DTU

Daniel Olesen, DTU Space

30540 – Photogrammetry (6)

Todays Lecture

- Recap on previous lectures (Camera Model, P3P and RANSAC)
- Light Detection and Ranging (LiDAR)
- Vision-based Navigation and Sensor Fusion with GNSS and INS (Research topic + student project proposals)

Pinhole Camera Model

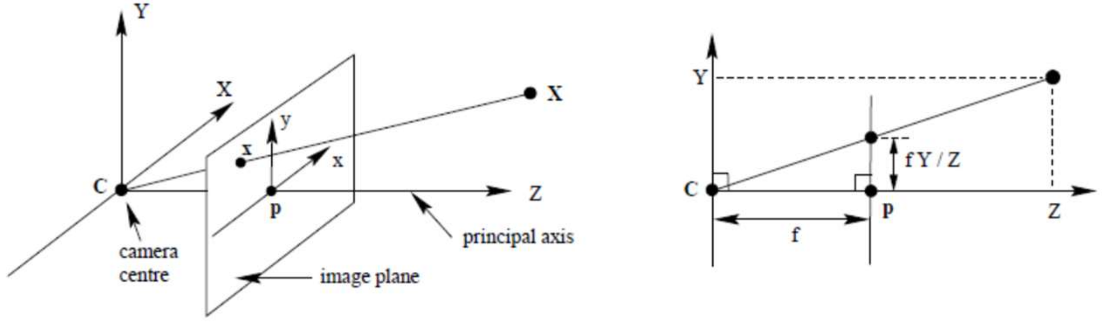


Fig. 6.1. **Pinhole camera geometry.** C *is the camera centre and* p *the principal point. The camera centre is here placed at the coordinate origin. Note the image plane is placed in front of the camera centre.*

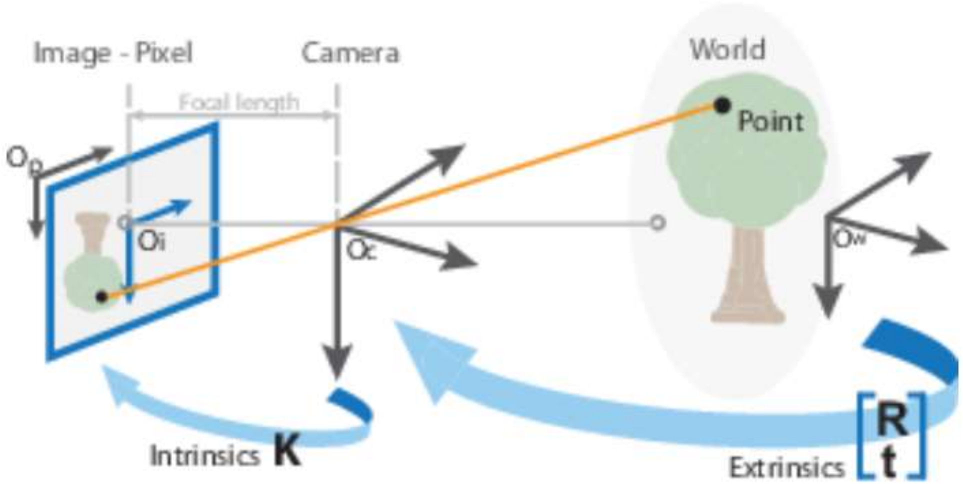
$$\boxed{(X, Y, Z)^{\mathsf{T}} \mapsto (fX/Z, fY/Z)^{\mathsf{T}}}$$ → Cartesian coordinates

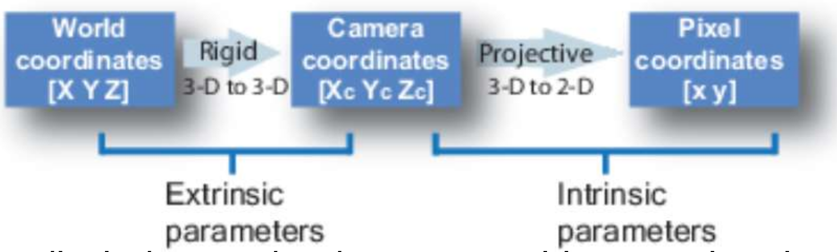
$$\boxed{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}$$ → Homogeneous coordinates

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

Source: Hartley & Zisserman: Multiple View Geometry

Extrinsic and intrinsic parameters





Source: https://se.mathworks.com/help/vision/ug/camera-calibration.html

Complete camera-model

$\mathbf{x} = \mathbf{P}\mathbf{X}$

Intrinsic (or calibration parameters)

$$\begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix}_{\text{imag}} = K[\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_{\text{world}}$$

Extrinsics (rotation + translation)

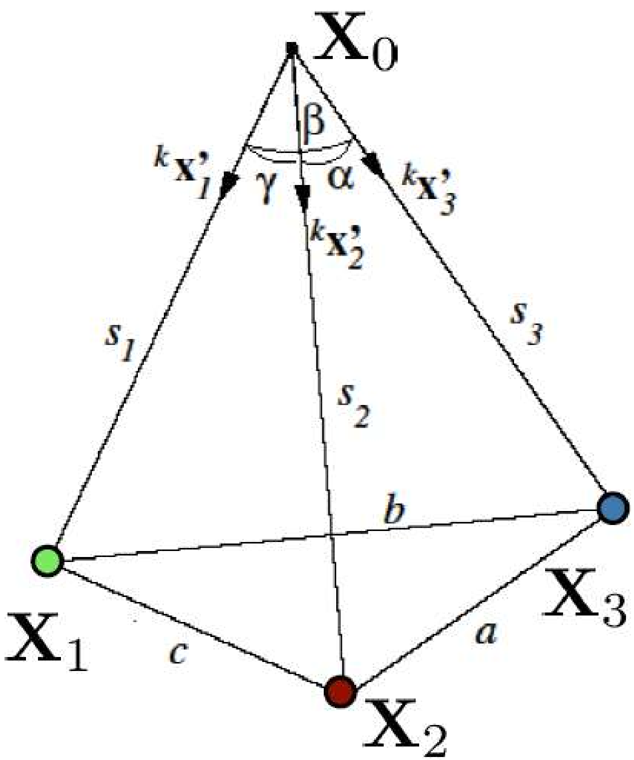
There is a total of 11 parameters in this model (5 intrinsics + 6 extrinsics

What are the intrinisic parameters?

Spatial Resection (P3P, PnP)

- Spatial resection refers to the problem of finding the coordinates of the camera centre and orientation w.r.t. the world given known correspondences of world-points and image points
- Fundamental for Camera-calibration and also widely-used in image-based navigation
- The problem was first treated in 1841 by Grünert but has received a lot of attention within Photogrammetry and Computer Vision.
- In the following the main concepts for Grünert's solution is given with 3 known world points
- We assume that the camera is calibrated (intrinsics and distortions are known).

Spatial Resection



$\mathbf{X_0}$ is the unknown camera-centre

$\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}$ is **known** world points

The angles between the points can be deduced as:

$$\alpha = \cos^{-1}(\overrightarrow{\boldsymbol{x_2}} \cdot \overrightarrow{\boldsymbol{x_3}}), \beta = \cos^{-1}(\overrightarrow{\boldsymbol{x_1}} \cdot \overrightarrow{\boldsymbol{x_3}}),$$

$$\gamma = \cos^{-1}(\overrightarrow{\boldsymbol{x_1}} \cdot \overrightarrow{\boldsymbol{x_2}})$$

The sides, a, b, c can be found as:
$a = |\mathbf{X_3} - \mathbf{X_2}|, b = |\mathbf{X_3} - \mathbf{X_1}|, c = |\mathbf{X_2} - \mathbf{X_1}|$

How do we determine s_1, s_2, s_3?

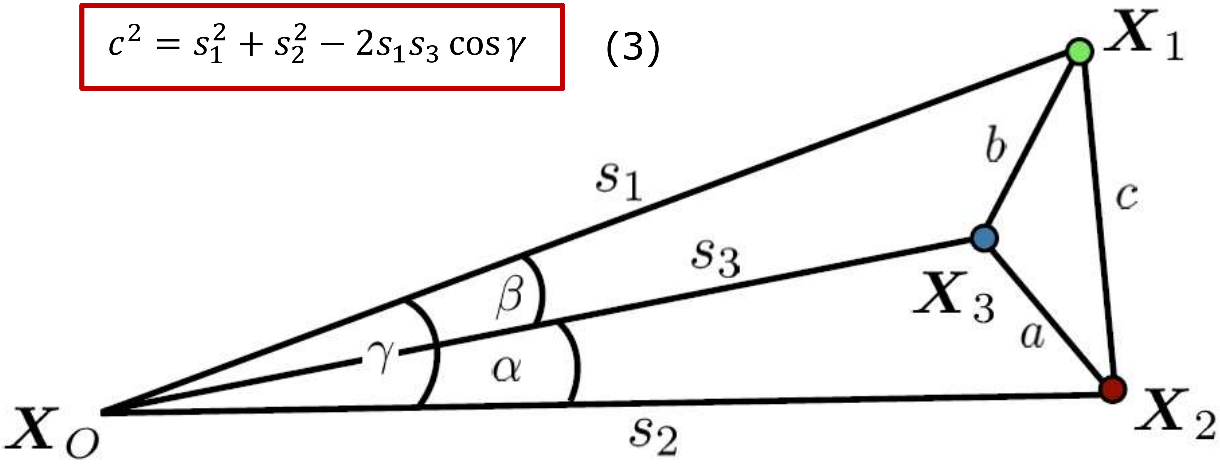
Spatial Resection

The Law of cosines!

