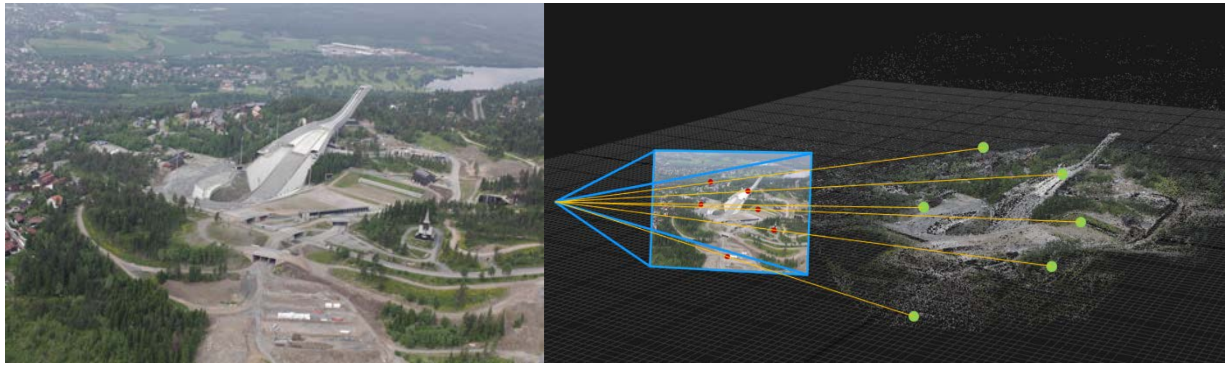
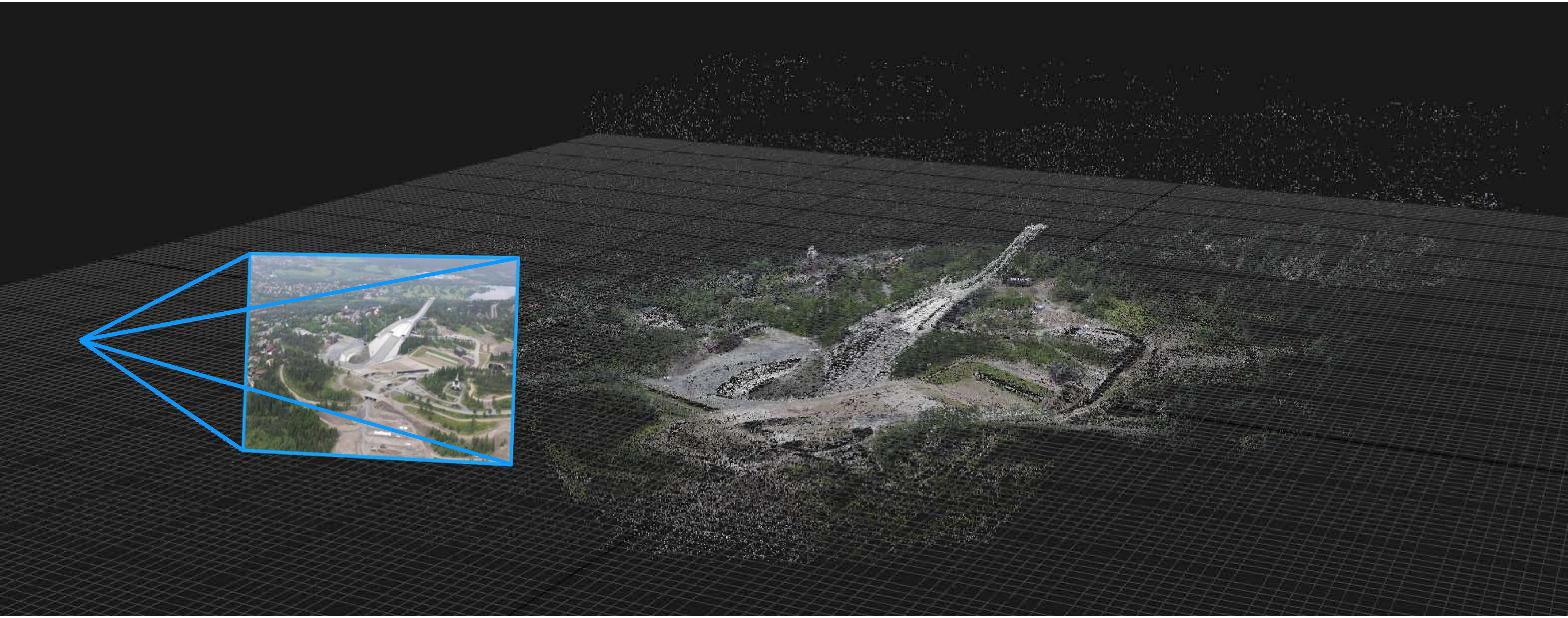
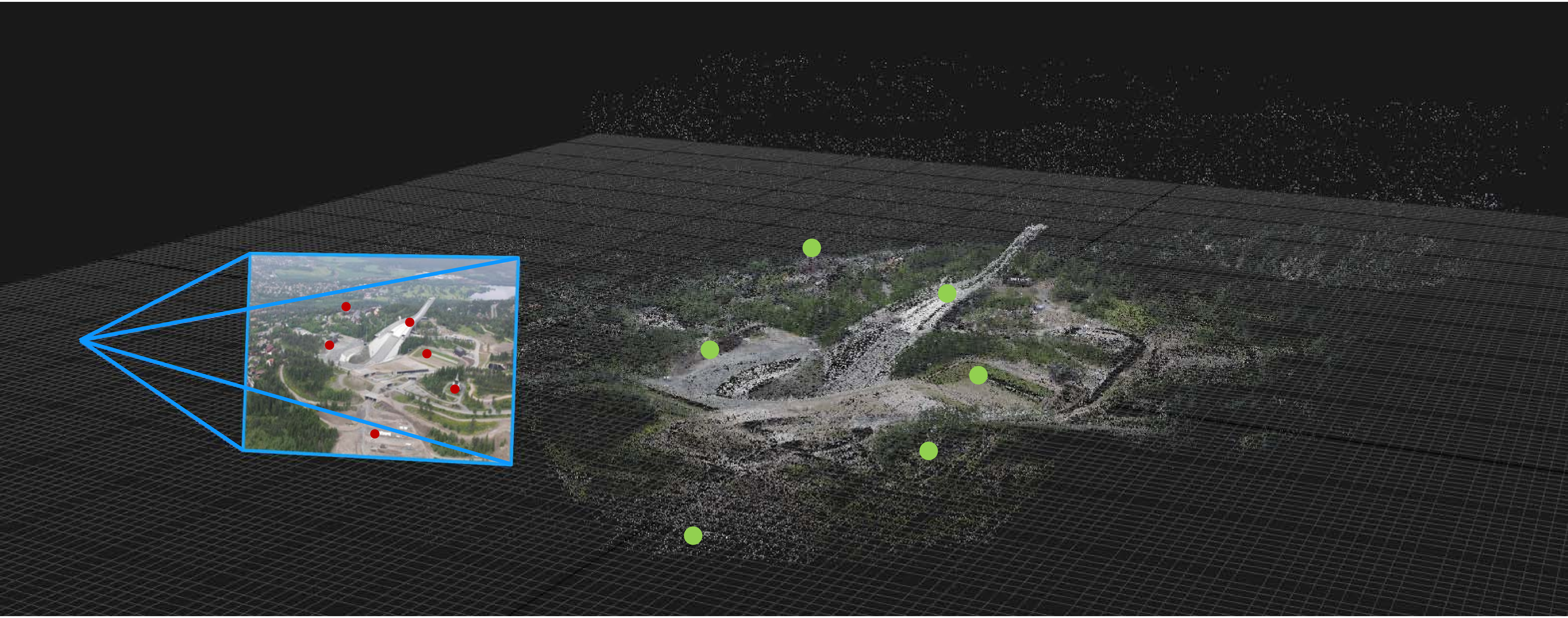
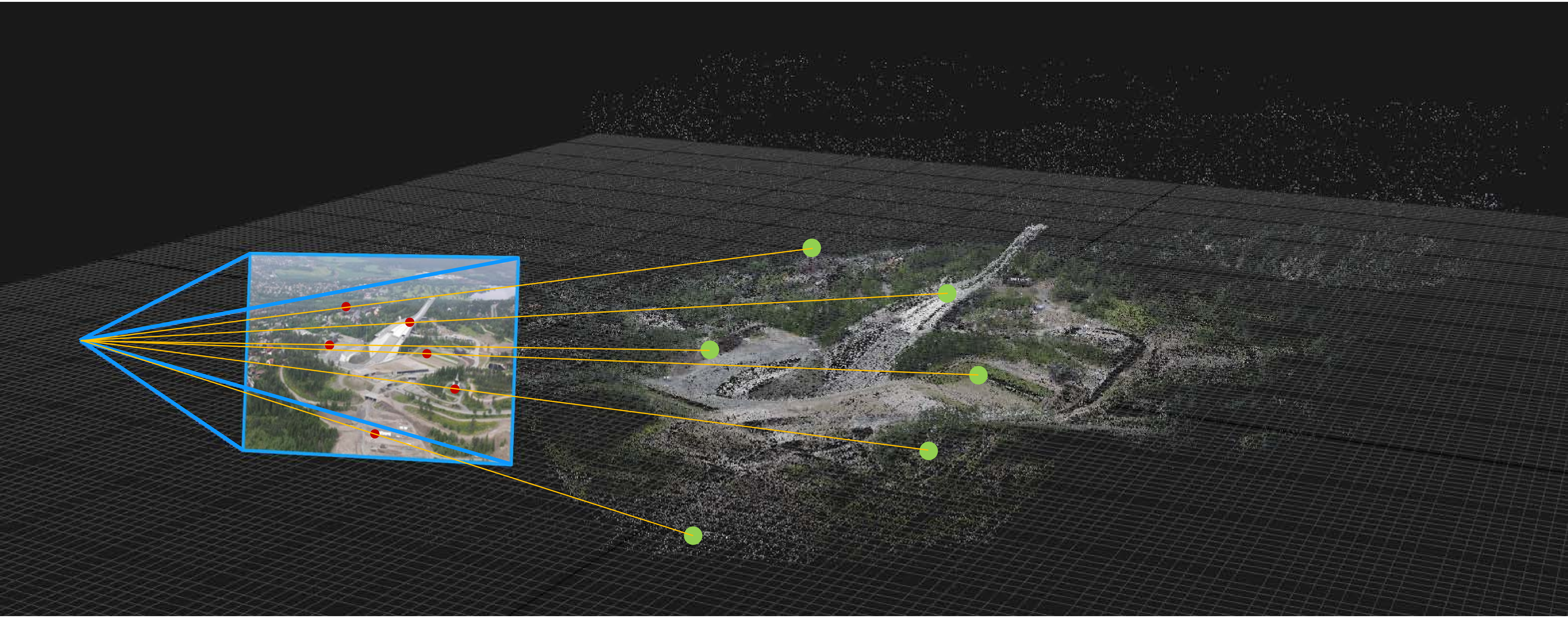


