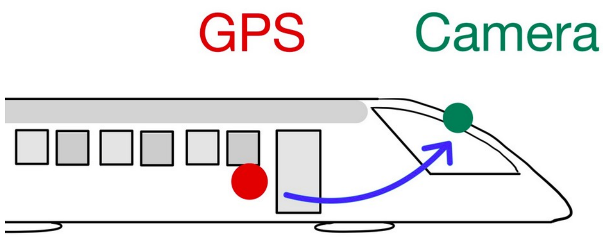
Semester Thesis

Professor Dr. Siegwart
Supervisors: Cornelius von Einem, David Hug
Nicolina Spiegelhalter
09.01.2023

Goal of Semester Thesis

- Online estimation of 6 DOF relative pose of camera to GPS, leveraging
  - Visual information
  - GPS and IMU measurements
  - Map information



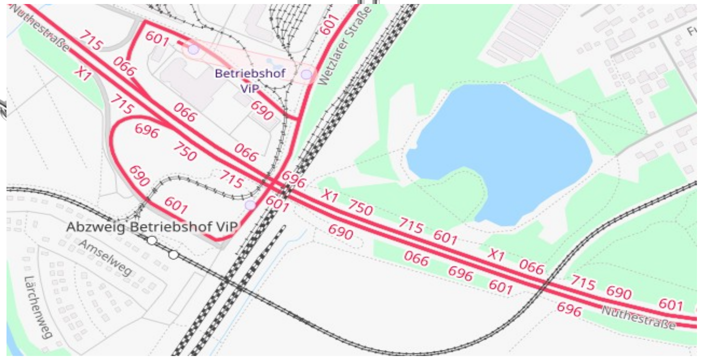
Img 1: Transformation between GPS and camera

- Purpose:
  - No additional prior calibration process
  - Labeled data generation for existing datasets
    - Finding ROI for obstacle detection

Img 2: Visual information

```
1 heading: -30.07068799440944
2 p_x: 371444.4318821279
3 p_y: 5804239.677466788
4 p_z: 40.456092834472656
5 q_w: 0.9657659850414041
6 q_x: 0.0
7 q_y: 0.0
8 q_z: -0.25941484563726575
```

Img 3: GPS and IMU measurements of train pose



Img 4: Map information on railway tracks

Approach based on literature review

- State of the art:
  - Pose estimate with calibration aids [1]
  - Solely rely on vanishing points if three are visible [2]

- Approach:

1)

Railway detection on straight tracks

→

2)

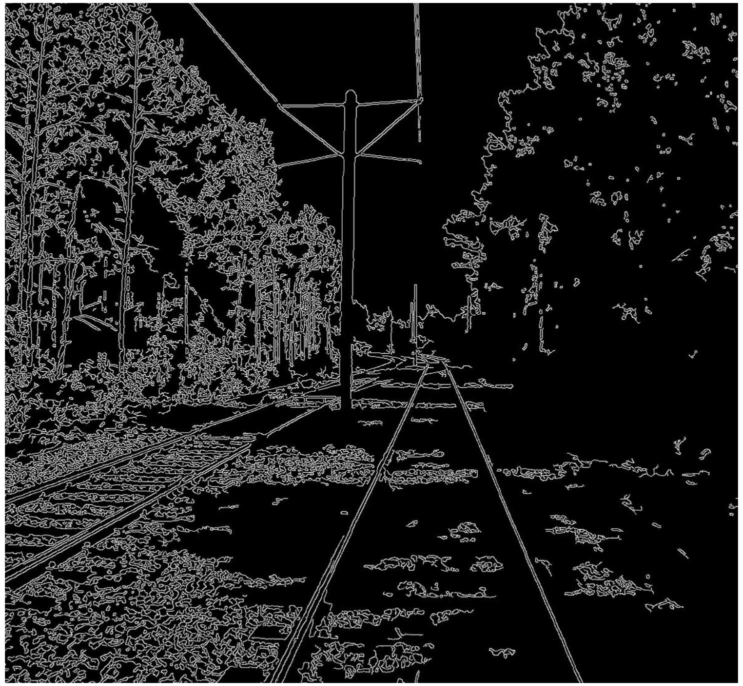
Multistaged parameter estimation:

- Leverage properties of vanishing point [3]
- Leverage detection of distinct landmarks [4]

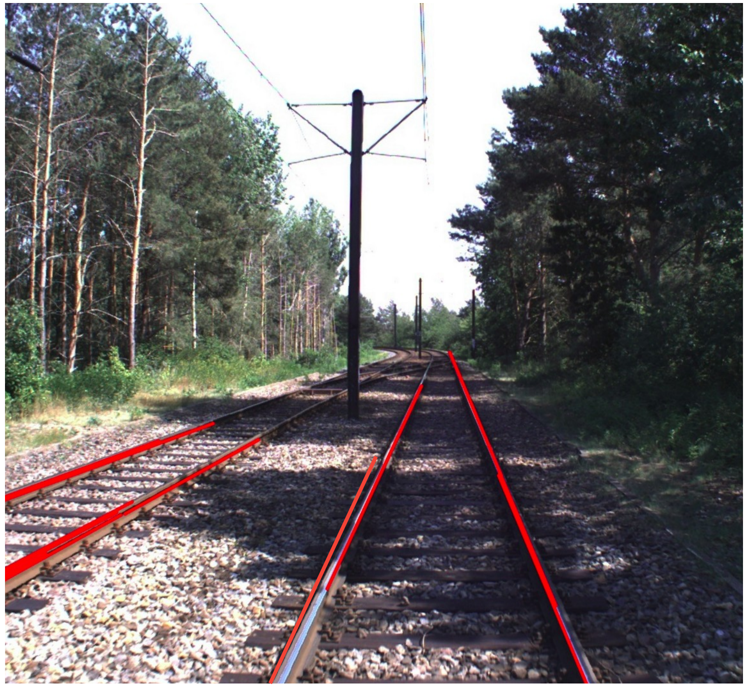
1. An Inverse Projective Mapping-based Approach for Robust Rail Track extraction, *Wang et al, 2015*
2. Using vanishing points for camera calibration, *Caprile, 1990*
3. Vision-based track estimation and turnout detectio using recursive estimation, *Ross, 2010*
4. Geolocation estimation of target vehicles using image processing and geometric computation, *Namazi, 2022*