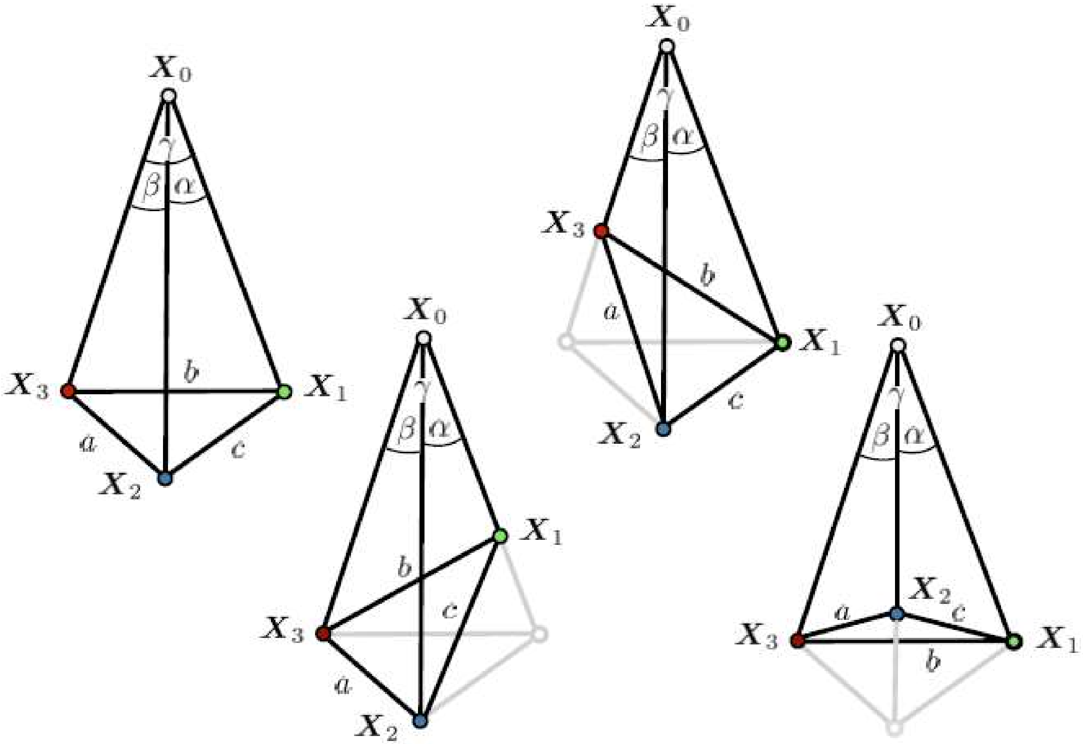
$$\boxed{a^2 = s_2^2 + s_3^3 - 2s_2 s_3 \cos\alpha} \quad (1)$$

$$\boxed{b^2 = s_1^2 + s_3^2 - 2s_1 s_3 \cos\beta} \quad (2)$$

$$\boxed{c^2 = s_1^2 + s_2^2 - 2s_1 s_3 \cos\gamma} \quad (3)$$



4 solutions to the polynomial



Finding the Camera centre…

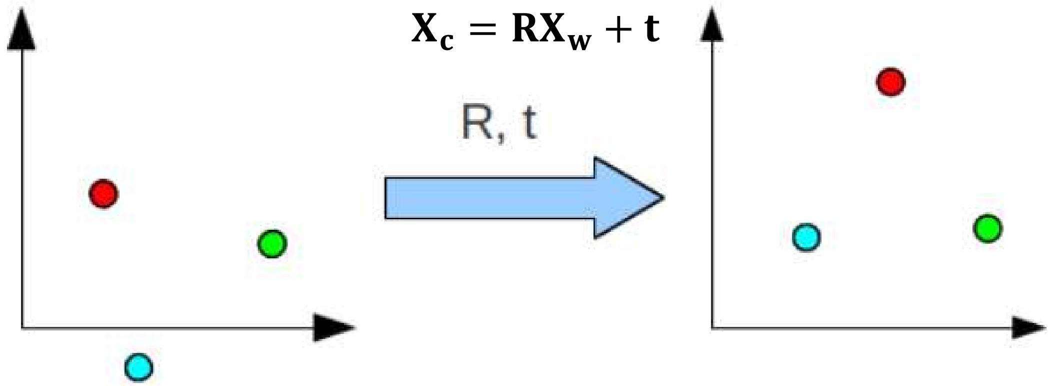
- From the previous slides, we can now express the known World-points in the camera-coordinate frame -> However the task was to identify the world-coordinates of the camera centre…

We first need to solve for $(\mathbf{R}, \mathbf{t})$,

Camera-centre in camera-cordinates

Camera-centre in world-cordinates

$$\mathbf{X_c} = \mathbf{R}\mathbf{X_w} + \mathbf{t}$$



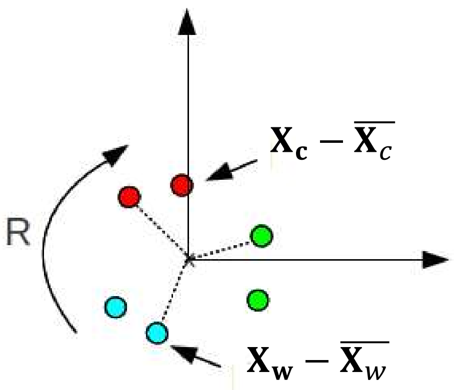
http://nghiaho.com/?page_id=671

Finding the Camera centre…

- First step, is to calculate the centroids (average point) from each representation (world and camera coords.)

$$\overline{\mathbf{X}_C} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_{C}^{\boldsymbol{i}},\ \overline{\mathbf{X}_w} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_{w}^{\boldsymbol{i}},$$

- Subtracting the centroids from the original point-sets, leaves only the Rotation unknown…



$$\mathbf{H} = \sum_{\boldsymbol{i}}^{\boldsymbol{N}} \left(\mathbf{X}_{w}^{\boldsymbol{i}}, -\overline{\mathbf{X}_w}\right)\left(\mathbf{X}_{c}^{\boldsymbol{i}}, -\overline{\mathbf{X}_c}\right)^{\boldsymbol{T}}$$

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \boldsymbol{SVD}(\mathbf{H})$$

$$\mathrm{d} = \mathrm{sgn}(\det(\mathbf{V}\mathbf{U}^{\mathbf{T}}))$$

$$\mathbf{R} = \mathbf{V}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mathrm{d} \end{bmatrix}\mathbf{U}^{\mathbf{T}}$$

http://nghiaho.com/?page_id=671

Finding the Camera centre…

- Finally, we can find the translation as:

$$\mathbf{t} = -\mathbf{R}\overline{\mathbf{X}_w} + \overline{\mathbf{X}_c}$$

The camera centre can hence be found as…

$$\mathbf{X_c} = \mathbf{R}\mathbf{X_w} + \mathbf{t} \Leftrightarrow \mathbf{X_w} = \mathbf{R}^{-1}(\underbrace{\mathbf{X_c}}_{[0 \quad 0 \quad 0]^T} - \mathbf{t}) = -\mathbf{R}^{-1}\mathbf{t} = -\mathbf{R}^{\mathbf{T}}\mathbf{t}$$

Note, that algorithm for finding the camera centre is known as the Kabsch algorithm. See also the paper from Arun et. al (PAMI-3DLS-1987.pdf)

RANSAC

Why do we need RANSAC?

- Feature matching result often contains a certain amount of false matches (outliers). Outliers will severally deteriorate or skew the estimation result.

- Generally speaking, RANdom SAmple Consensus (RANSAC) is an iterative method for estimating model parameters from observations containing outliers. The model could be a line, a parabola, an ellipse, etc.

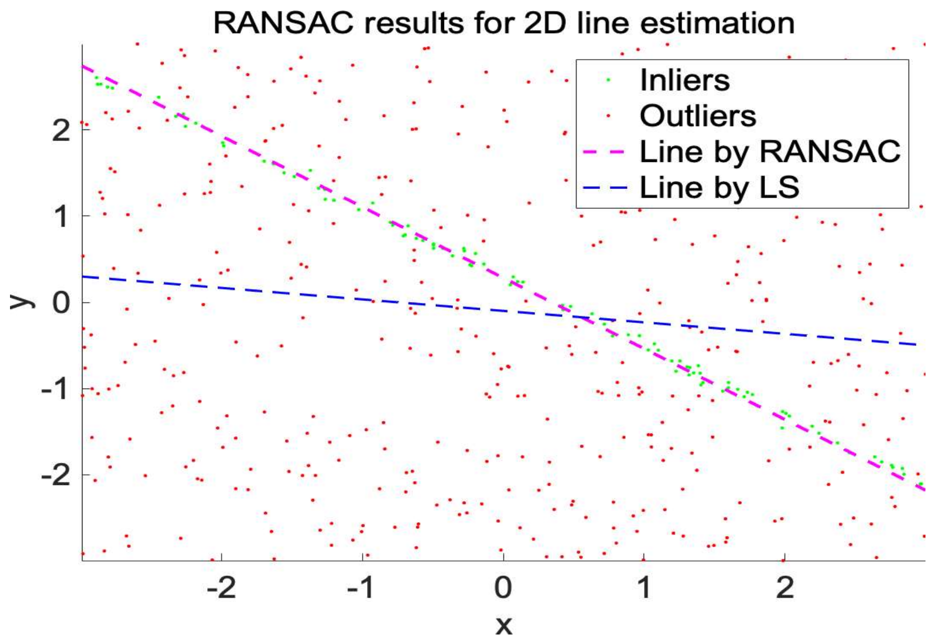


Figure: Line estimation results.

RANSAC Pipeline

- Randomly select n samples from observations, n is the minimum number of samples needed for model estimation, e.g. n = 2 for 2D line, n = 3 for 3D plane, etc.
- Model estimation, could be LS, SVD, etc.
- Compute Consensus: apply a model-specific loss function to each observation and the model obtained, the response could serve as the consensus, e.g. point-line distance, point-plane distance.
- Classifier inliers and outliers using a predefined threshold and log the model with the maximum number of inliers.
- Iterate

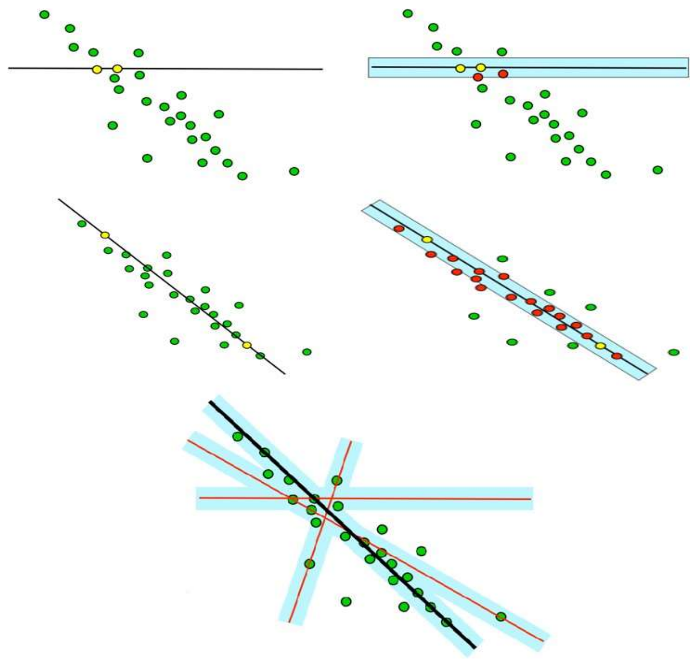


Figure: RANSAC Procedure.