Lecture 5.2
Pose from known 3D points

Trym Vegard Haavardsholm







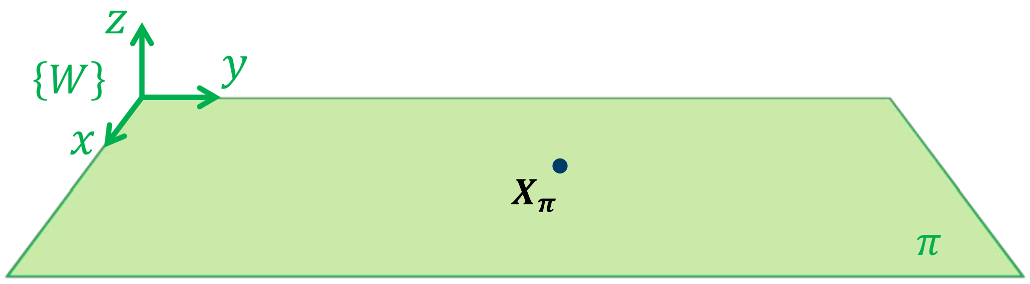
World geometry from correspondences

	Structure (scene geometry)	Motion (camera geometry)	Measurements
Pose estimation	Known	Estimate	3D to 2D correspondences
Triangulation, Stereo	Estimate	Known	2D to 2D correspondences
Reconstruction, Structure from Motion	Estimate	Estimate	2D to 2D correspondences

Pose estimation relative to a world plane

- Choose the world coordinate system so that the xy-plane corresponds to a plane π in the scene