Railway detection

1) Canny Edge Detection



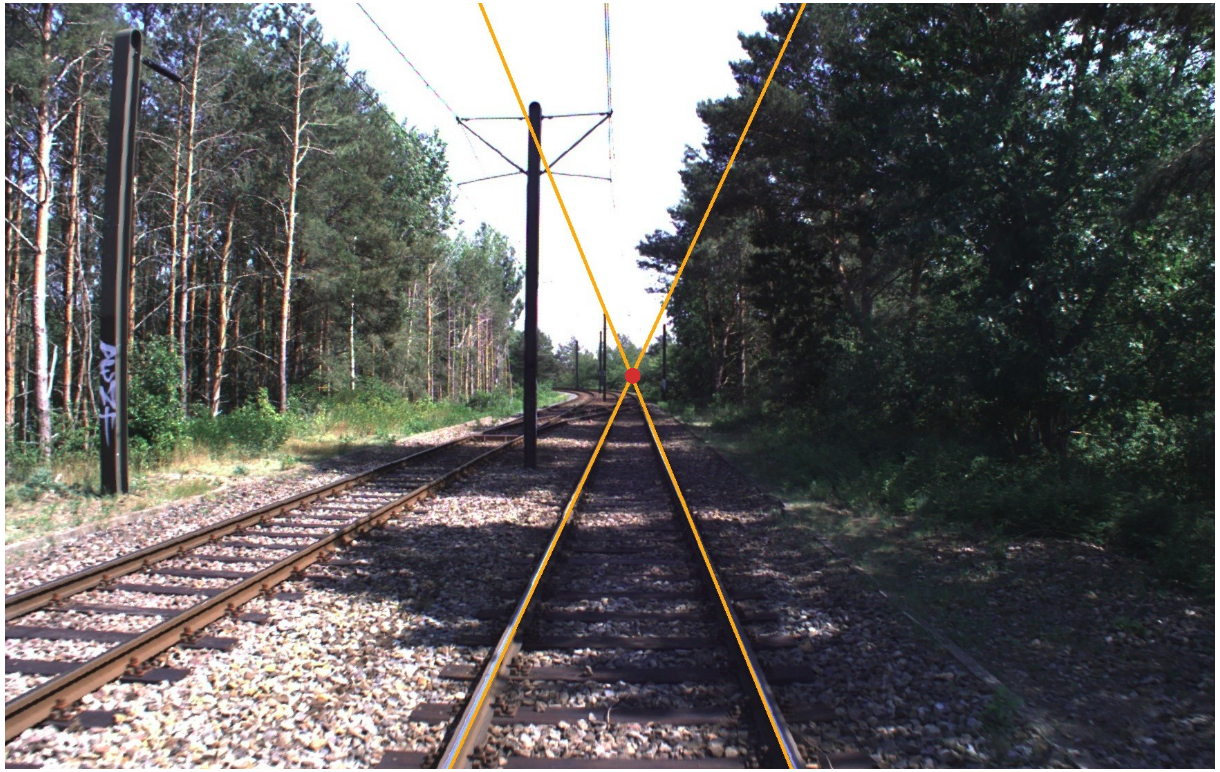
2) Probabilistic Hough Transform



3) Geometrical prior knowledge to filter for correct lines

Vanishing point

- Projective geometry:
  - parallel lines in world converge to one point in image
  - Three vanishing points – one per axis
- Tracks converge to one horizontal vanishing point



Img 1: Vanishing point at crossection of railway tracks

Order of Multistaged Parameter Estimation

Pitch

Yaw

Roll

↓

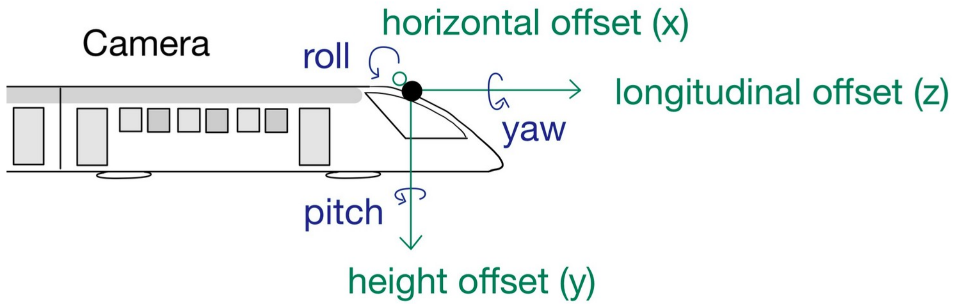
Height offset

↓

Horizontal offset

↓

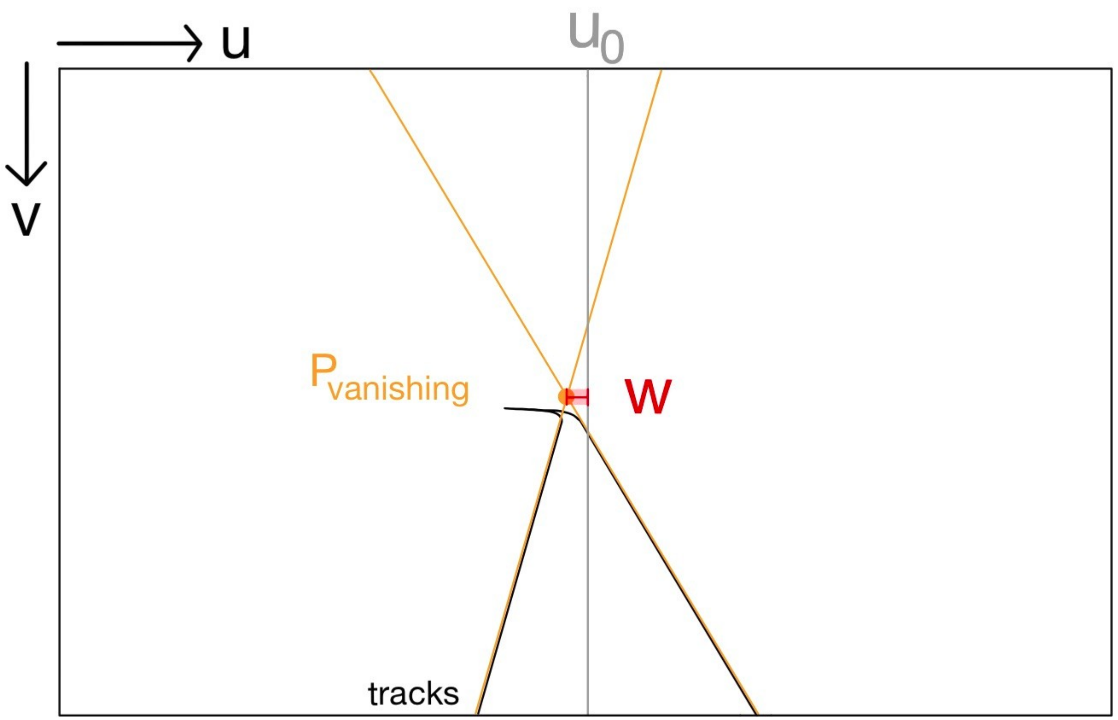
Longitudinal offset



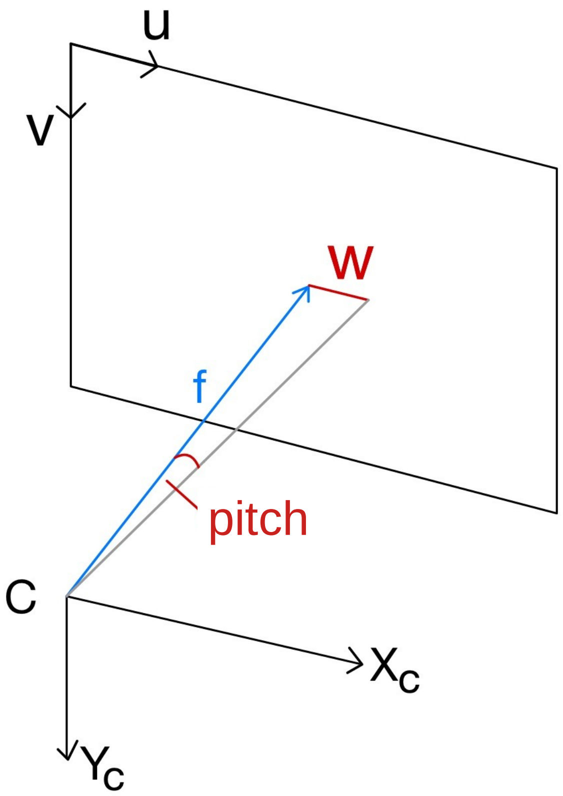
Img 1: Axes label visualization for camera

Pitch estimate

- Leverage vanishing point



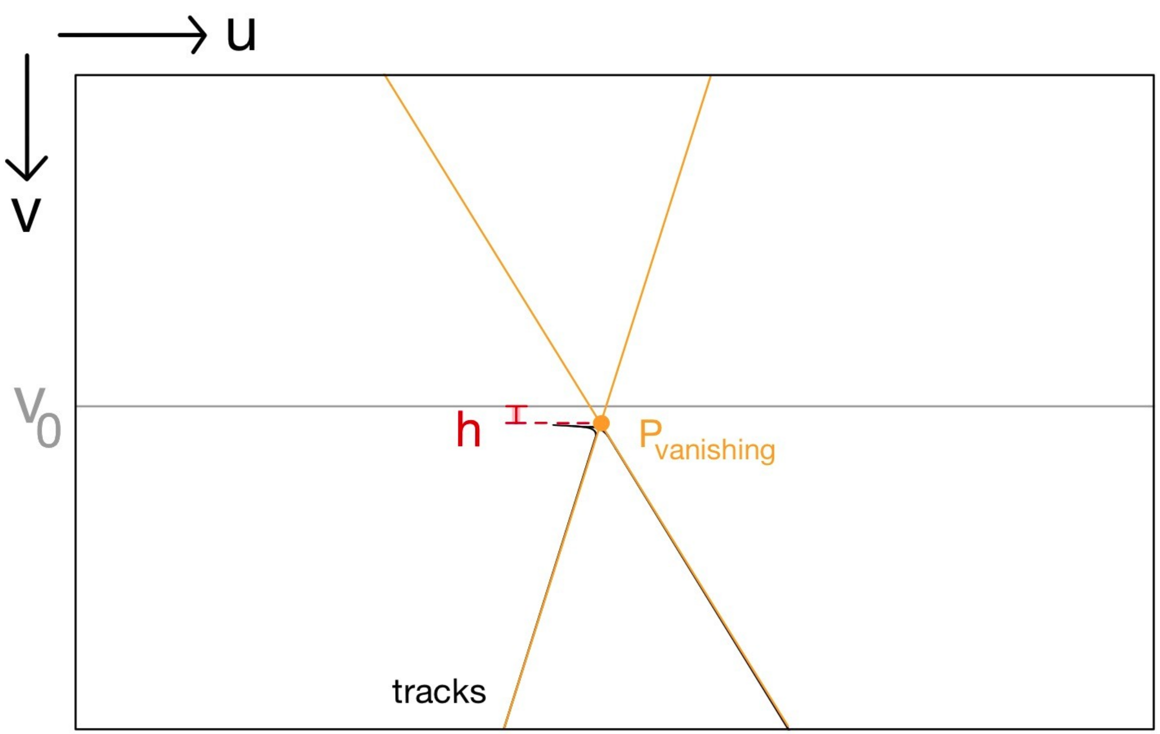
Img 1: Visualization of vanishing point and offset *w*