How many iterations to choose?

Let p_{out} be the probability that one point is an outlier, n be the minimum number of samples for model estimation, N be the number of iterations, p be the desired probability that we get a good sample:

$$p = 1 - (1 - (1 - p_{out})^n)^N$$

- $1 - p_{out}$: Probability of an inlier.
- $(1 - p_{out})^n$: Probability of choosing n inliers.
- $(1 - (1 - p_{out})^n)$: Probability of at least one items in the sample being outliers for one iteration.
- $(1 - (1 - p_{out})^n)^N$: Probability that N iterations are contaminated.
- $1 - (1 - (1 - p_{out})^n)^N$: Probability that at least one iteration is not contaminated.

Usually, we set p and then compute N backwardly as $N = \frac{\log(1-p)}{\log(1-(1-p_{out})^n)}$.

How many iterations to choose?

A look-up table for N when p is set to 0.99.

				p_{out}			
n	5%	10%	20%	25%	30%	40%	50%
2	2	3	5	6	7	11	17
3	3	4	7	9	11	19	35
4	3	5	9	13	17	34	72
5	4	6	12	17	26	57	146
6	4	7	16	24	37	97	293
7	4	8	20	33	54	163	588
8	5	9	26	44	78	272	1177

Example of RANSAC

Figure: Feature matching before applying RANSAC.

As you can see, false matches get hypothesized.Let’s apply RANSAC + epipolar constraint.

Example of RANSAC

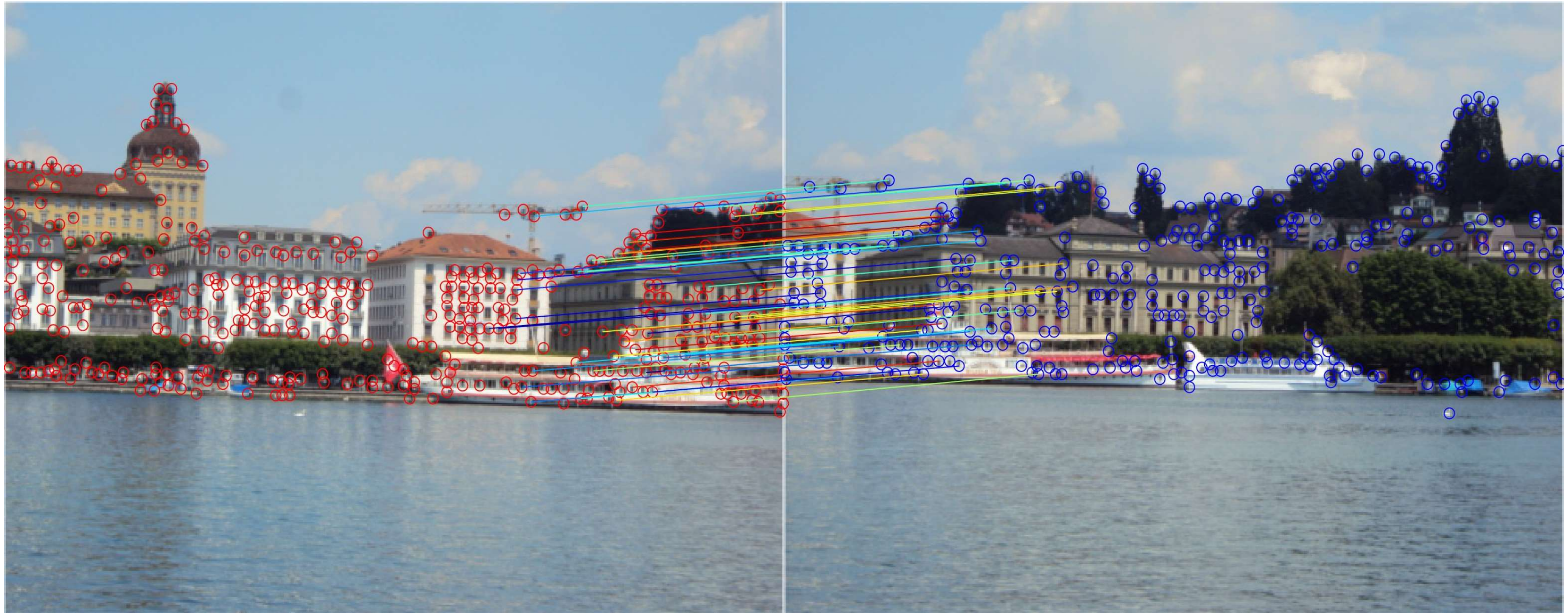


Figure: Feature matching after applying RANSAC.

As you can see, false matches are removed.

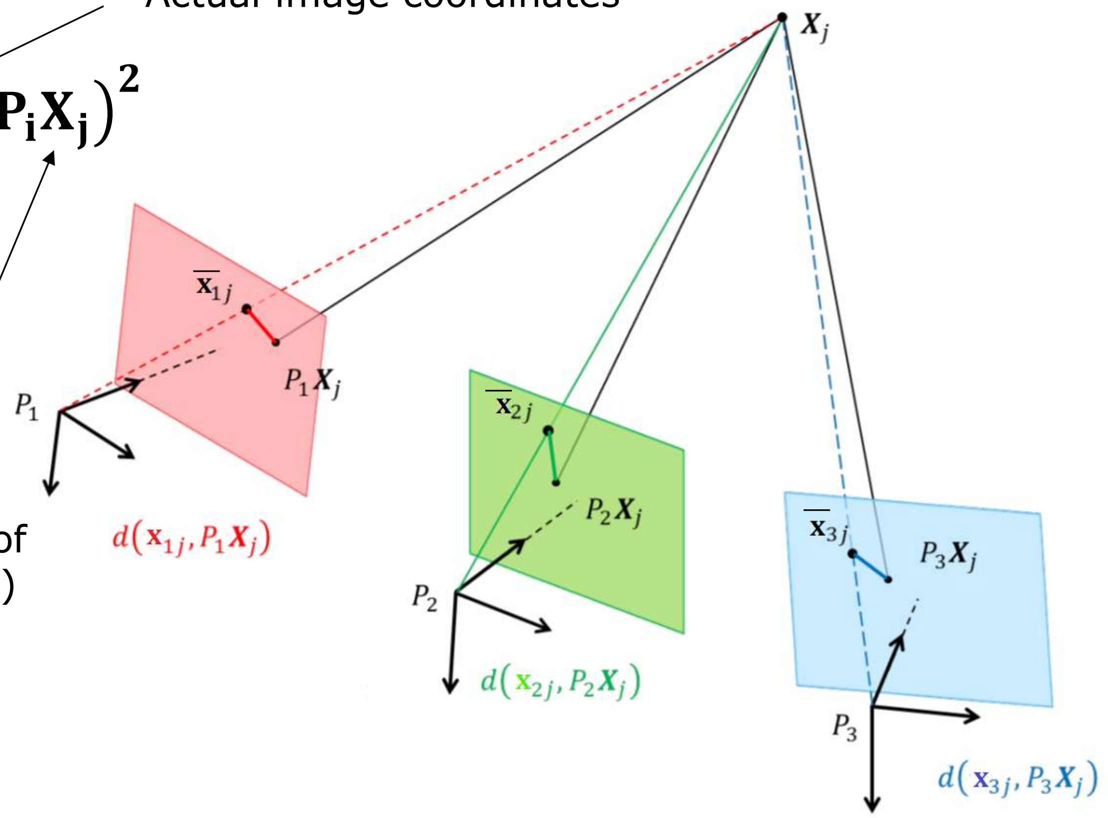
Bundle Adjustment

Actual image coordinates

$$\min_{\mathbf{P_i},\,\mathbf{X_j}} \sum_{i=1}^{m} \sum_{j=1}^{n} d\left(\overline{\mathbf{x_{ij}}}, \mathbf{P_i}\mathbf{X_j}\right)^{\mathbf{2}}$$

Camera projection matrix

Estimate of 3D (world) point



Bundle Adjustment

- Bundle adjustment is statistical optimal
  - Maximum-Likelihood estimator under Gaussian noise

- Exploit all observations and considers uncertainties and correllation

- Requires an initial estimate

- Needs fewer control points compared to P3P (which needs at least 4 per image)

Aerial Triangulation

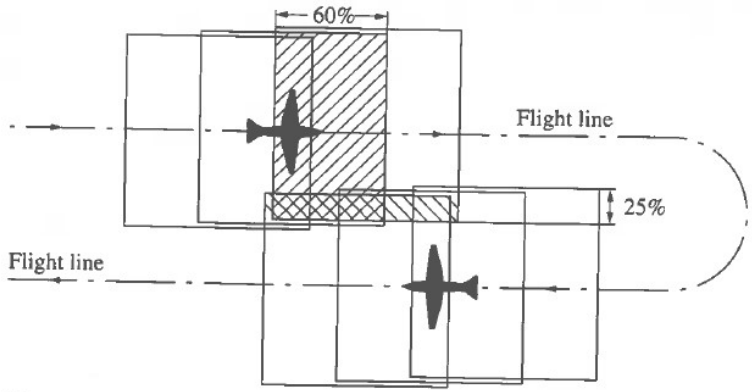


Figure 1-6 Photographic overlap.