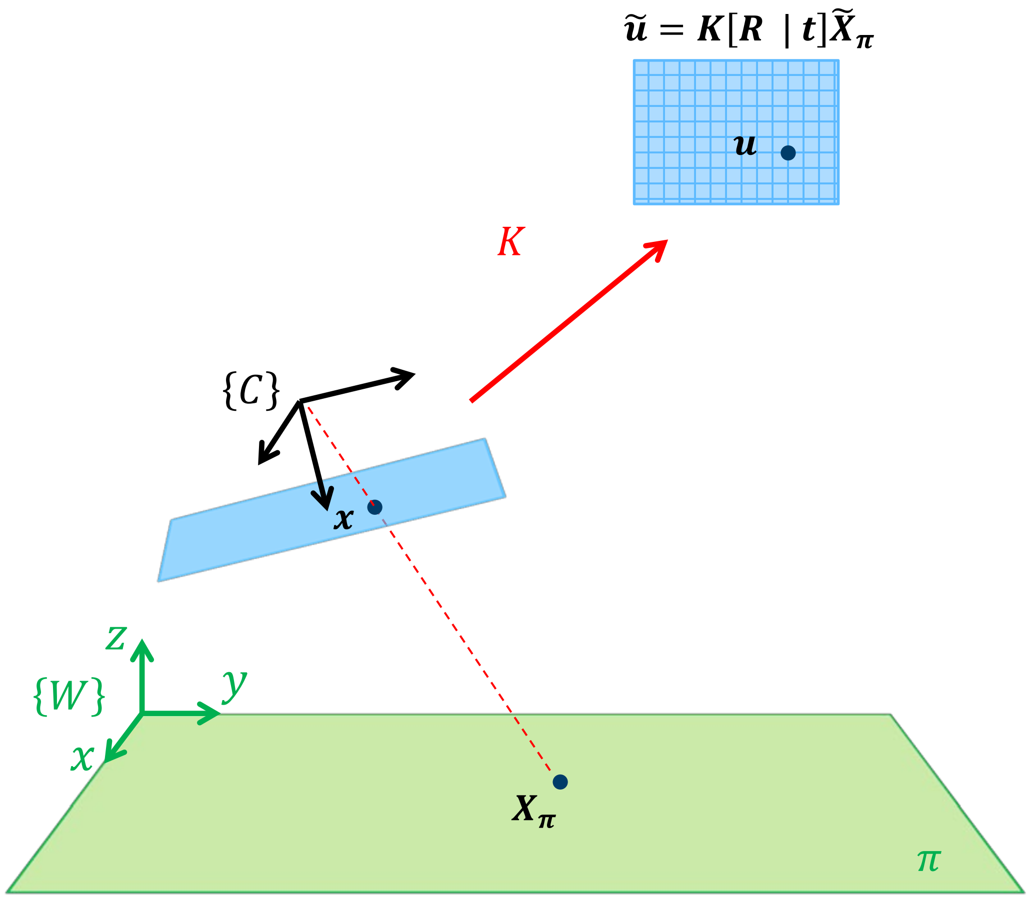
$$\boldsymbol{X}_\pi = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \qquad \boldsymbol{x}_\pi = \begin{bmatrix} x \\ y \end{bmatrix}$$



Pose estimation relative to a world plane

- We can map points on the world plane into image coordinates by using the perspective camera model

$$\tilde{\boldsymbol{u}} = \boldsymbol{K}\left[\boldsymbol{R} \mid \boldsymbol{t}\right]\tilde{\boldsymbol{X}}_{\pi} \qquad \boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$



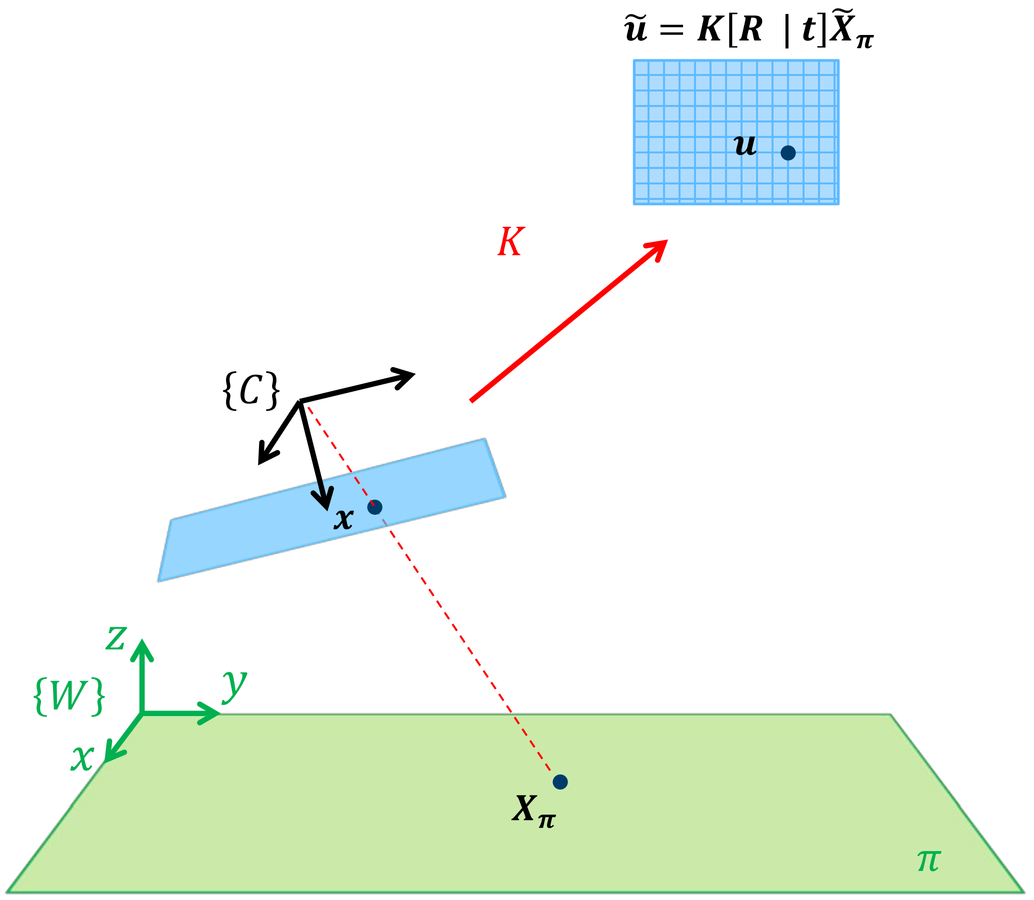
Pose estimation relative to a world plane