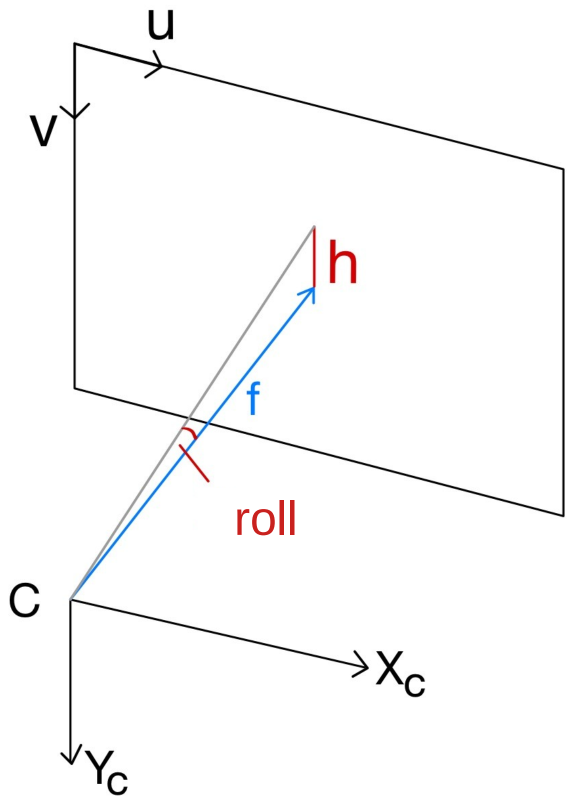
Img 2: Visualization of focal length *f* and offset *w*

Roll estimate

- Leverage vanishing point



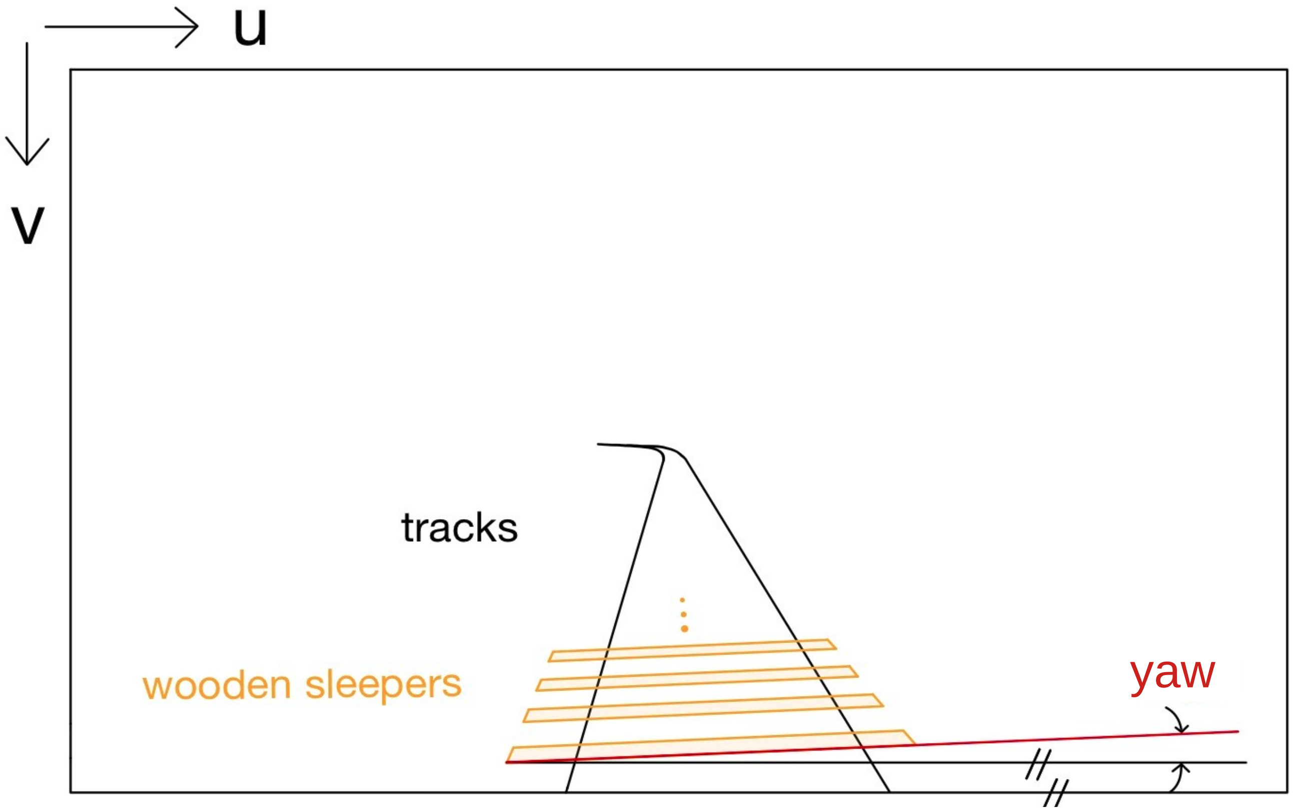
Img 1: Visualization of vanishing point and offset *h*



Img 2: Visualization of focal length *f* and offset *h*

Yaw estimate

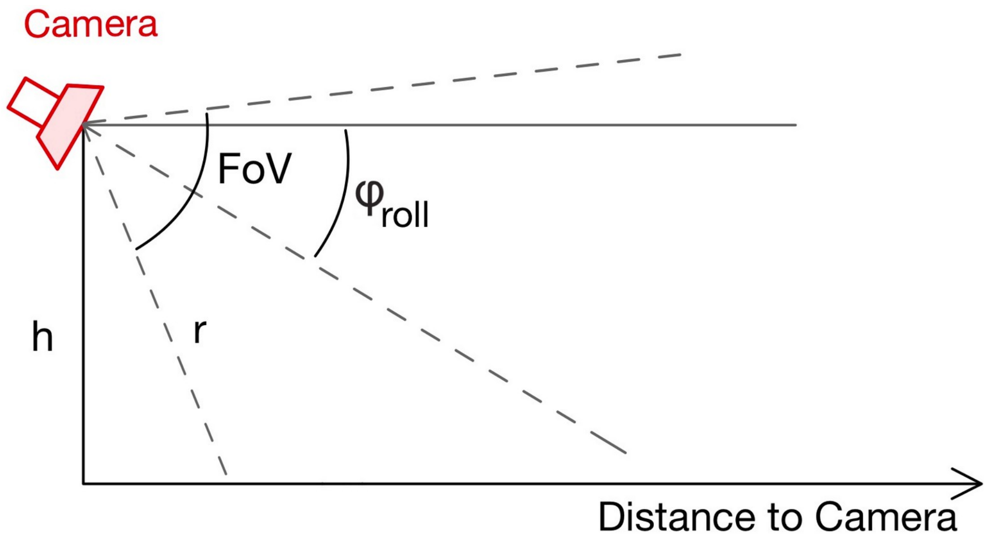
- Visual detection of wooden sleepers
- Compute the angle of their contours (in red)