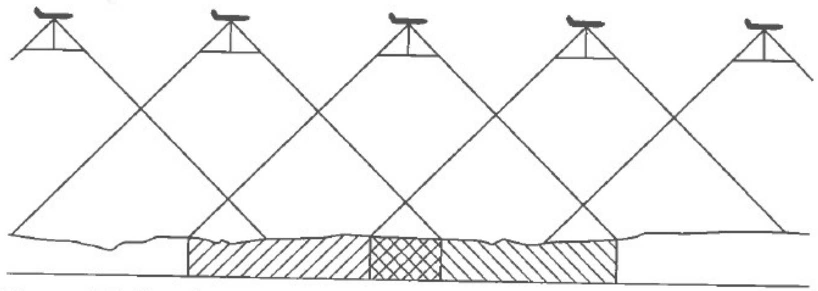
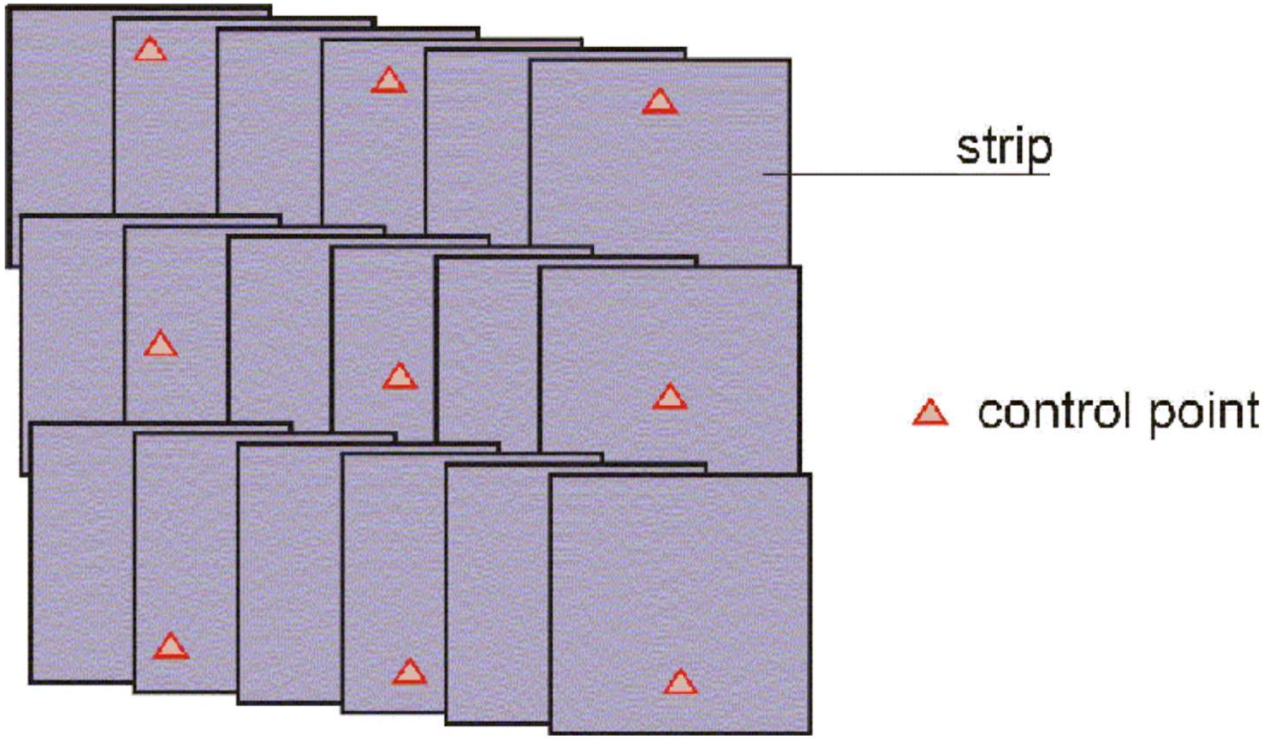


Figure 1-7 Overlap along flight line.

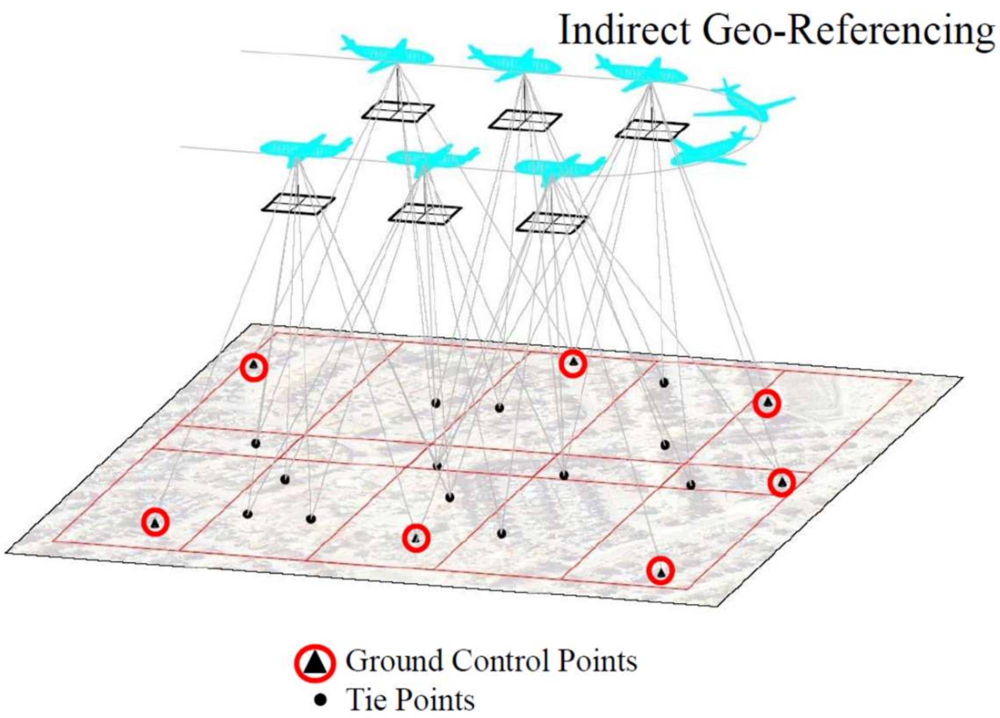
AT is the task of estimating the 3D location of points using aerial images

Bundle-adjustment is the default framework for this process

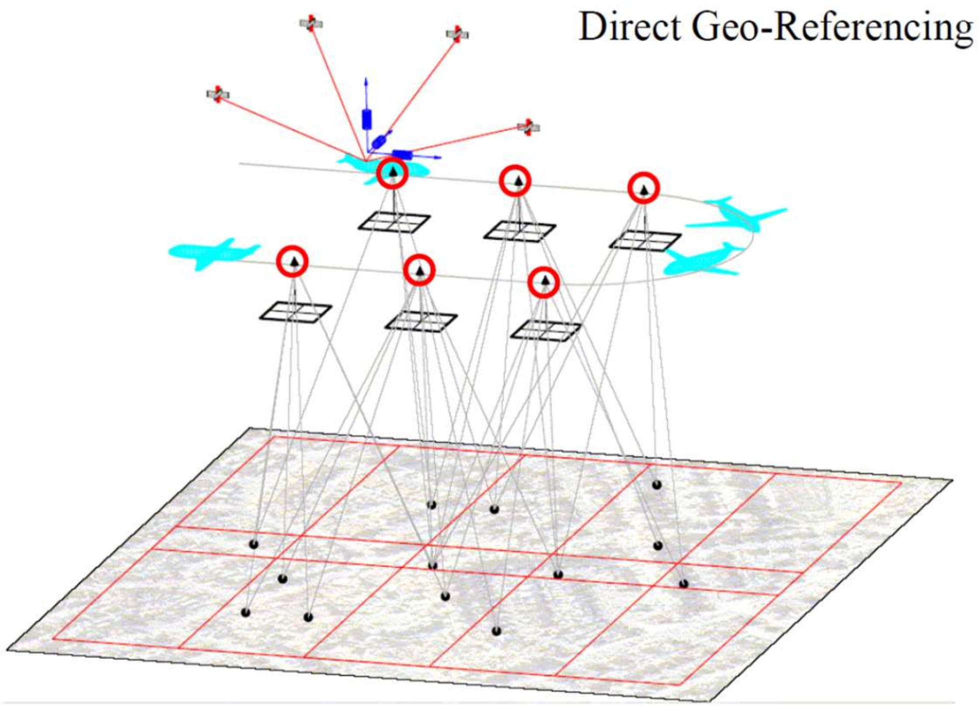
Aerial Triangulation



Aerial Triangulation



Aerial Triangulation



Todays Lecture

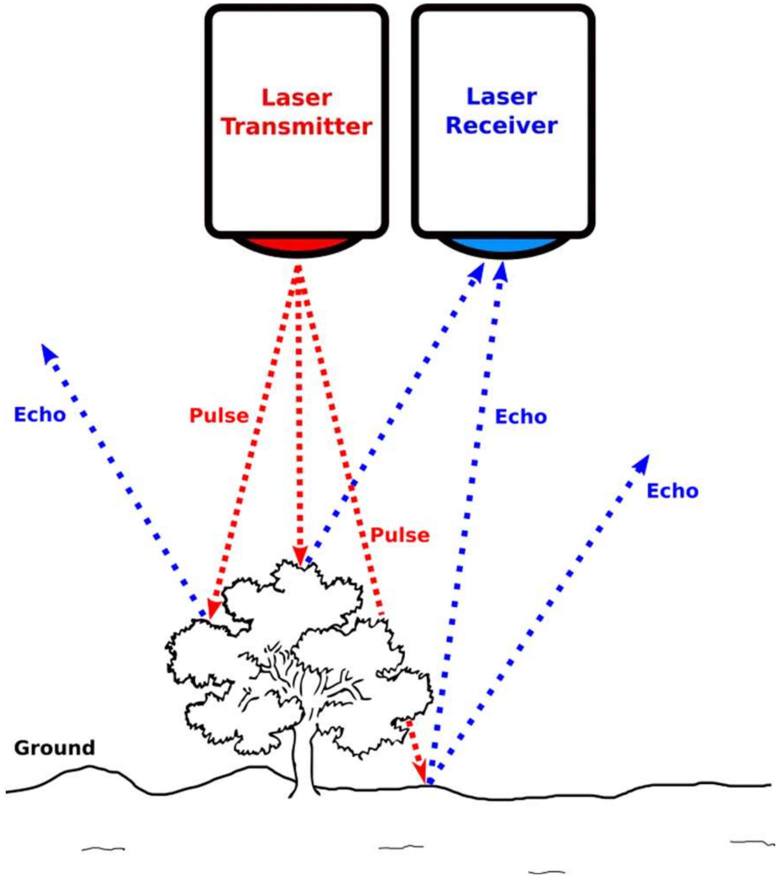
✓Recap on previous lectures (Camera Model, P3P and RANSAC)

- Light Detection and Ranging (LiDAR)

- Vision-based Navigation and Sensor Fusion with GNSS and INS (Research topic + student project proposals)

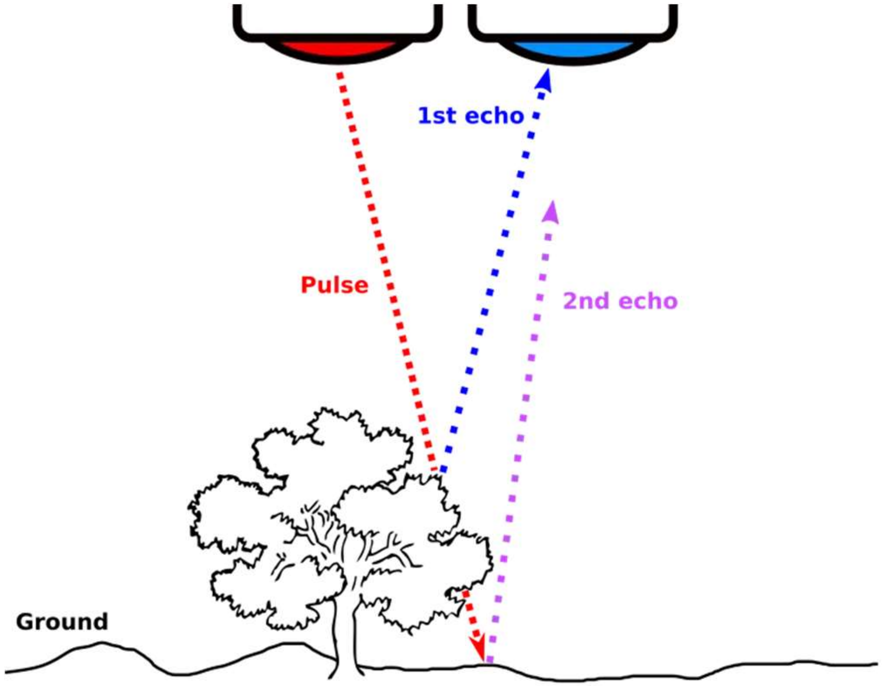
LiDAR basics

- LiDAR: Light detection and ranging; i.e a system that performs remote distance measurements using light - **Active sensor**
- Laser pulse sent towards the target by LiDAR. When the pulse hits an object, the returned echo will be measured.
- Time between transmitted/received pulse is measured and translated into distance to object.