- We can map points on the world plane into image coordinates by using the perspective camera model

$$\tilde{\boldsymbol{u}} = \boldsymbol{K}\left[\boldsymbol{R} \mid \boldsymbol{t}\right]\tilde{\boldsymbol{X}}_{\pi}$$

$$= \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{t}\right]\begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$



Pose estimation relative to a world plane

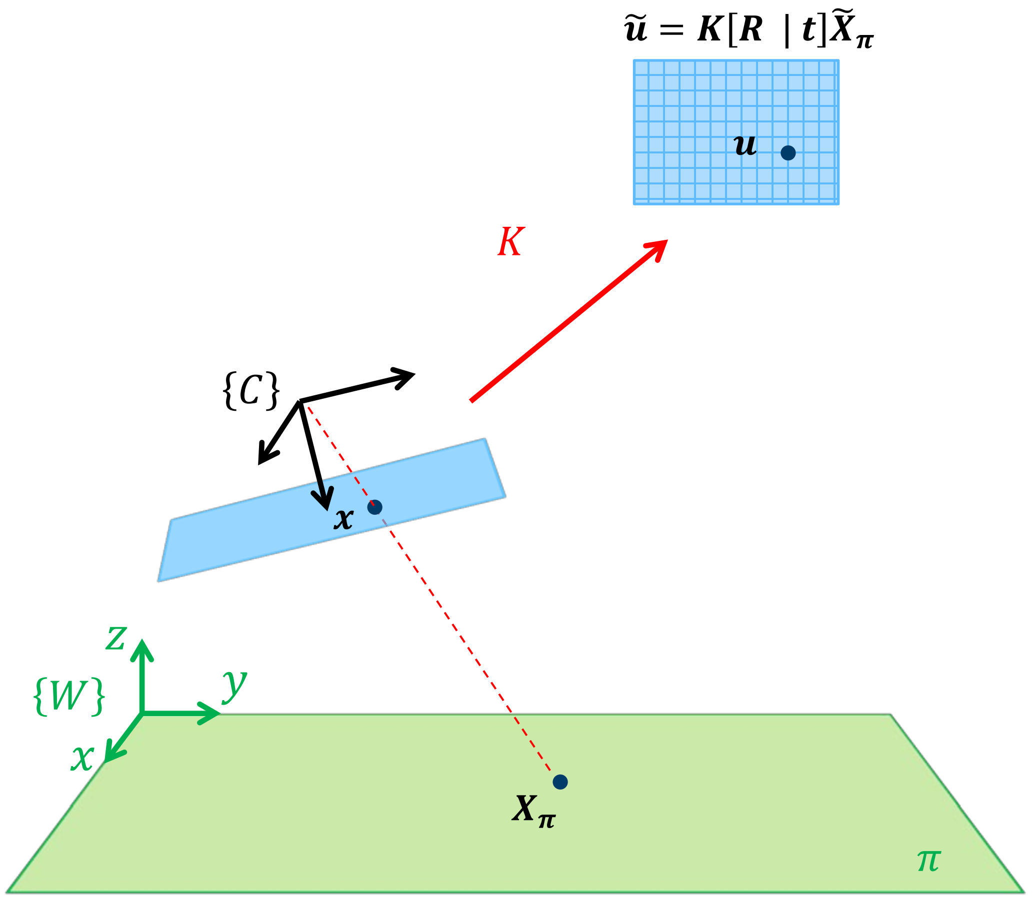
- We can map points on the world plane into image coordinates by using the perspective camera model

$$\tilde{\boldsymbol{u}} = \boldsymbol{K}\left[\boldsymbol{R} \mid \boldsymbol{t}\right]\tilde{\boldsymbol{X}}_{\pi}$$

$$= \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{t}\right]\begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}$$

$$= \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

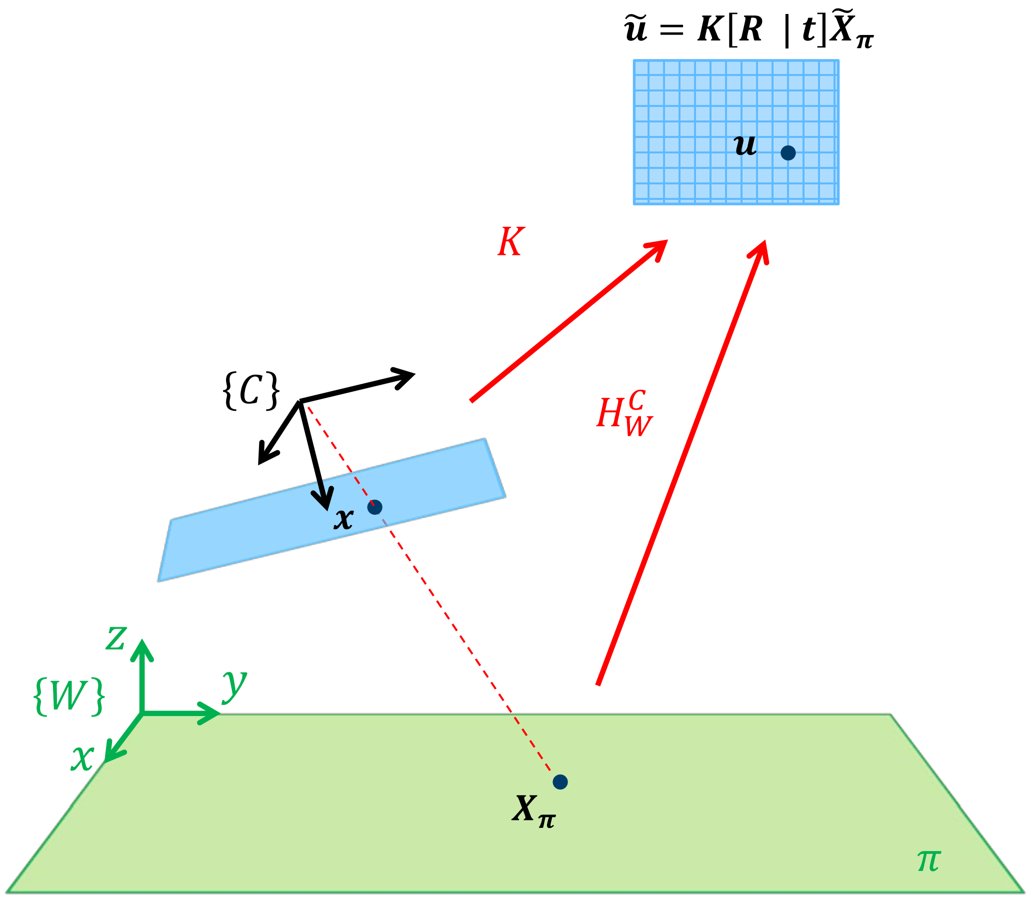


Pose estimation relative to a world plane

- We can map points on the world plane into image coordinates by using the perspective camera model

$$
\begin{aligned}
\tilde{\boldsymbol{u}} &= \boldsymbol{K}\left[\boldsymbol{R} \mid \boldsymbol{t}\right]\tilde{\boldsymbol{X}}_{\pi} \\
&= \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{t}\right]\begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix} \\
&= \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \\
&= \boldsymbol{H}_W^C \tilde{\boldsymbol{x}}_{\pi}
\end{aligned}
$$