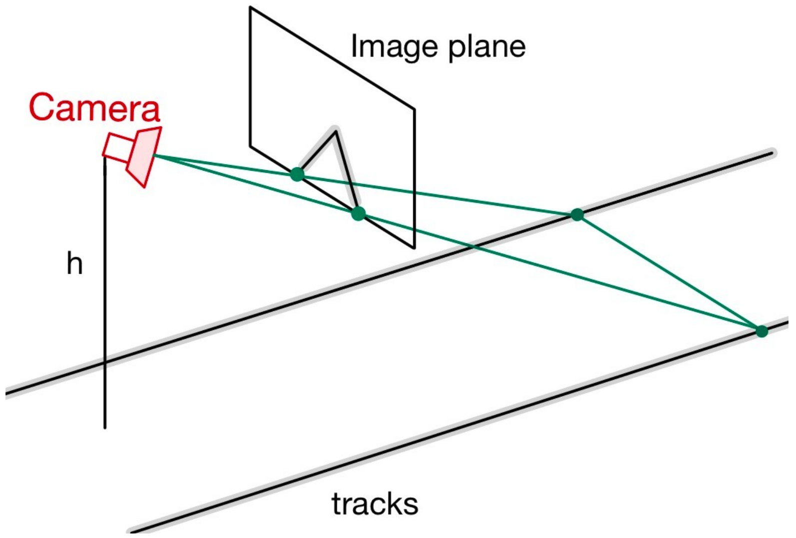
Img 1: Visualization of wooden sleepers between the tracks

Height offset estimate

- Leverage relation of perceived width of tracks to known real world width
- Trigonometry to determine height

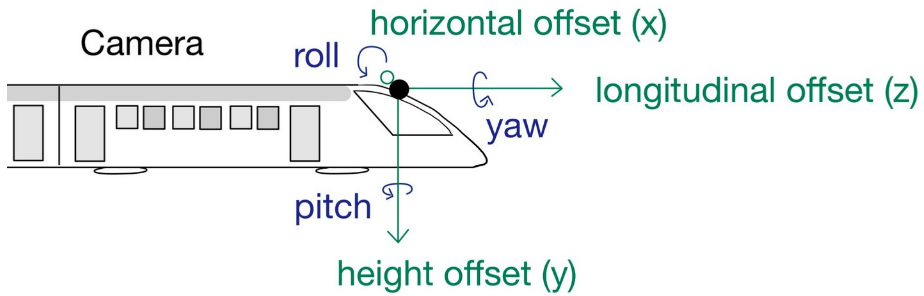


Img 1: Relation between parameters *r*, camera roll and height *h*



Img 2: Visualization of perceived width of tracks at bottom frame

Horizontal and longitudinal offset process



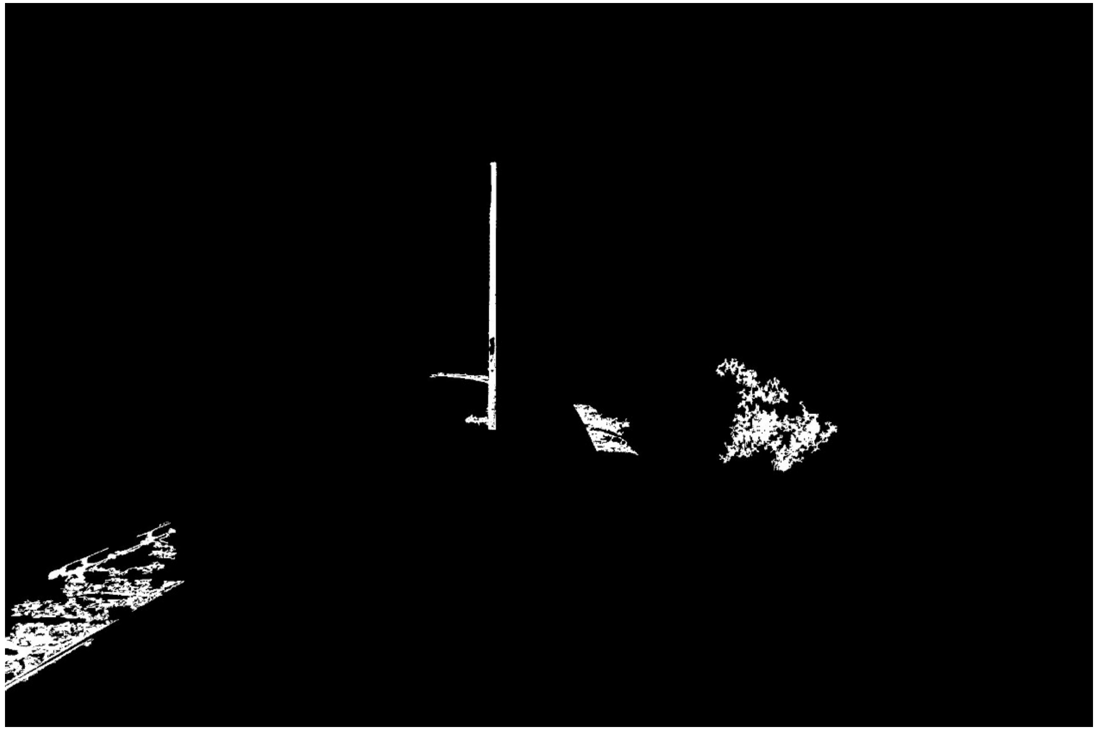
Img 1: Axes labels visualization

1) Distinctive landmark detection: electricity pole
2) Compute depth of bottom pixel (orange)
3) Convert bottom pixel into GPS frame
4) Compare visual detection to known location

Img 2: Electricity poles visual in image with ground contact point highlighted

Pole detection

- Image processing
  - HSV color filters
  - Component analysis for large areas
- Identify pole as component with vertical line
- Ground contact point: lowest pixel of pole



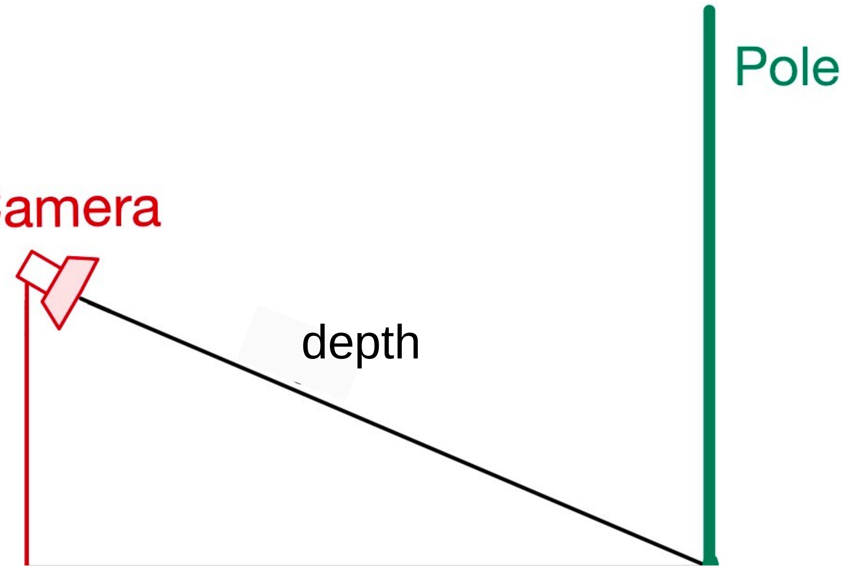
Img 1: Image after color-filtering and component analysis for larger areas

Depth computation[1]

- Extract equation from homogeneous transformation matrix

$$\begin{pmatrix} X_w \\ Y_w \\ \boxed{Z_w} \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 \\ h_5 & h_6 & h_7 & h_8 \\ h_9 & h_{10} & h_{11} & h_{12} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