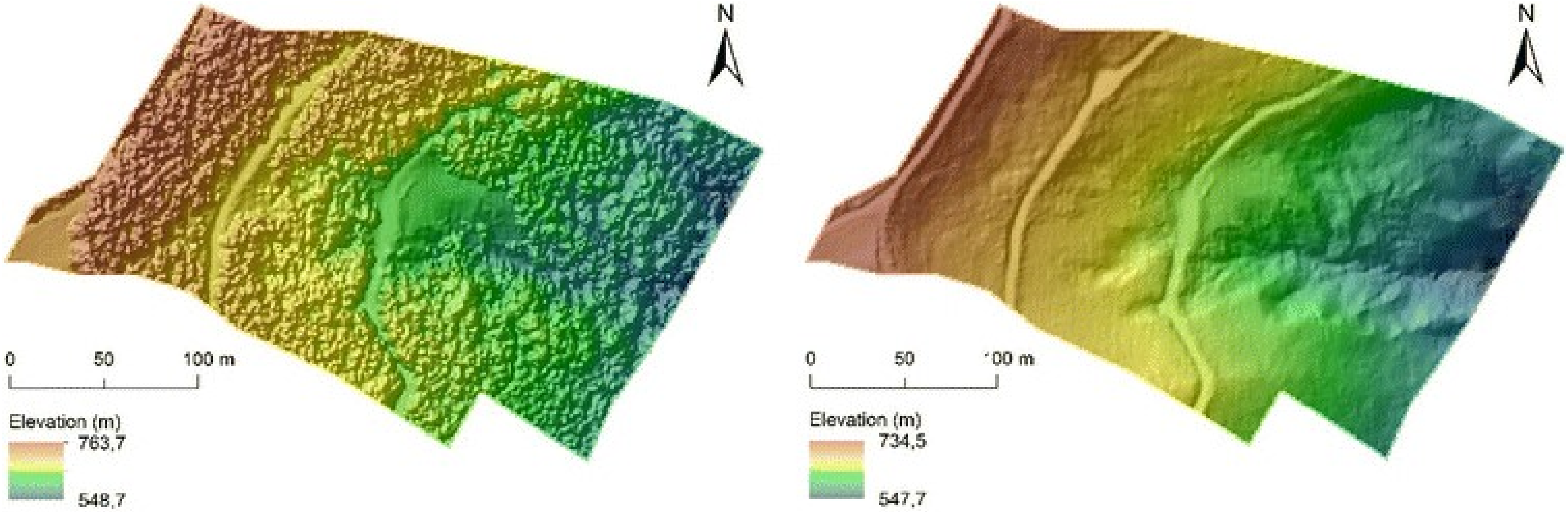


LiDAR return types

- A LiDAR is often able to differentiate between multiple echoes originating from the same pulse. The first echo (strongest) will be reflected by the top of the object closest to the sensor, eg. a rooftop or the top of a tree.
- The following echoes will be returned by the objects that are lower, such as the leaves branches or the ground.
- The Velodyne VLP-16 LiDAR is able to sense between two different echoes.



LiDAR DSM vs LiDAR DTM

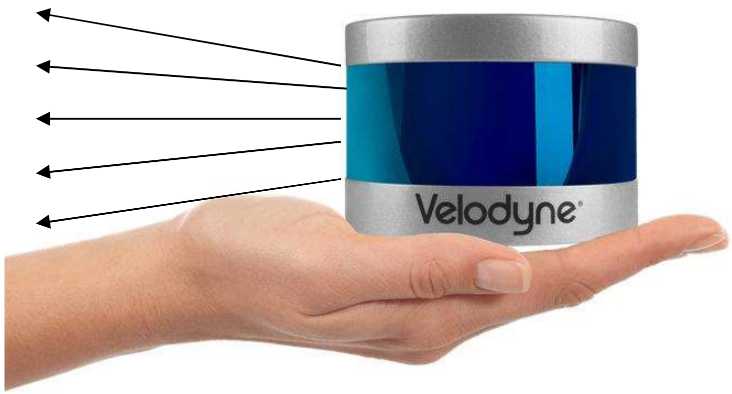


https://www.researchgate.net/figure/LIDAR-DSM-left-and-LIDAR-DTM-right_fig2_321917526

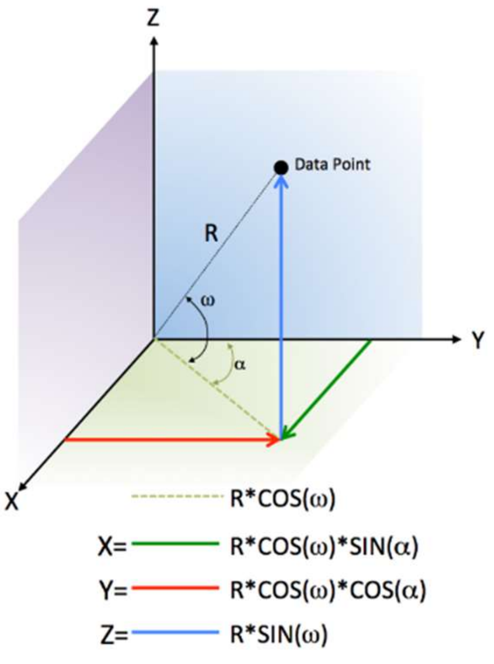
Velodyne VLP-16 Specifications

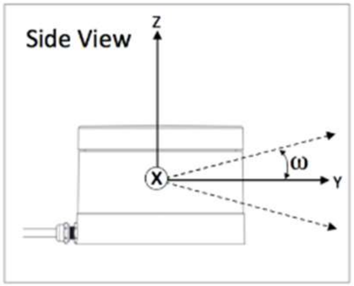
Velodyne VLP-16

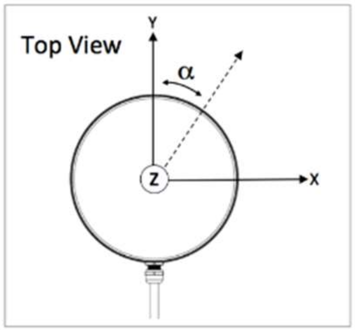
- **16 laser channels**
- **Measures range and calibrated reflectivity (intensity)**
- **Range accuracy: +/- 3 cm**
- **Measurement range: 1-100 meters**
- **Wave length: 903 nm**
- Field of view:
  – 360° horizontal
  – 30° vertical field of view, with ±15° up and down.
- Scan rate: 5-20Hz
- Data rate: 300.000 pts/sec
- Beam width: 0.17

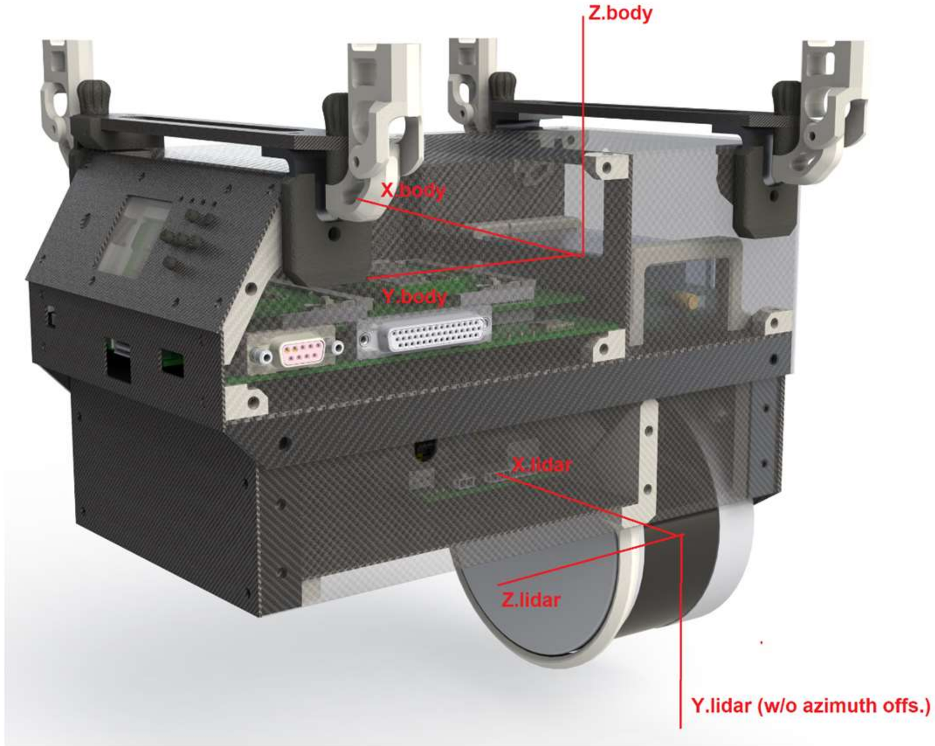


LiDAR payload for UAV









LiDAR vs Photogrammetry

- LiDAR surveys has the advantage that we often can recover terrain in addition to the surface model
- LiDAR surveys can work at night (independent to light conditions)
- LiDAR scans generally use direct georeferencing and require that a high-end GNSS/INS system is installed for georeferencing
- It cost roughly 10 times more to buy a UAV system for LiDAR surveys than for classical photogrammetry with GCPs
- Photogrammetry typically deliver a denser 3D point cloud than LiDAR