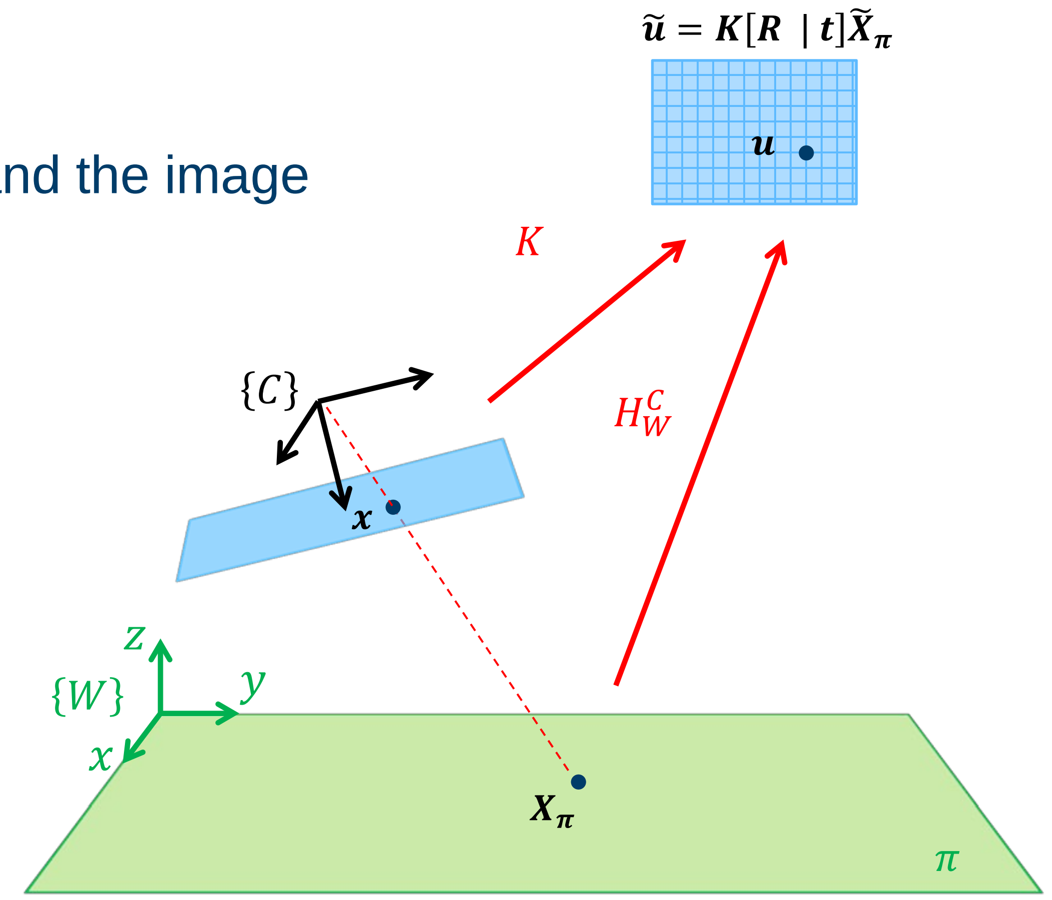
$$
\boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}
$$



Pose estimation relative to a world plane

⇒ For a calibrated camera,
we have a relation between the camera pose
and the homography between the world plane and the image

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] \qquad \boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

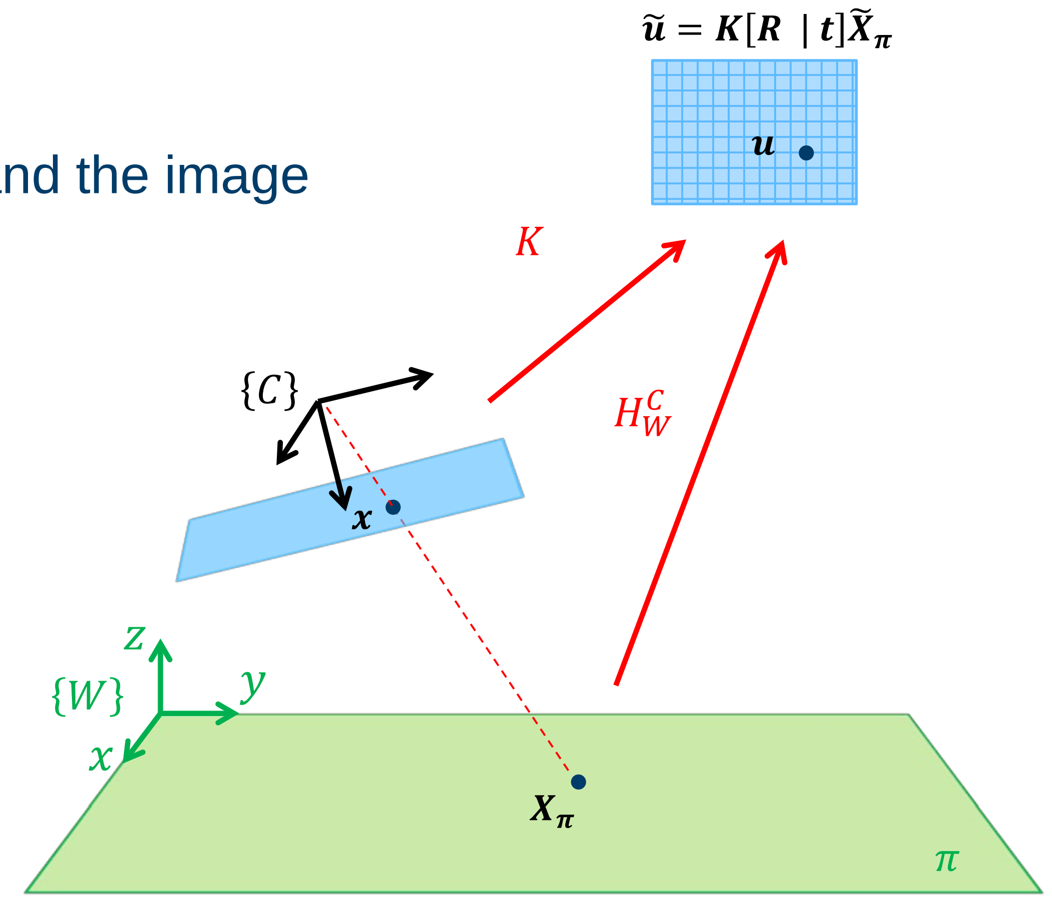


Pose estimation relative to a world plane

⇒ For a calibrated camera,
we have a relation between the camera pose
and the homography between the world plane and the image

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] \qquad \boldsymbol{T}_W^C = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

Can we use this to get camera pose
given a homography?