- Z_W = floor elevation of ground contact point
- Solve for depth Z_C

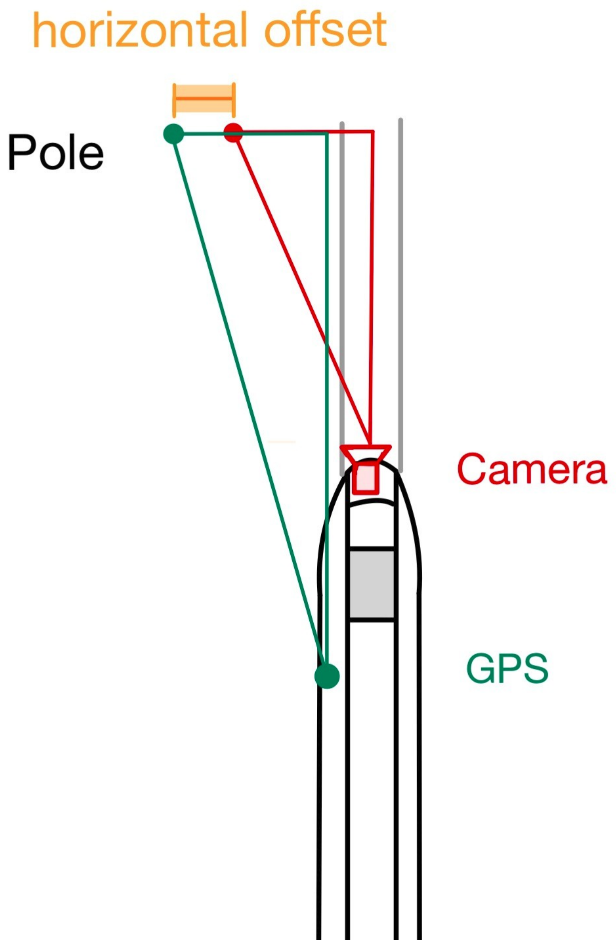


Img 1: Geometry used to compute depth

1. Geolocation estimation of target vehicles using image processing and geometric computation, *Namazi, 2022*

Horizontal offset

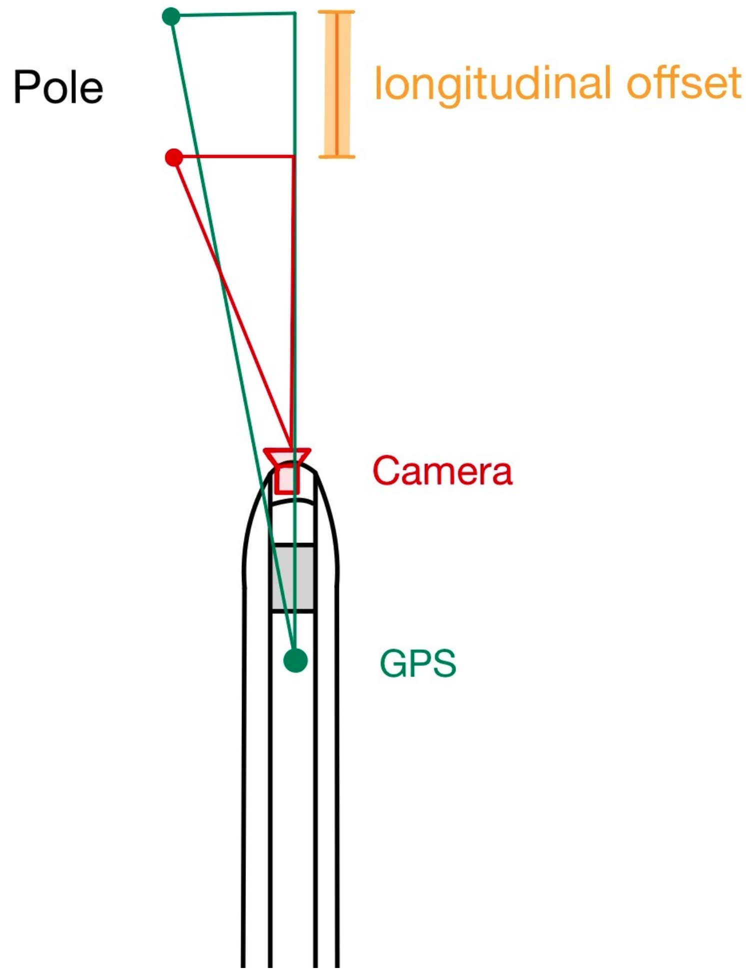
- Conversion of bottom pixel into GPS frame
- Horizontal difference of visual detection to known location



Img 1: Difference in pole perception by camera to known location

Longitudinal offset

- After succesful pole detection and depth computation
- Conversion of bottom pixel into GPS frame
- Longitudinal difference of visual detection to known location



Img 1: Difference in pole perception by camera to known location

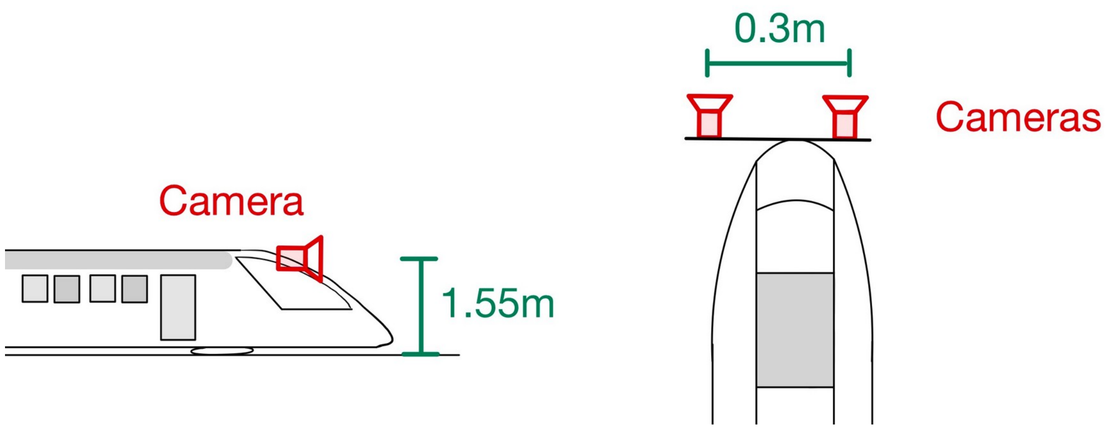
Results

- Camera setup at front of train
- Comparison of my translational and rotational estimates to prior knowledge

Img 1: Camera setups at the front of the train

Results

- Camera setup at front of train
- Comparison of my translational estimates to prior knowledge

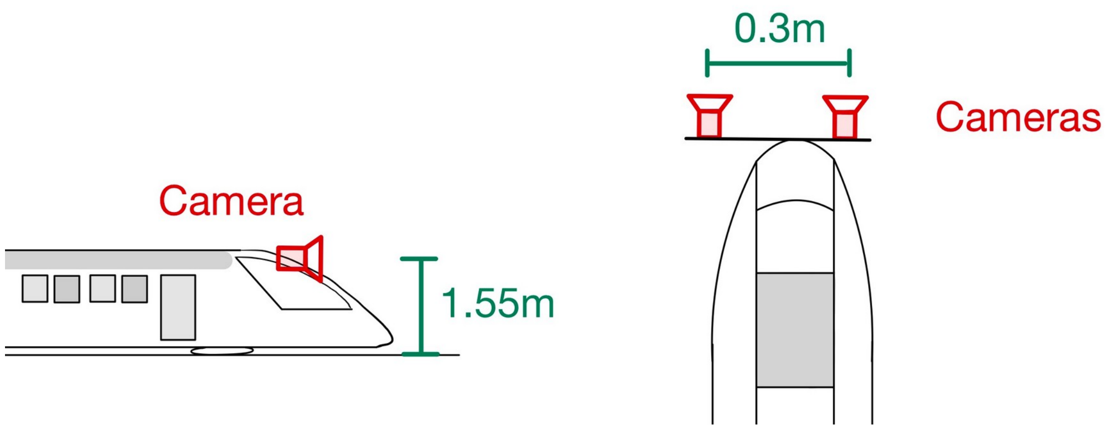


Img 1: a) Side view of camera setup in front of train b) Top view of camera setup

Estimate medians wrt GPS	Horizontal offset	Height	Longitudinal offset
Right camera	1.34m	1.47m	12.81m
Left camera	1.19m	1.5m	10.76m

Results

- Camera setup at front of train
- Comparison of my translational estimates to prior knowledge