Typical UAV Platforms for Photogrammetry and LiDAR surveys

DJI Phantom 4 Pro ~ 15.000 DKK

DJI Matrice 300 RTK ~75.000 DKK

DJI Zenmuse L1 LiDAR sensor (with INS): ~100.000 DKK

Todays Lecture

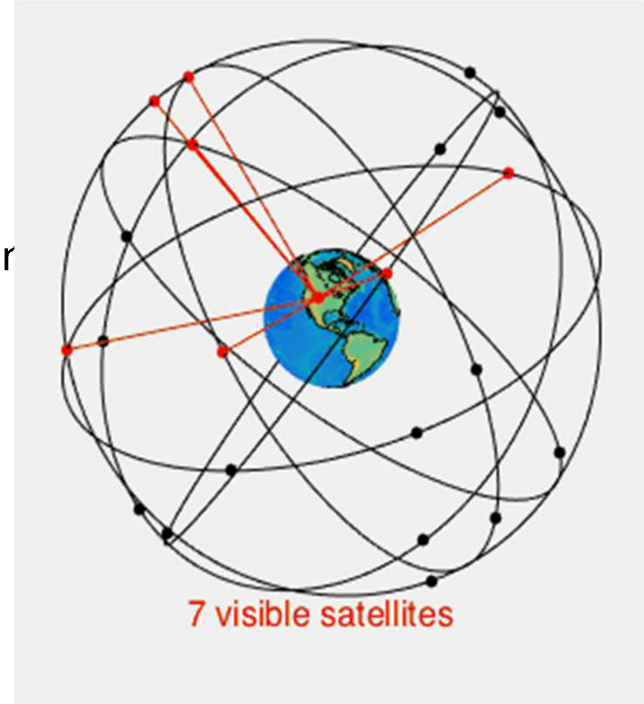
✓Recap on previous lectures (Camera Model, P3P and RANSAC)

✓Light Detection and Ranging (LiDAR)

- Vision-based Navigation and Sensor Fusion with GNSS and INS (Research topic + student project proposals)

Global Navigation Satellite Systems (GNSS)

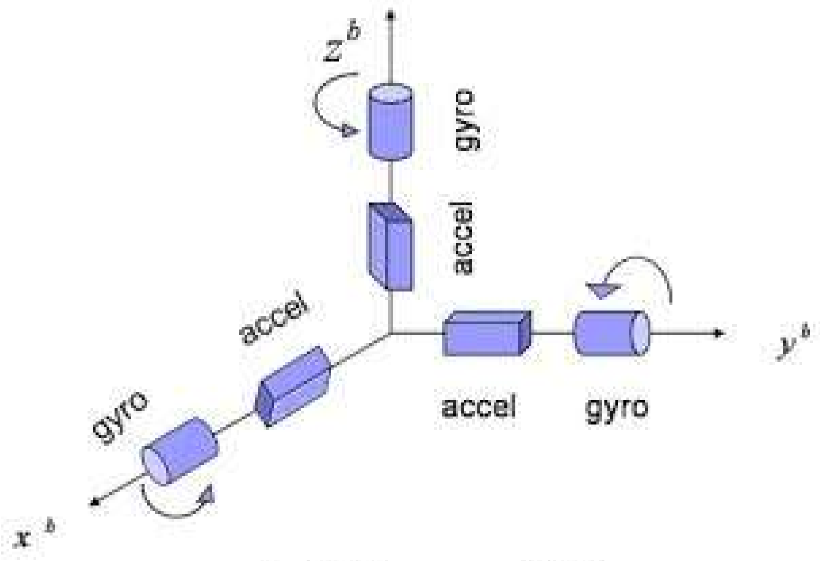
- GNSS = **<u>GPS</u>** + GLONASS + Galileo + Beidou
- More than 100 satellites in total.
- Radio signals are transmitted from satellites that orbits the earth ir approximate 20.000 km altitude
- Satellites positions are always precisely known
- A GNSS receiver computes a position from relative arrival of signals from a minimum 4 satellites
- GNSS requires a good (unobstructed view) to satellites for optimum performance, i.e. does not work well indoors, in urban canyons or under vegation.



GPS orbit simulation
(Wikimedia commons)

Inertial Navigation System (INS)

Inertial Navigation systems are based on an IMU, which measures angular-rate, ω, and specific force, $\bar{f}$, measurements along three orthogonal axes and integrate these into motion equations.



Inertial Measurement Unit
3 accelerometers, 3 gyroscopes

Simplified, INS mechanizations is based on laws from **classical mechanics**:

Linear motion

$$v(t_1) = \int_{t_0}^{t_1} \bar{f}(t)\,dt + v(t_0)$$

$$s(t_1) = \int_{t_0}^{t_1} v(t)\,dt + s(t_0)$$

Rotational motion

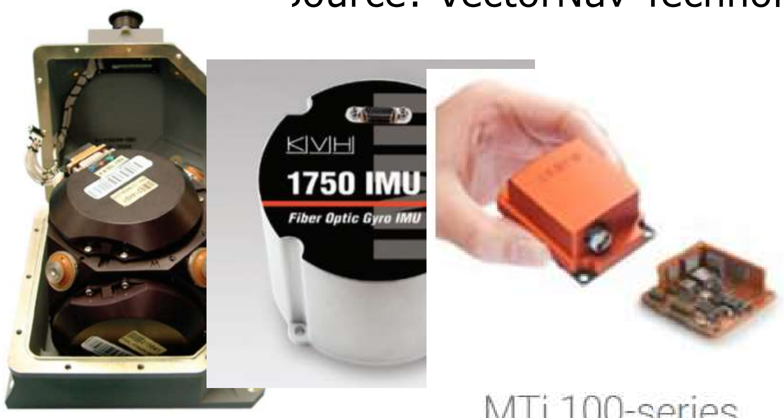
$$\theta(t_1) = \int_{t_0}^{t_1} \omega(t)\,dt + \theta(t_0)$$

Inertial Navigation System (INS)

Grade	Accelerometer Bias Error [mg]		Horizontal Position Error [m] 1s	10s	60s	1hr
Navigation	0.025		0.13 mm	12 mm	0.44 m	1.6 km
Tactical	0.3		1.5 mm	150 mm	5.3 m	19 km
Industrial	3		15 mm	1.5 m	53 m	190 km
Automotive	125		620 mm	60 m	2.2 km	7900 km

>100.000$
10.000-100.000$
1.000-10.000$

Source: VectorNav Technologies



HG9900 IMU

e IMU
Industrial IMU

)
Ex. Xsense Mti-100

Fiber-Optic Gyro (FOG)
Technology: MEMS

0$
Price ~1.000$

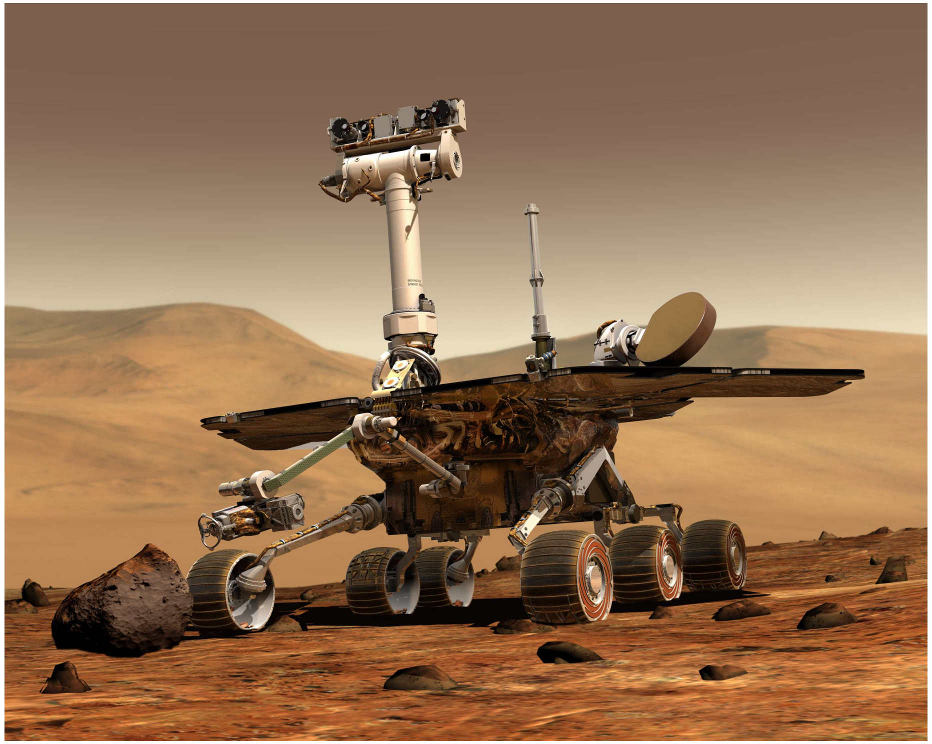
kg
Weight: 55 gram

Vision-based Navigation and Sensor-fusion

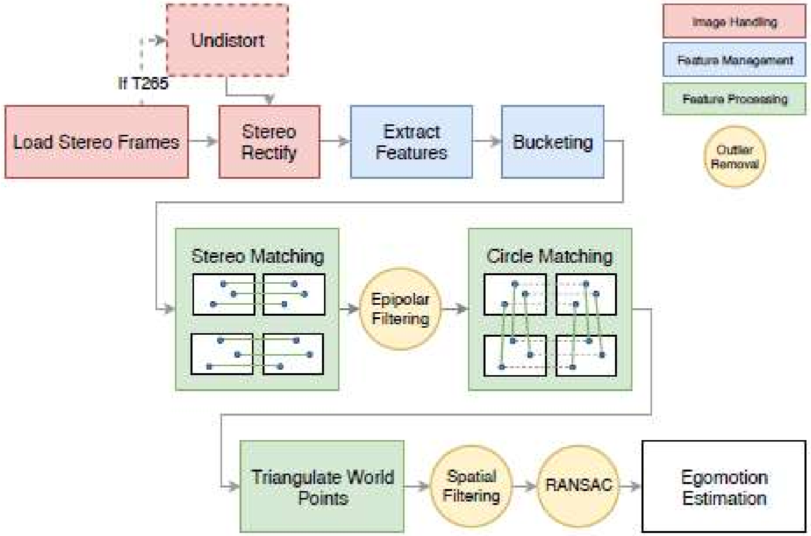
- Navigation with cameras (Visual Odometry) has increased in popularity in recent years due to technological advances (better and cheaper hardware) and algorithm developments

- The Mars Exploration Rovers, Spirit and Opportunity (2003) was among the first platforms to use an onboard (real-time) VO system.