Pose estimation relative to a world plane

- Assume a perfect, noise-free homography between the world plane and the image:

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]$$

Pose estimation relative to a world plane

- Assume a perfect, noise-free homography between the world plane and the image:

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]$$

- Then, because of scale ambiguity:

$$\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] \sim \boldsymbol{K}^{-1}\boldsymbol{H}_W^C = \boldsymbol{M}$$

Pose estimation relative to a world plane

- Assume a perfect, noise-free homography between the world plane and the image:

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]$$

- Then, because of scale ambiguity:

$$\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] \sim \boldsymbol{K}^{-1}\boldsymbol{H}_W^C = \boldsymbol{M}$$

- Since the columns of rotation matrices have unit norm,
  we can also find a scale factor λ so that the first two columns of $\boldsymbol{M}$ get unit norm.
  We then have the two possible solutions:

$$\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] = \pm\lambda\boldsymbol{M}$$

Pose estimation relative to a world plane

- Assume a perfect, noise-free homography between the world plane and the image:

$$\boldsymbol{H}_W^C = \boldsymbol{K}\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right]$$

- Then, because of scale ambiguity:

$$\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] \sim \boldsymbol{K}^{-1}\boldsymbol{H}_W^C = \boldsymbol{M}$$

- Since the columns of rotation matrices have unit norm,
  we can also find a scale factor λ so that the first two columns of ***M*** get unit norm.
  We then have the two possible solutions:

$$\left[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}\right] = \pm\lambda\boldsymbol{M}$$

- The last column in ***R*** is given by the cross product of the two first columns:

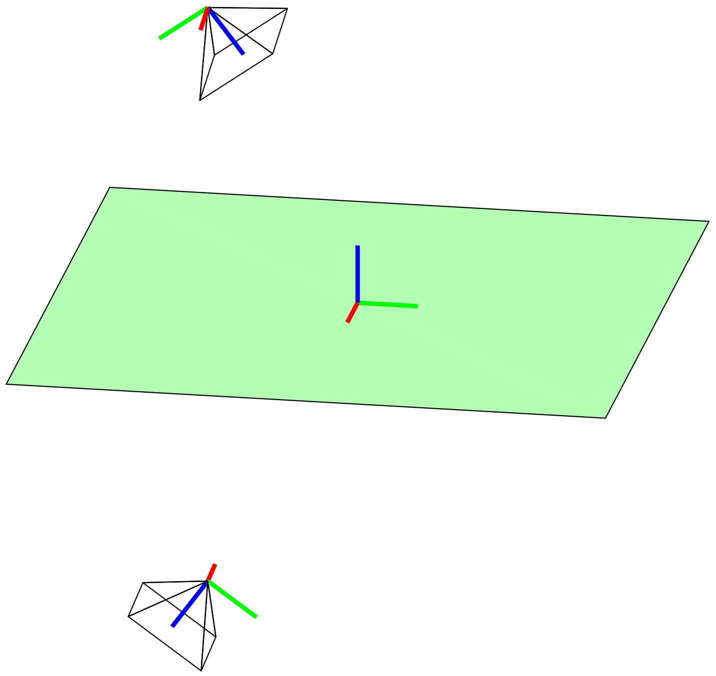
$\boldsymbol{r}_3 = \pm\left(\boldsymbol{r}_1 \times \boldsymbol{r}_2\right)$ where the sign is chosen so that $\det\left(\boldsymbol{R}\right) = 1$

Pose estimation relative to a world plane

- We are now able to reconstruct the camera pose in the world coordinate system for each of the two solutions:

$$\boldsymbol{T}_C^W = \boldsymbol{T}_W^{C^{-1}} = \begin{bmatrix} [\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3] & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}^{-1}$$

- It is in practice simple find the correct solution because only one side of the plane is typically visible



Pose estimation relative to a world plane

- With a homography estimated from point correspondences,
  this approach will typically not give proper rotation matrices because of noise

Pose estimation relative to a world plane

- With a homography estimated from point correspondences,
  this approach will typically not give proper rotation matrices because of noise
  - But it is possible to find the closest rotation matrix in the Frobenius-norm sense with SVD!

Pose estimation relative to a world plane

- With a homography estimated from point correspondences,
  this approach will typically not give proper rotation matrices because of noise
  - But it is possible to find the closest rotation matrix in the Frobenius-norm sense with SVD!

- Let $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{M}}$ be the matrices with the two first columns of $\boldsymbol{R}$ and $\boldsymbol{M}$:

$$\bar{\boldsymbol{R}} = [\boldsymbol{r}_1, \boldsymbol{r}_2] \quad \bar{\boldsymbol{M}} = [\boldsymbol{m}_1, \boldsymbol{m}_2]$$

Pose estimation relative to a world plane

- With a homography estimated from point correspondences, this approach will typically not give proper rotation matrices because of noise
  - But it is possible to find the closest rotation matrix in the Frobenius-norm sense with SVD!

- Let $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{M}}$ be the matrices with the two first columns of $\boldsymbol{R}$ and $\boldsymbol{M}$:

$$\bar{\boldsymbol{R}} = [\boldsymbol{r}_1, \boldsymbol{r}_2] \quad \bar{\boldsymbol{M}} = [\boldsymbol{m}_1, \boldsymbol{m}_2]$$

- With SVD we can get the decomposition $\bar{\boldsymbol{M}} = \boldsymbol{U}_{3\times 2}\boldsymbol{\Sigma}_{2\times 2}\boldsymbol{V}_{2\times 2}^{T}$. The optimal first two columns $\bar{\boldsymbol{R}}$ of a proper $\boldsymbol{R}$, and the corresponding scale λ is then:

$$\bar{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^T \qquad \lambda = \frac{\operatorname{trace}\left(\bar{\boldsymbol{R}}^T \bar{\boldsymbol{M}}\right)}{\operatorname{trace}\left(\bar{\boldsymbol{M}}^T \bar{\boldsymbol{M}}\right)} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{2} R_{ij} M_{ij}}{\sum_{i=1}^{3}\sum_{j=1}^{2} M_{ij}^2}$$

Pose estimation relative to a world plane

- With a homography estimated from point correspondences, this approach will typically not give proper rotation matrices because of noise
  - But it is possible to find the closest rotation matrix in the Frobenius-norm sense with SVD!