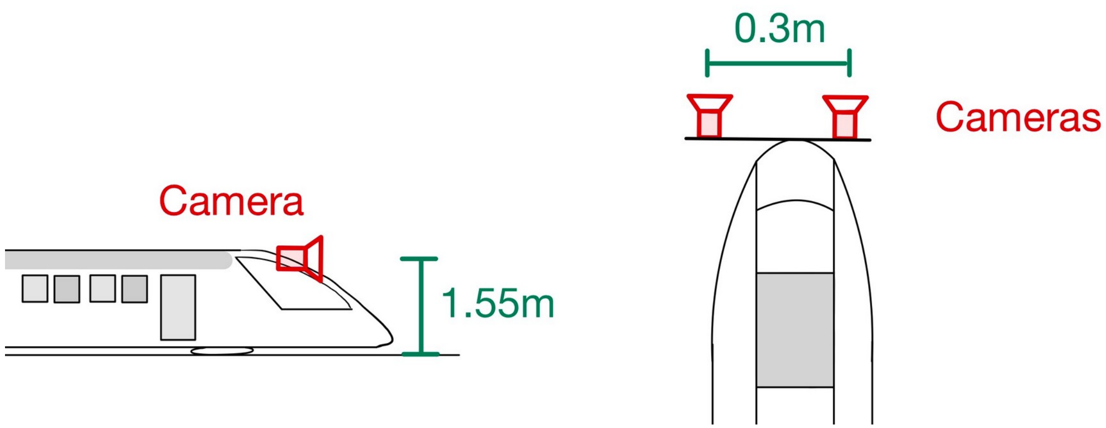
Img 1: a) Side view of camera setup in front of train b) Top view of camera setup

Estimate medians wrt GPS	Horizontal offset	Height	Longitudinal offset
Right camera	1.34m	1.47m	12.81m
Left camera	1.19m	1.5m	10.76m

Distance between cameras from my estimates: 0.15m
Ground truth: 0.3m

Results

- Camera setup at front of train
- Comparison of my translational estimates to prior knowledge



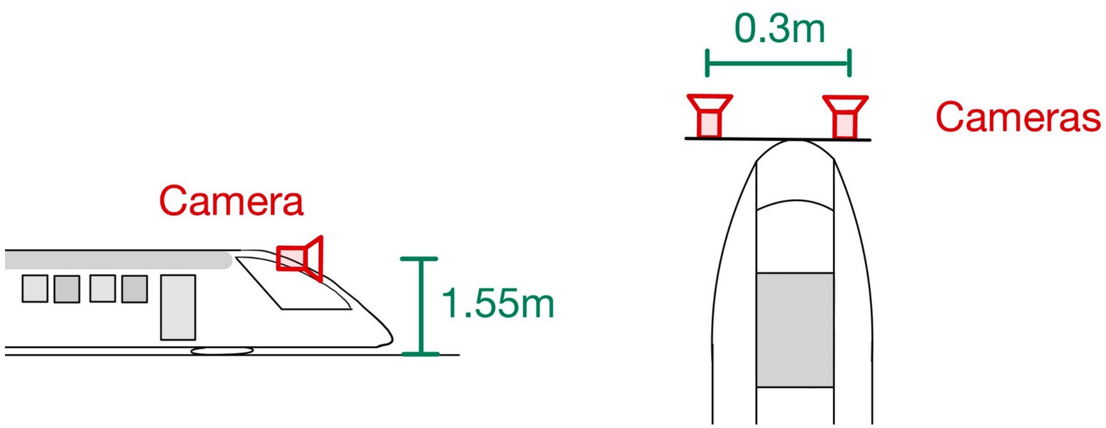
Img 1: a) Side view of camera setup in front of train b) Top view of camera setup

Estimate medians wrt GPS	Horizontal offset	Height	Longitudinal offset
Right camera	1.34m	1.47m	12.81m
Left camera	1.19m	1.5m	10.76m

Ground truth: 1.55m

Results

- Camera setup at front of train
- Comparison of my translational estimates to prior knowledge



Img 1: a) Side view of camera setup in front of train b) Top view of camera setup

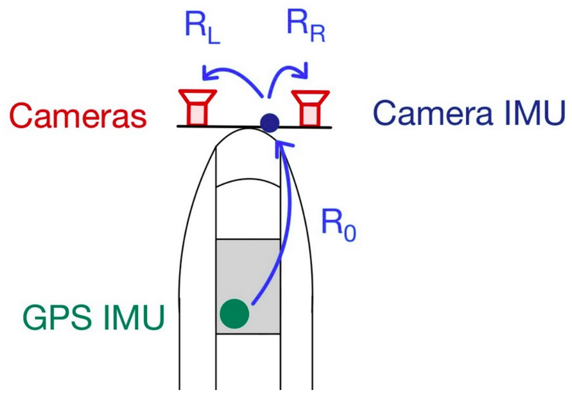
Estimate medians wrt GPS	Horizontal offset	Height	Longitudinal offset
Right camera	1.34m	1.47m	12.81m
Left camera	1.19m	1.5m	10.76m

Ground truth: should have same values

Results

- Camera setup at front of train

- Comparison of my rotational estimates to:
  - values obtained by rotation from GPS IMU to camera via camera IMU
  - Relative rotation between cameras

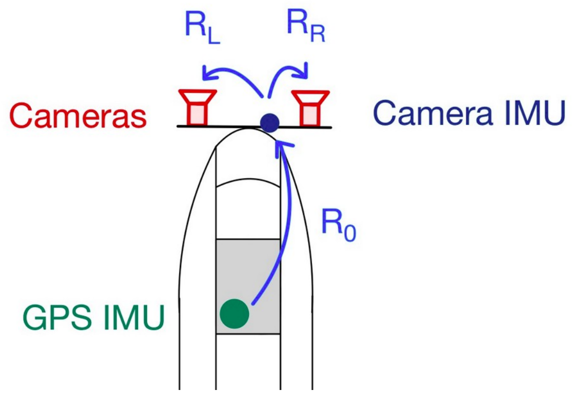


Img 1: Setup with GPS IMU and camera IMU

Estimate medians wrt GPS	Roll	Pitch	Yaw
Right camera	-90.0°	89.49°	0.42°
Left camera	-90.02°	91.58°	-0.65°

Results

- Camera setup at front of train
- Comparison of my rotational estimates to:
  - Euler angles from rotation from GPS IMU to camera



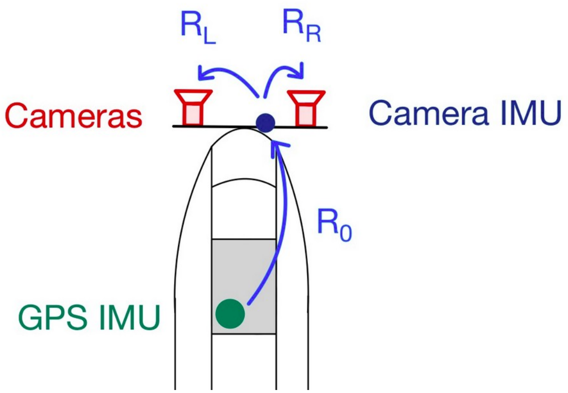
Img 1: Setup with GPS IMU and camera IMU

Estimate medians wrt GPS	Roll	Pitch	Yaw
Right camera	-90.0°	89.49°	0.42°
Left camera	-90.02°	91.58°	-0.65°

Summed difference: 1.1°

Results

- Camera setup at front of train
- Comparison of my rotational estimates to:
  - Euler angles from rotation from GPS IMU to camera



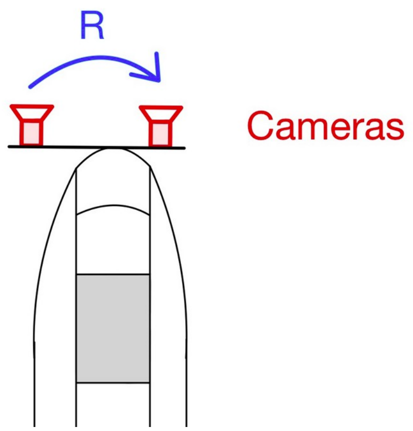
Img 1: Setup with GPS IMU and camera IMU

Estimate medians wrt GPS	Roll	Pitch	Yaw
Right camera	-90.0°	89.49°	0.42°
Left camera	-90.02°	91.58°	-0.65°

Summed difference: 2.44°

Results

- Camera setup at front of train
- Comparison of my rotational estimates to:
  - Relative rotation between cameras



