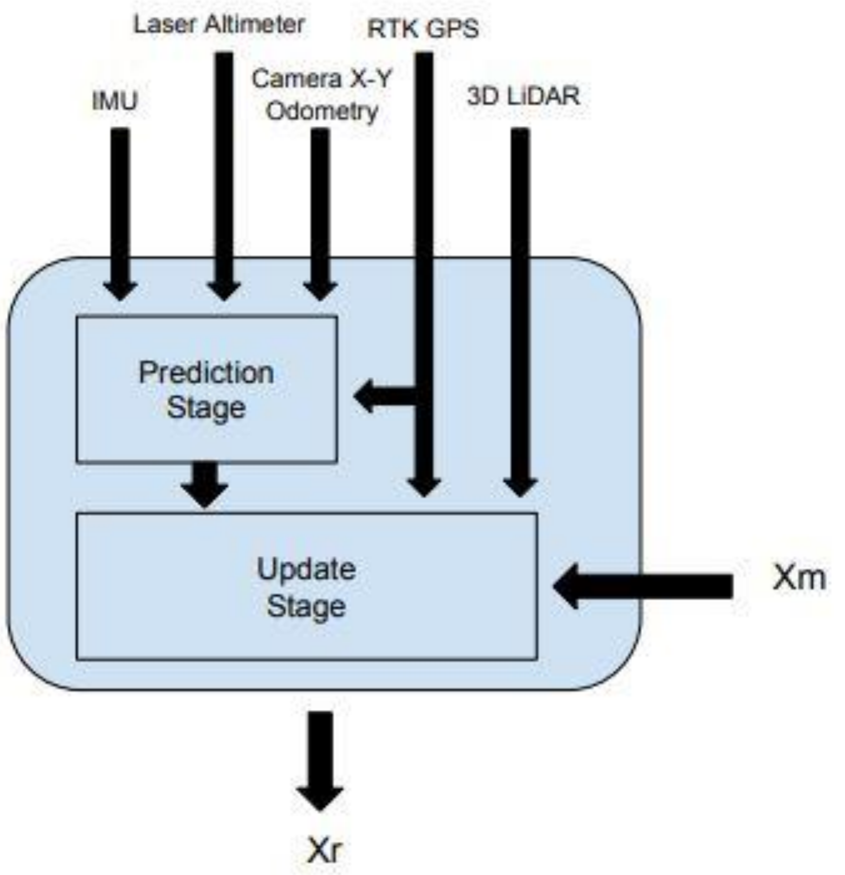
- INPUT: measurements from multiple sensors.
- OUTPUT: 3D geometrical map X_m and the 3D drone pose estimation X_r.



Artificial Intelligence & Information Analysis Lab

Data fusion in drone localization

- INPUT: measurements from multiple sensors + 3D geometrical map.
- OUTPUT: 3D drone pose estimation X_r.

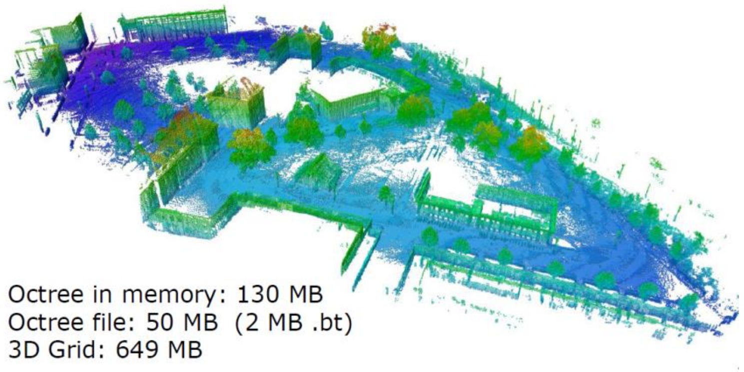


Final result: 3D maps

- Formats:
  - 3D triangle mesh.
  - 3D Octomap.