- Let $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{M}}$ be the matrices with the two first columns of $\boldsymbol{R}$ and $\boldsymbol{M}$:

$$\bar{\boldsymbol{R}} = \left[\boldsymbol{r}_1, \boldsymbol{r}_2\right] \quad \bar{\boldsymbol{M}} = \left[\boldsymbol{m}_1, \boldsymbol{m}_2\right]$$

- With SVD we can get the decomposition $\bar{\boldsymbol{M}} = \boldsymbol{U}_{3\times 2}\boldsymbol{\Sigma}_{2\times 2}\boldsymbol{V}_{2\times 2}^{T}$.
The optimal first two columns $\bar{\boldsymbol{R}}$ of a proper $\boldsymbol{R}$, and the corresponding scale λ is then:

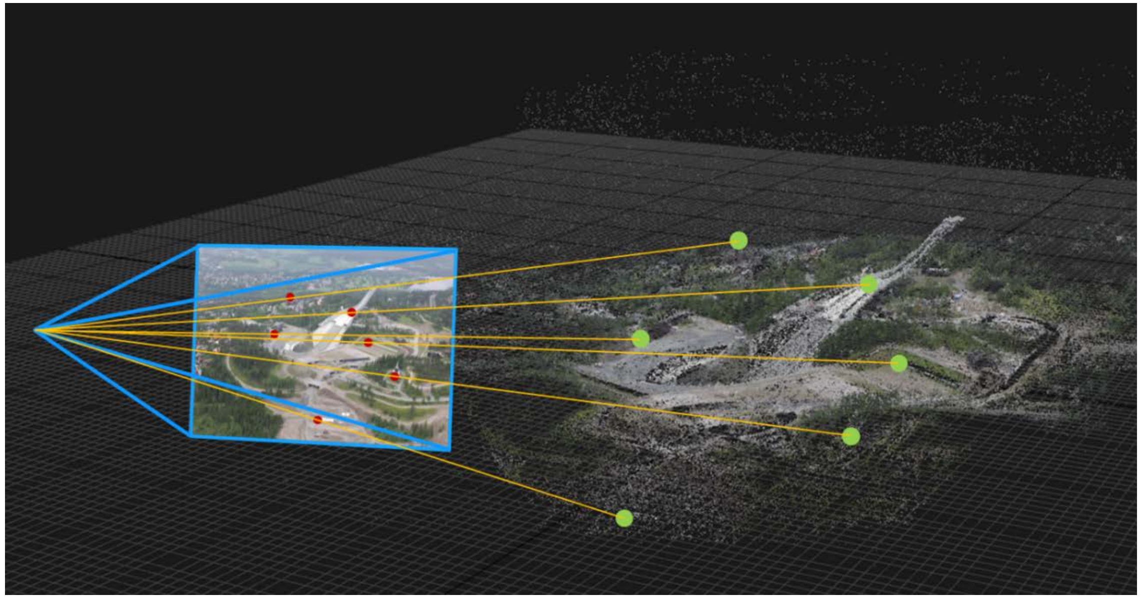
$$\bar{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^T \qquad \lambda = \frac{\operatorname{trace}\left(\bar{\boldsymbol{R}}^T \bar{\boldsymbol{M}}\right)}{\operatorname{trace}\left(\bar{\boldsymbol{M}}^T \bar{\boldsymbol{M}}\right)} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{2} R_{ij} M_{ij}}{\sum_{i=1}^{3}\sum_{j=1}^{2} M_{ij}^2}$$

- The corresponding pose with ambiguity can then be found as before

Pose estimation relative to known 3D points

- Iterative nonlinear estimation
  - Measurement model

$$\boldsymbol{u}_l = h(\boldsymbol{T}; l, \boldsymbol{K}) + \boldsymbol{\eta} \qquad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

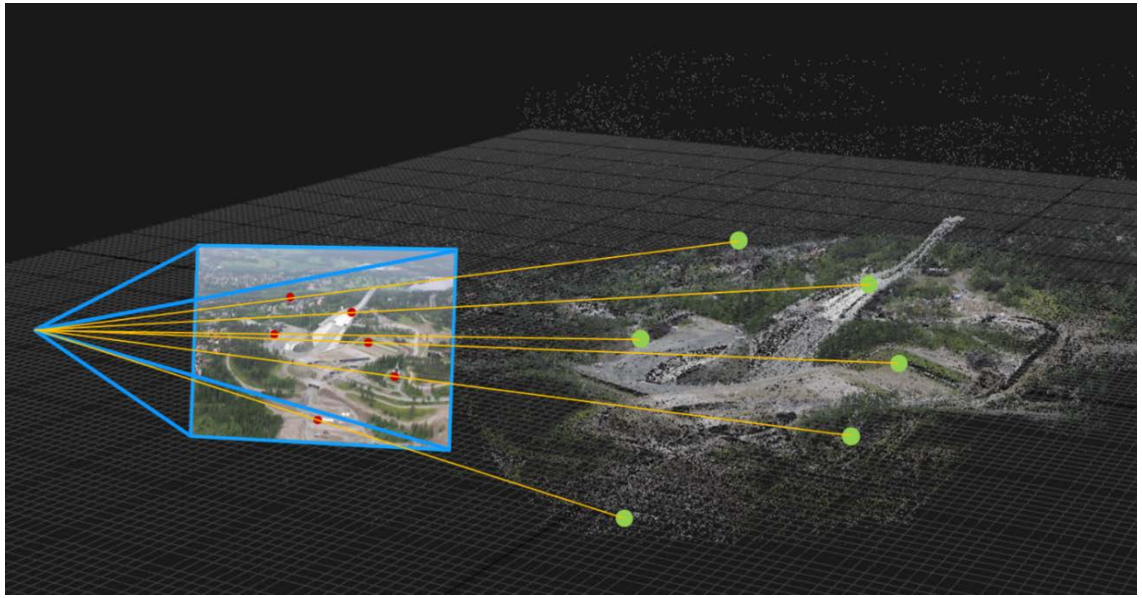


Pose estimation relative to known 3D points

- Iterative nonlinear estimation
  - Measurement model

$$\boldsymbol{u}_l = h(\boldsymbol{T}; l, \boldsymbol{K}) + \boldsymbol{\eta} \qquad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

$$= \boldsymbol{K}\left[\boldsymbol{R} \mid \boldsymbol{t}\right]\boldsymbol{X}_l + \boldsymbol{\eta}$$