Img 1: Relative rotation between cameras

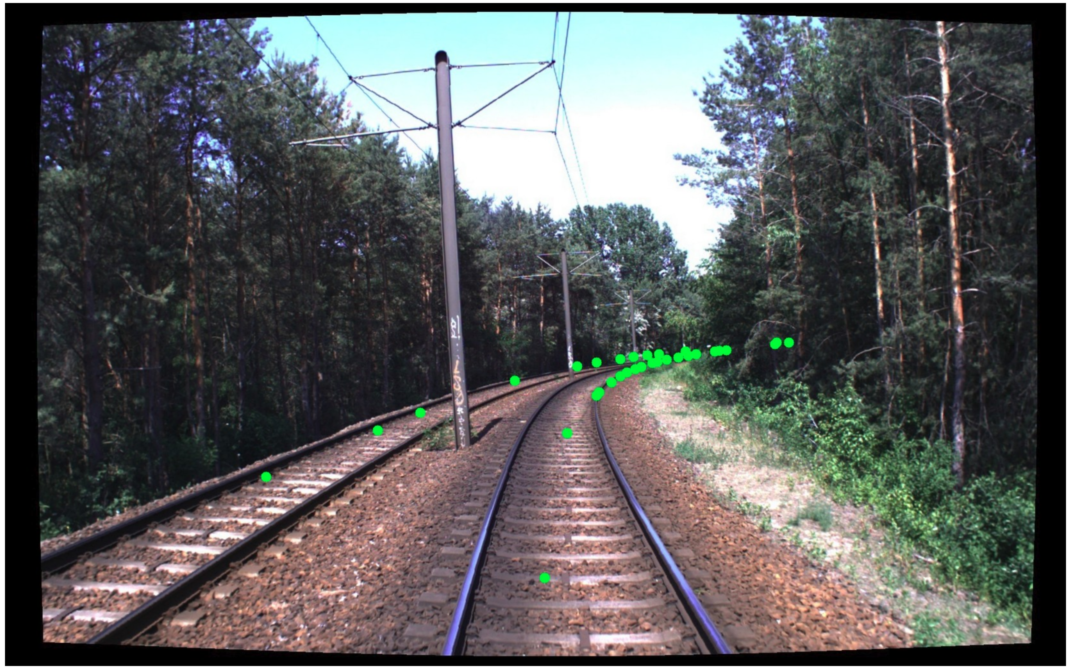
Estimate medians wrt GPS	Roll	Pitch	Yaw
Right camera	-90.0°	89.49°	0.42°
Left camera	-90.02°	91.58°	-0.65°

Summed difference: 0.46°

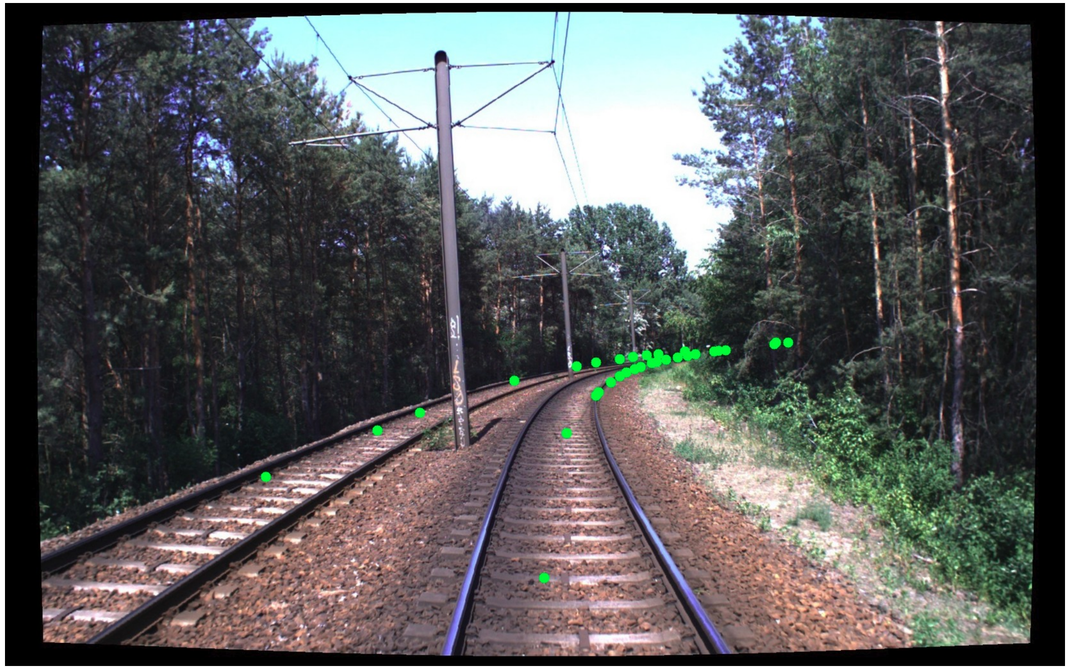
Results

- Projection of railway track nodes from OpenStreetMap into image frames:

- Left camera:



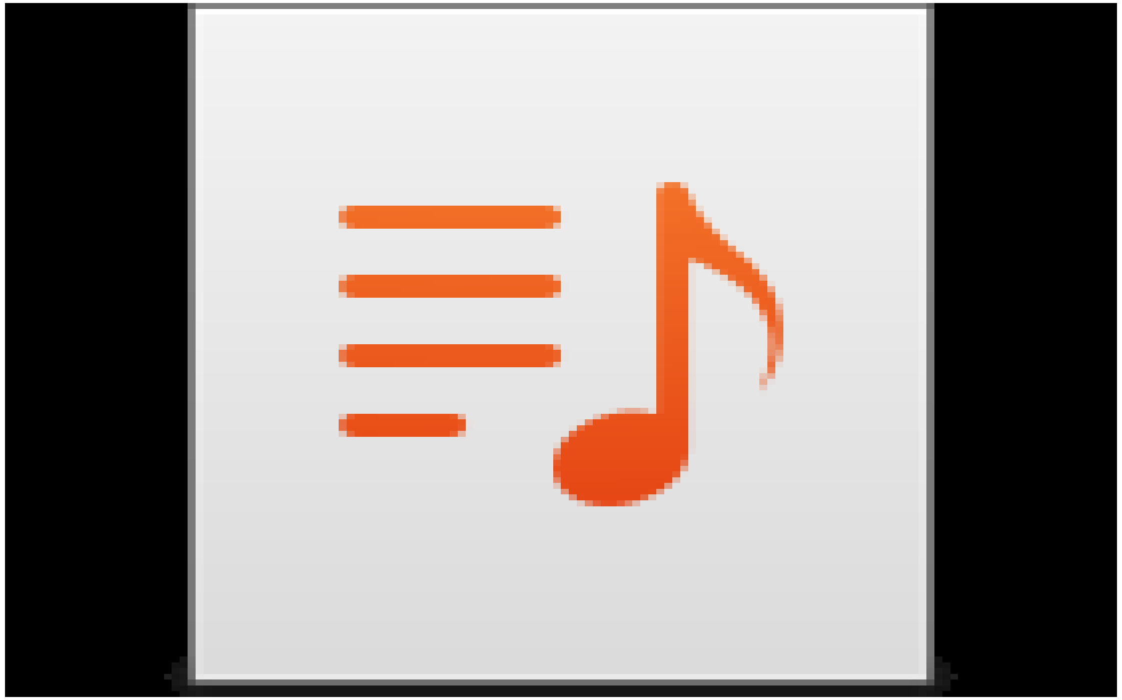
- Right camera:



Results

- Projection of railway track nodes from OpenStreetMap into image frames:

- Left camera:



- Right camera:

Conclusion

- Importance of accurate angles
- Dependencies have large impact
  - Camera height
  - Horizontal offset
  - Longitudinal offset
- Automate key frame selection with e.g. Schneider [1]

1. Visual-inertial Self-calibration on Informative Motion Segments, *Schneider et al, 2017*

Questions?