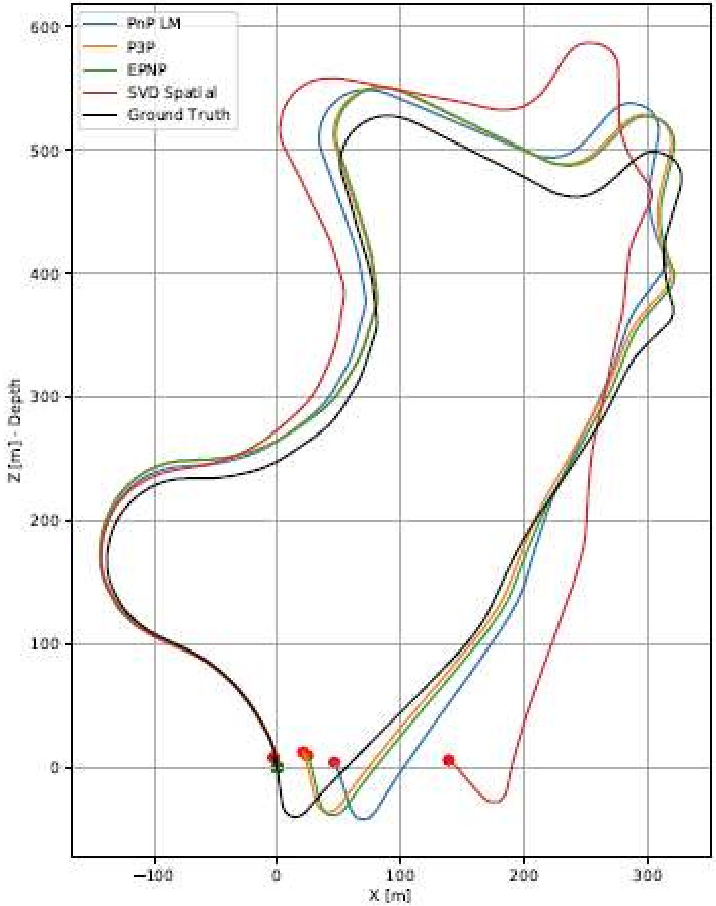
- VO systems can be using a simple camera (monocular VO) or stereo cameras (Stereo VO). In the mono case, an additional sensor would normally be used to help resolve the scale ambiguity.



Stereo Visual Odometry



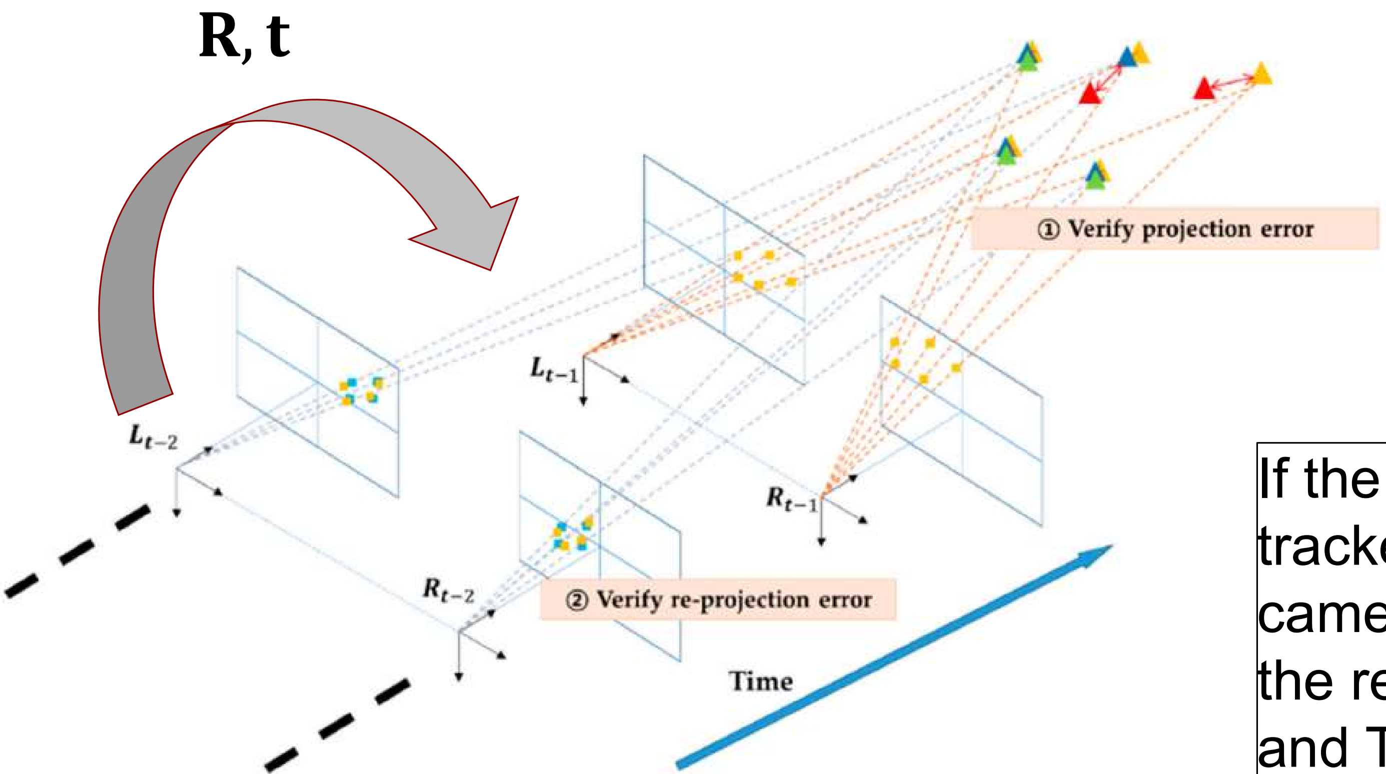
Example of an Stereo VO pipeline



P3P is used for estimating camera-movement

VO Performance on KITTI dataset

Feature Tracking (Circle Matching)



S-J Yoon & T. Kim (2019)

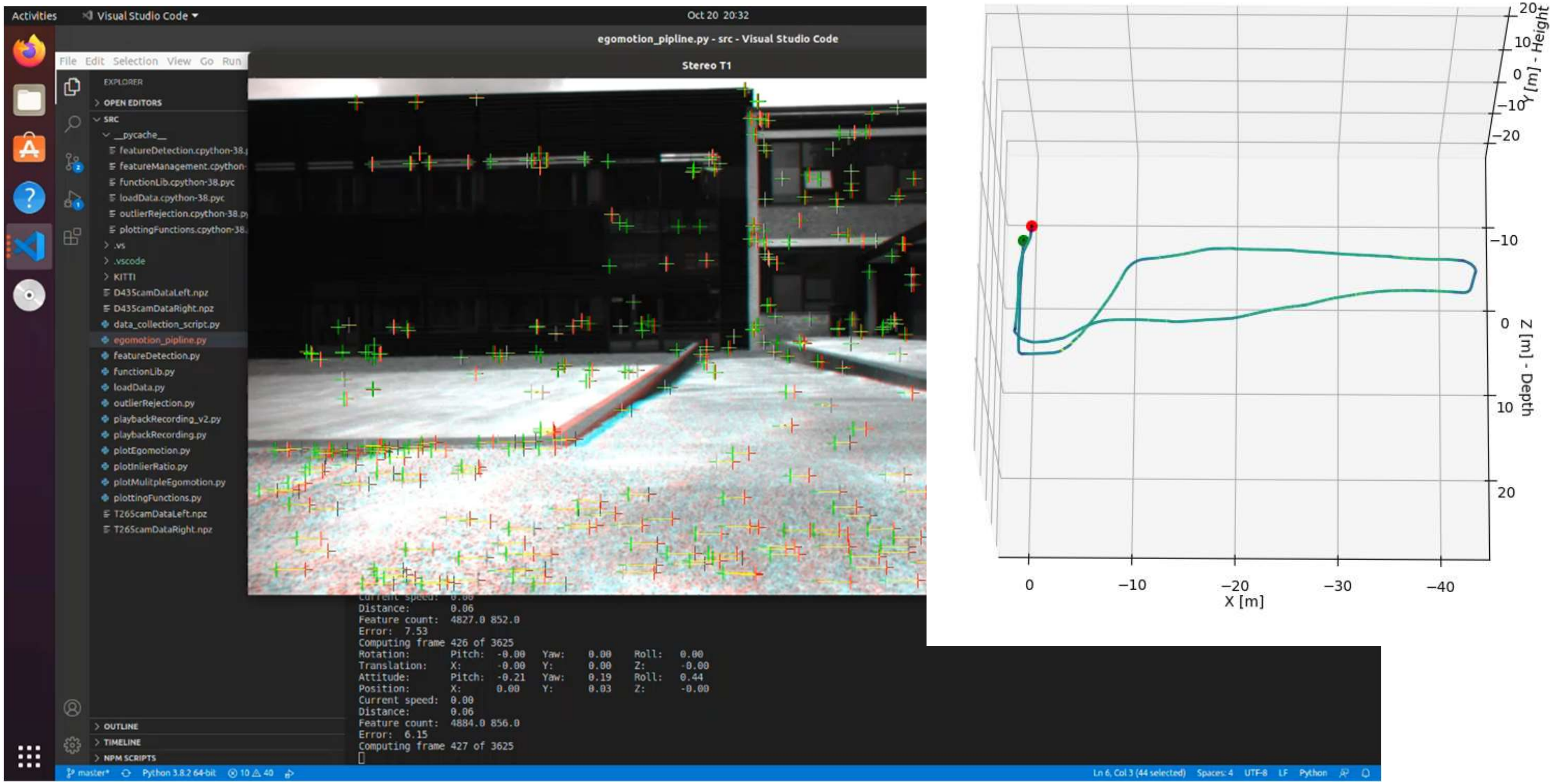
If the same feature points are tracked by the left and right cameras for multiple frames the relative motion (Rotation and Translation can be recovered)

UGV test platform

AION robotics UGV platform with pixhawk open source flight controller

- High-end GNSS antenna and receiver (capable of cm’s accuracy)
- MEMS based IMU
- Front-facing stereo camera (Intel Realsense D455)
- Upward facing fish-eye camera for situational awareness and sensing GNSS obstacles