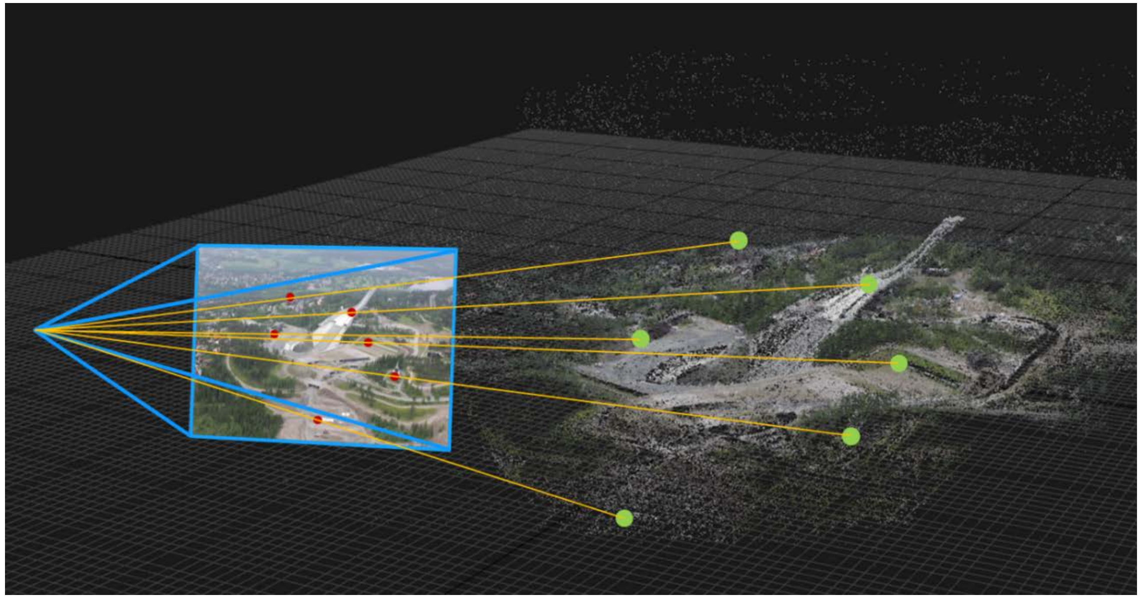
Pose estimation relative to known 3D points

- Iterative nonlinear estimation
  - Measurement model

$$\begin{aligned} \boldsymbol{u}_l &= h(\boldsymbol{T}; l, \boldsymbol{K}) + \boldsymbol{\eta} \qquad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \\ &= \boldsymbol{K}\left[\boldsymbol{R} \,|\, \boldsymbol{t}\right]\boldsymbol{X}_l + \boldsymbol{\eta} \end{aligned}$$

  - Minimize error over all 3D-2D landmark correspondences

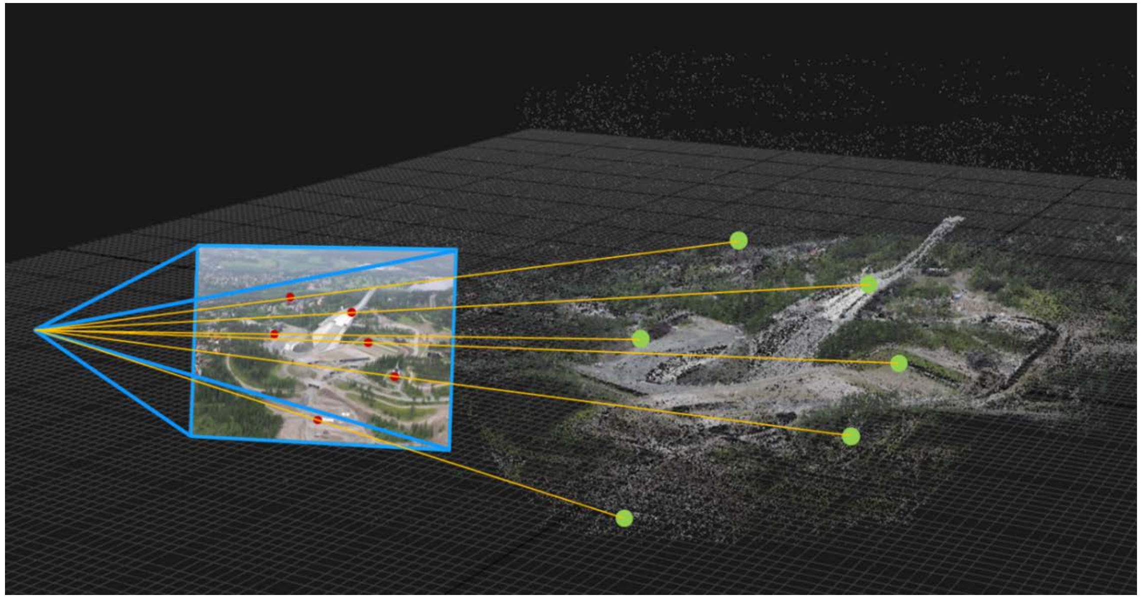
$$\boldsymbol{T}^{MAP} = \arg\min_{\boldsymbol{T}} \sum_i \left\| h_i(\boldsymbol{T}) - \boldsymbol{u}_i \right\|_{\boldsymbol{\Sigma}_i}^2$$



Pose estimation relative to known 3D points

- *n*-Point Pose Problem (P*n*P)
  - Typically fast non-iterative methods
  - Minimal in number of points
  - Accuracy comparable to iterative methods

- Examples
  - P3P, EPnP
  - P4Pf
    - Estimate pose and focal length
  - P6P
    - Estimates *P* with DLT
  - R6P
    - Estimate pose with rolling shutter



Summary

- Pose estimation relative to a world plane
  - Pose from homography

- Pose estimation relative to known 3D points
  - Iterative methods
  - P*n*P

- Further reading:
  - Torstein Sattler,
    CVPR 2015 Tutorial on Large-Scale Visual Place Recognition and Image-Based Localization