Octomap:

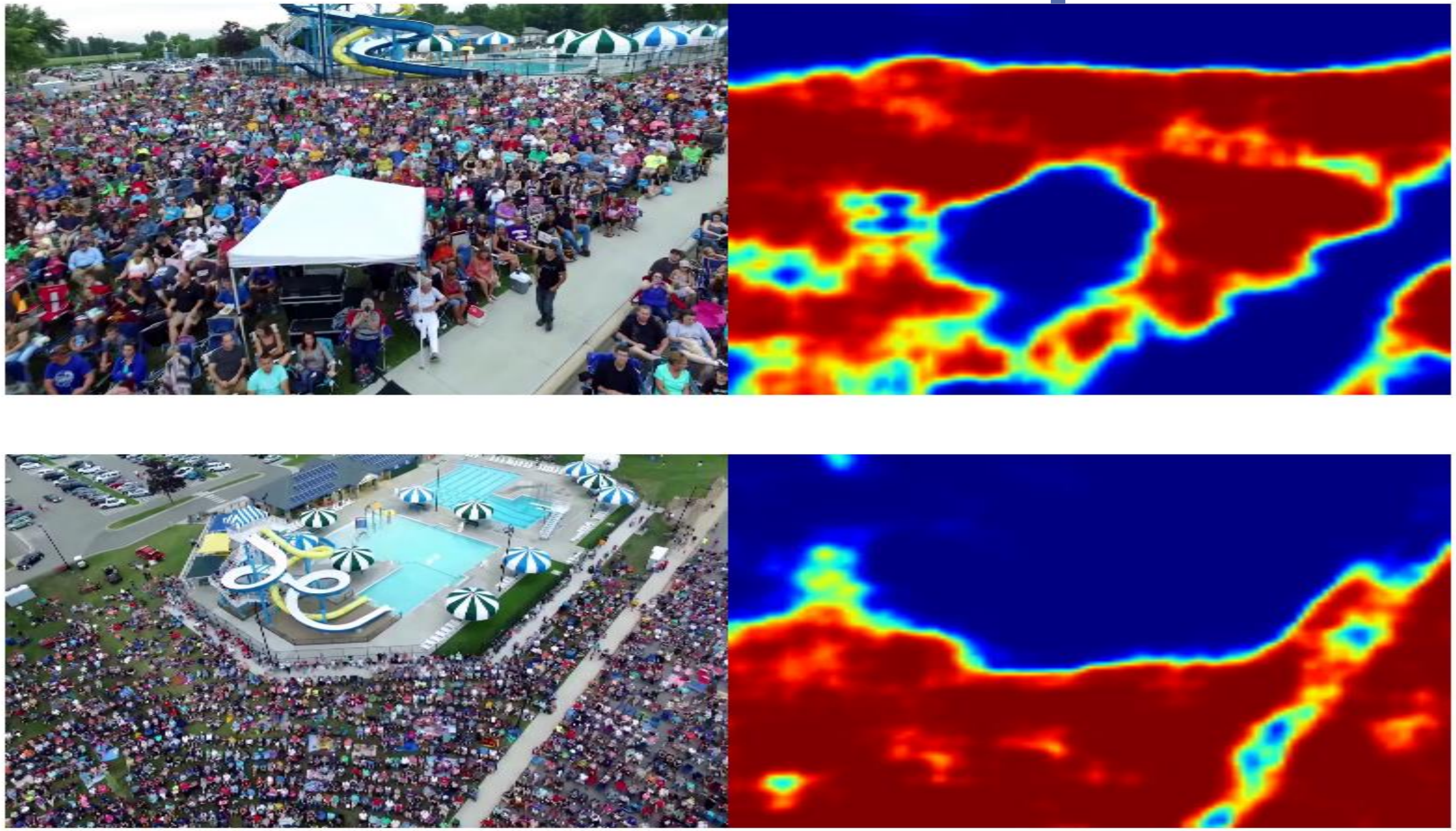
- It is a full 3D model representing the 3D environment.
- It provides a volumetric representation of space, namely of the occupied, free and unknown areas.
- It is based on octrees and uses probabilistic occupancy estimation.