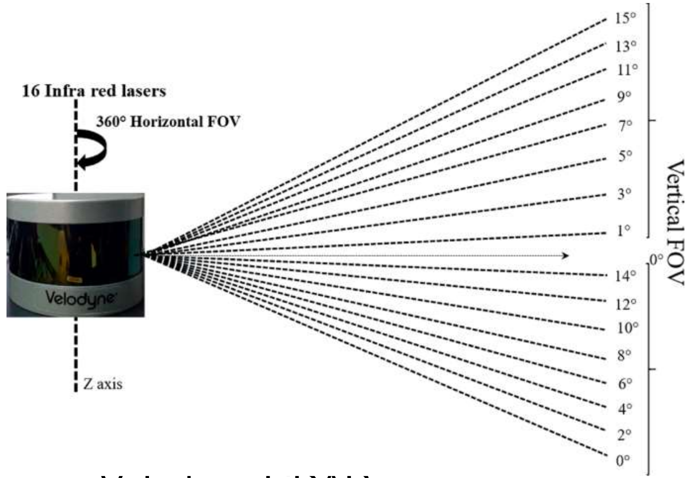
Stereo Visual Odometry



LIDAR Odometry and Mapping

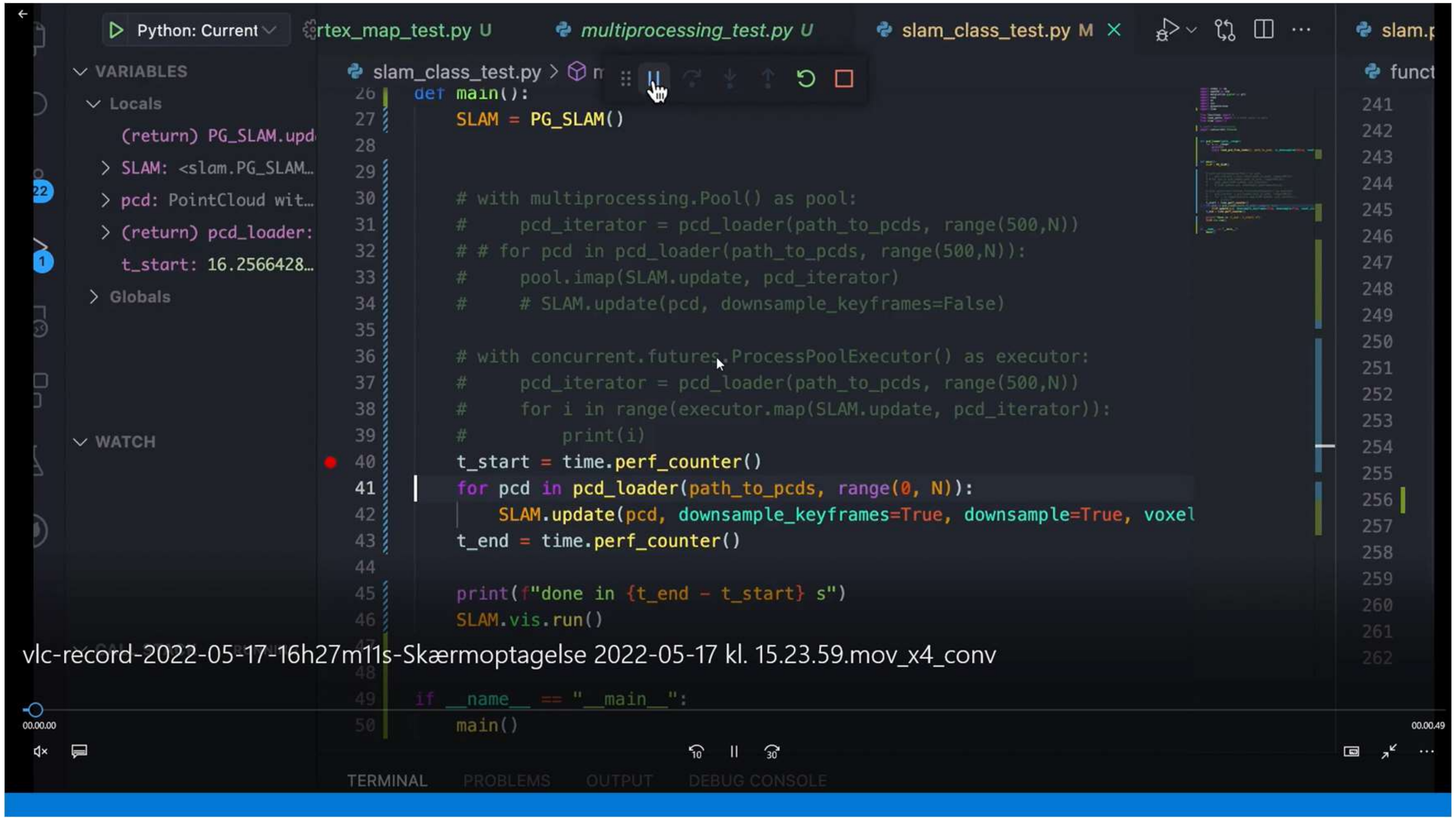
- LiDAR Odometry and Mapping is a family of algorithms that performs concurrent 3D mapping and Odometry

- LOAM algorithms can also incorporate loop-closures if the same points are crossed more than once, which can limit drift of the odometry estimates.



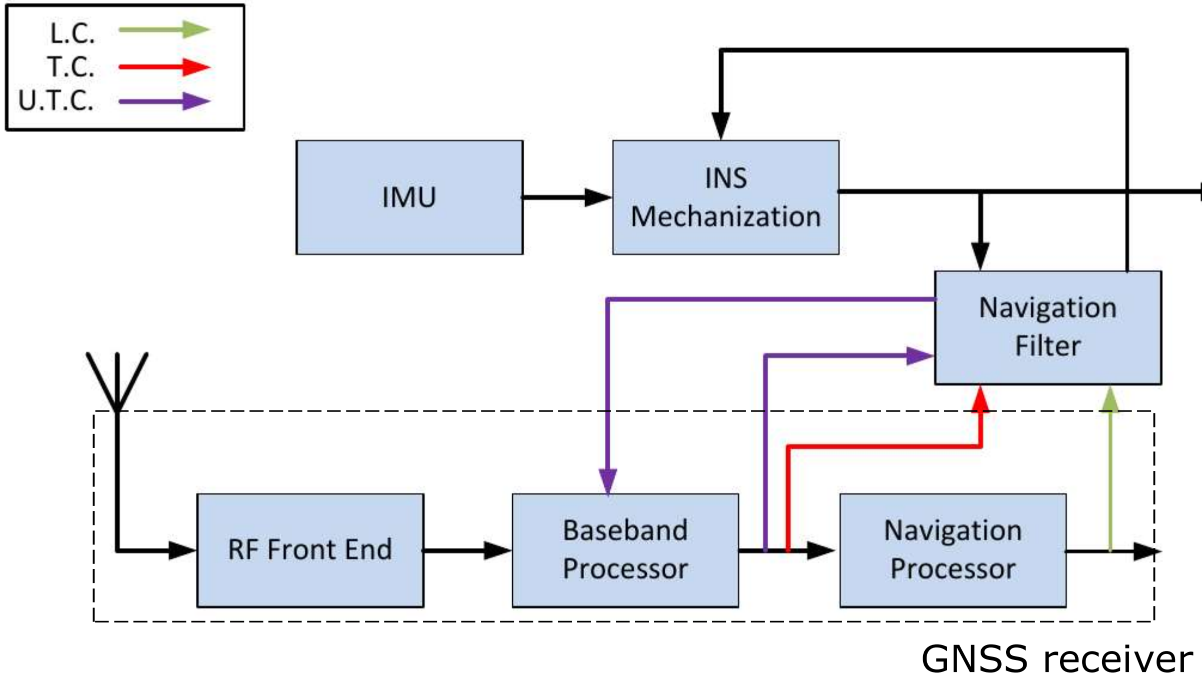
Velodyne LiDAR

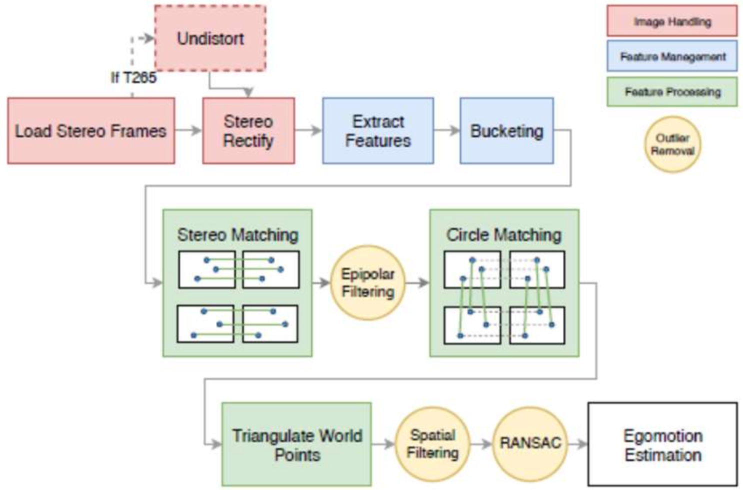
LIDAR Odometry and Mapping (LOAM)



Jakob Hedemann, DTU Space

GNSS, INS and VO integration





GNSS, INS and VO integration

- GNSS, INS and VO is complimentary systems and all has their own benefits and drawbacks
  - Inertial Navigation Systems (INS) are excellent for short-term (relative) navigation but is heavily degraded over longer timer intervals
  - GNSS performs very well in uninterrupted outdoors environments, but is troublesome in e.g. forested environments or urban canyons
  - VO is typically very good when features are static and light-conditions are optimal. Exhibits a linear drift over time, as translation and rotation errors eventually will build up

Other courses

- 30554 Global Navigation Satellite Systems
  - Understand how a GNSS receicer works (RF signals, baseband processing and positioning algorithms)
  - Learn how to determine a position from GNSS measurements
  - Apply different positioning techniques, code- and carrier based measurement to obtain accuracies down to centimeters.
  - Introduction to sensor-fusion using Kalman filters

Selected project proposals

- Visual inertial navigation –> Fusing IMU measurements with monocular or stereo cameras for robust positioning in GNSS degraded environments for UGVs and UAVs
- Upward facing fisheye camera for automatic classification of LOS GNSS signals
  - Automatic classification of sky vs obstacles
  - Create a 3D model of surrounding using images
  - Monocular VO
  - SLAM using existing 3D models
- Upward facing fisheye camera for tracking Aurora borealis in order to compensate ionospheric delays in GNSS signals
- Develop vision and sensor fusion techniques for a UAV to follow the center-line of a river