Nicolina Spiegelhalter
spiegeln@ethz.ch

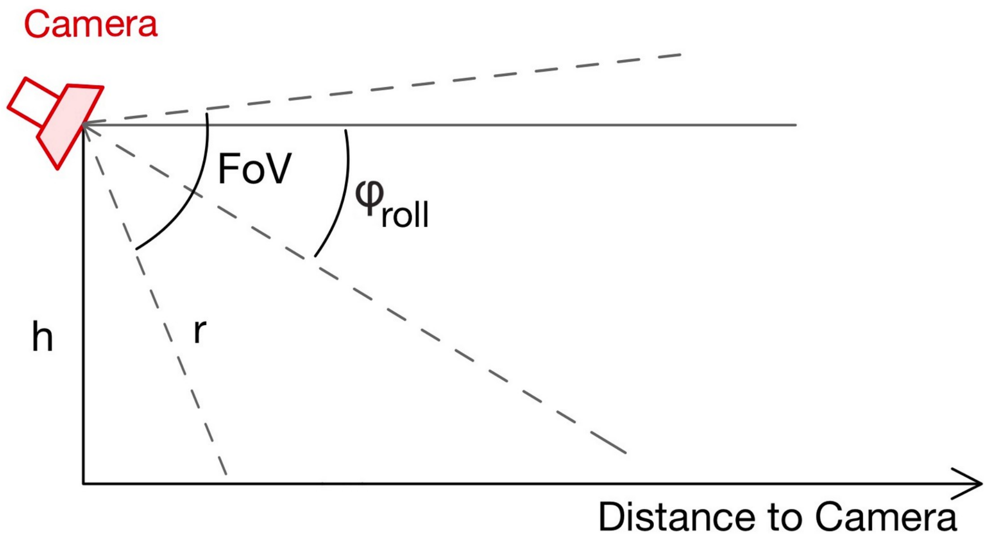
Height estimate detail

- Leveraging
  - Camera pitch: -x-rotation
  - opening angle of camera alpha = 46.4°
  - normed width of tracks G= 1.435m
  - perceived width of tracks in pixel w_0: 542 pixel

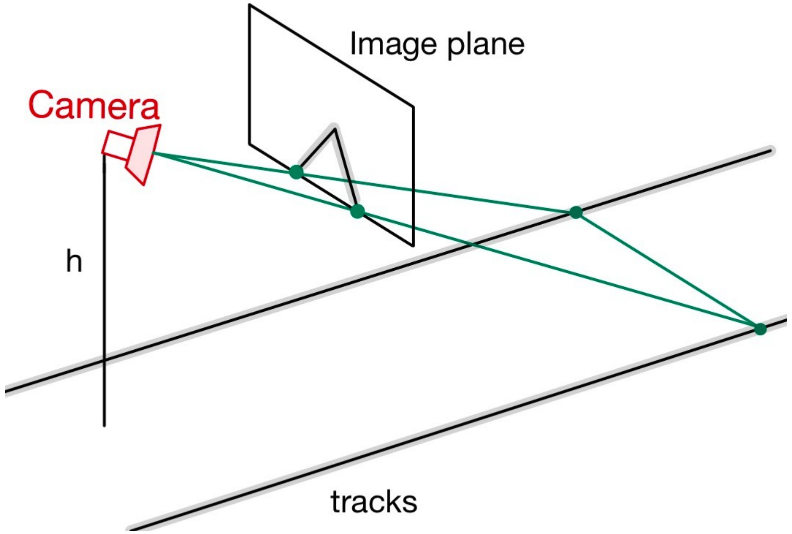
- Combination of following equations[1] to find h:

$$w_0 = \frac{f}{r} G \qquad g = \frac{h}{\tan(\phi_{\text{pitch}} + \frac{\alpha}{2})}.$$

$$r = (g^2 + h^2)^{1/2}$$



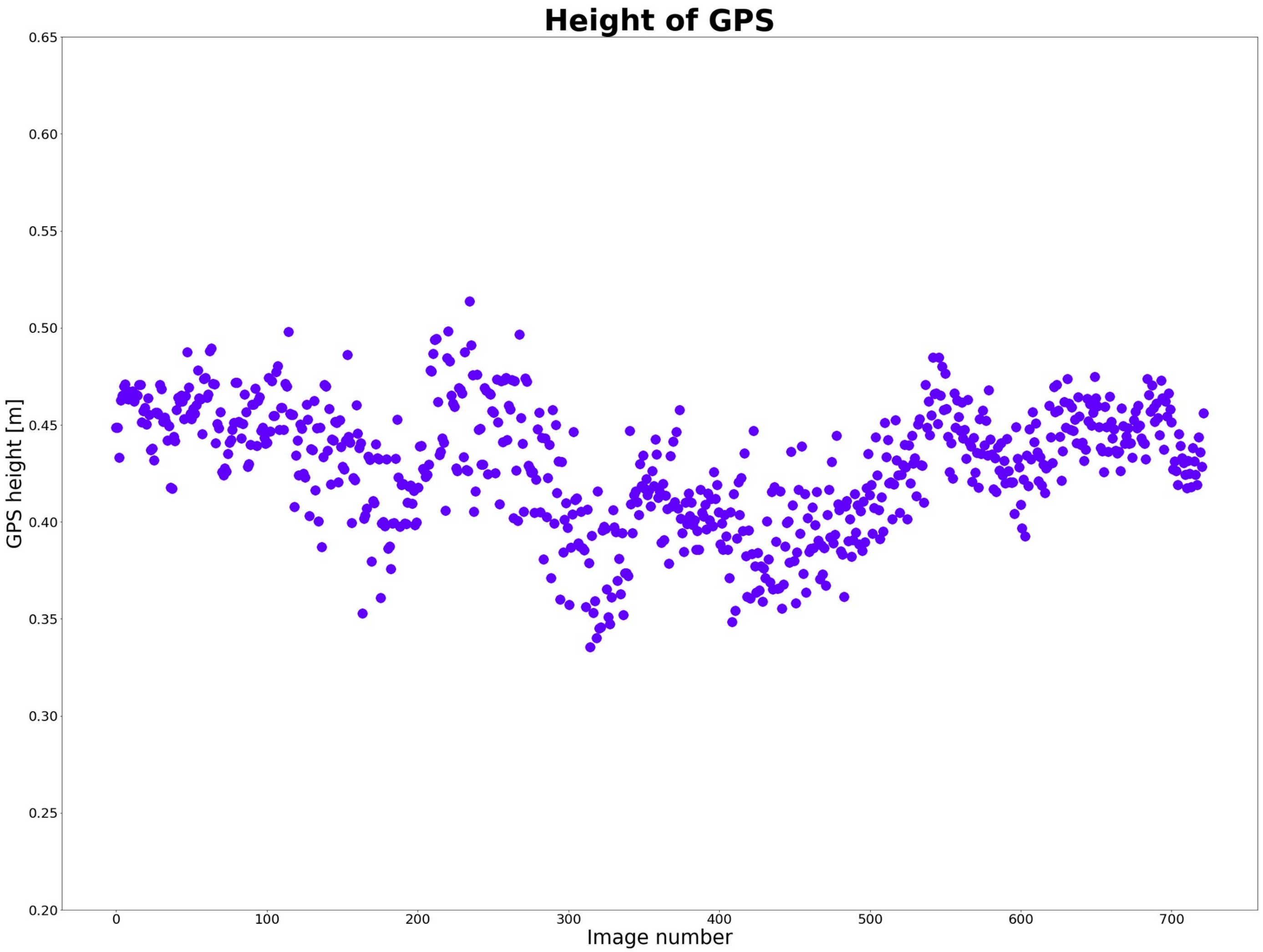
Img 1: Relation between parameters *r*, camera roll and height *h*



Img 2: Visualization of perceived width of tracks at bottom frame

GPS height consistency

- Consistency of RTK-GPS height with geobrooker.geobasis.de for elevation information:
  - Median: 0.432
  - Standard deviation: 0.03



Horizontal and longitudinal offset detail

- Express X_C, Y_C dependent of Z_C:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}$$

$$X_c = \frac{Z_c(u - u_0)}{\alpha_u} \quad Y_c = \frac{Z_c(v - v_0)}{\alpha_v} \quad \lambda = Z_c$$

- Use H_{w_cam} including estimates, substitute X_C, Y_C, Z_C for upper definitions.

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = H_{cam\_w}^{-1} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 \\ h_5 & h_6 & h_7 & h_8 \\ h_9 & h_{10} & h_{11} & h_{12} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

- Replace Z_w for known elevation at pole location, take third row and solve for Z_C:

$$Z_w = h_{elevation}$$

$$Z_c = (h_{elevation} - h_{12})/(h_9 \frac{(u - u_0)}{\alpha_u} + h_{10} \frac{(v - v_0)}{\alpha_v} + h_{11})$$