Semantic 3D Map Annotation

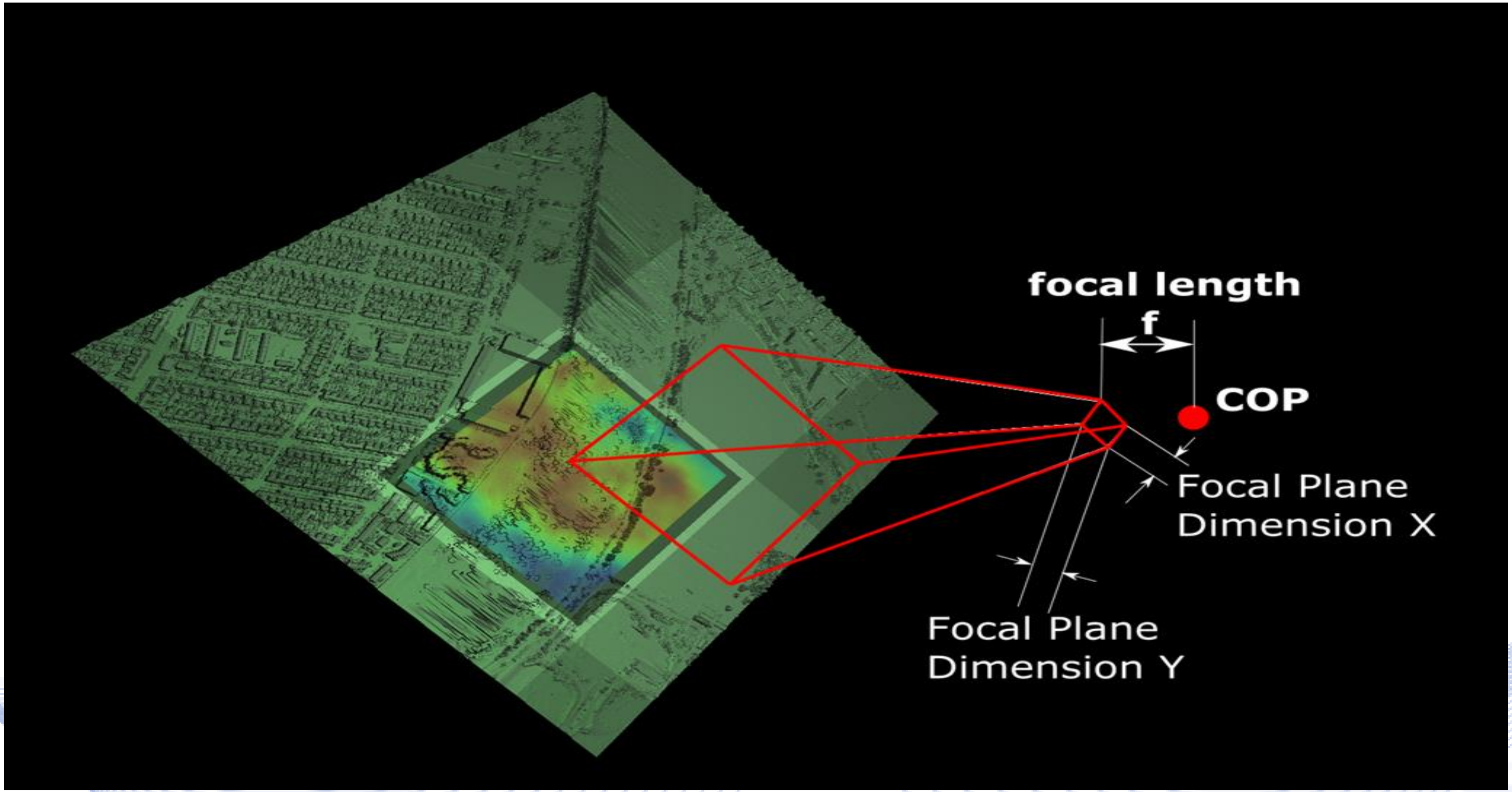
- Dynamic annotations derived through drone video analysis are projected on the 3D map.
- 3D scene models: 3D Mesh or Octomap.
- Assumes that we know the camera extrinsic and intrinsic parameters.

Projection of crowd location onto the 3D map



Crowd detection and mapping.

Semantic 3D Mesh Map Annotation



Crowd region projection on a 3D map.

Bibliography

[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.

[PIT2017] I. Pitas, “Digital video processing and analysis” , China Machine Press, 2017 (in Chinese).

[PIT2013] I. Pitas, “Digital Video and Television” , Createspace/Amazon, 2013.

[NIK2000] N. Nikolaidis and I. Pitas, “3D Image Processing Algorithms”, J. Wiley, 2000.

[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr

Artificial Intelligence &
Information Analysis Lab