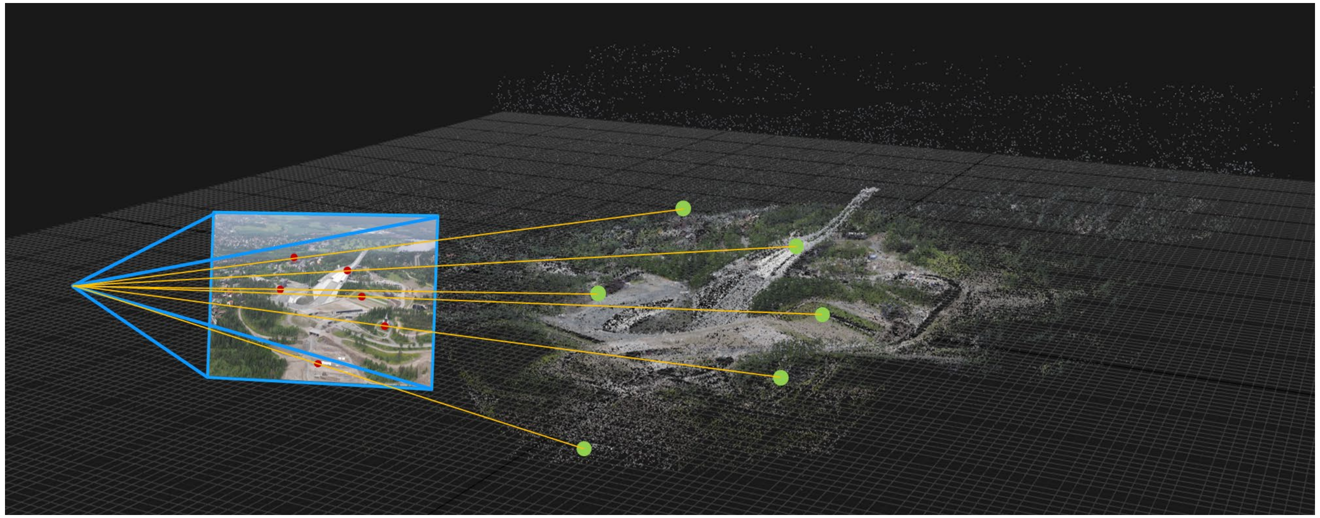
UiO : Department of Technology Systems
University of Oslo

Estimating camera pose from a single image and a known map

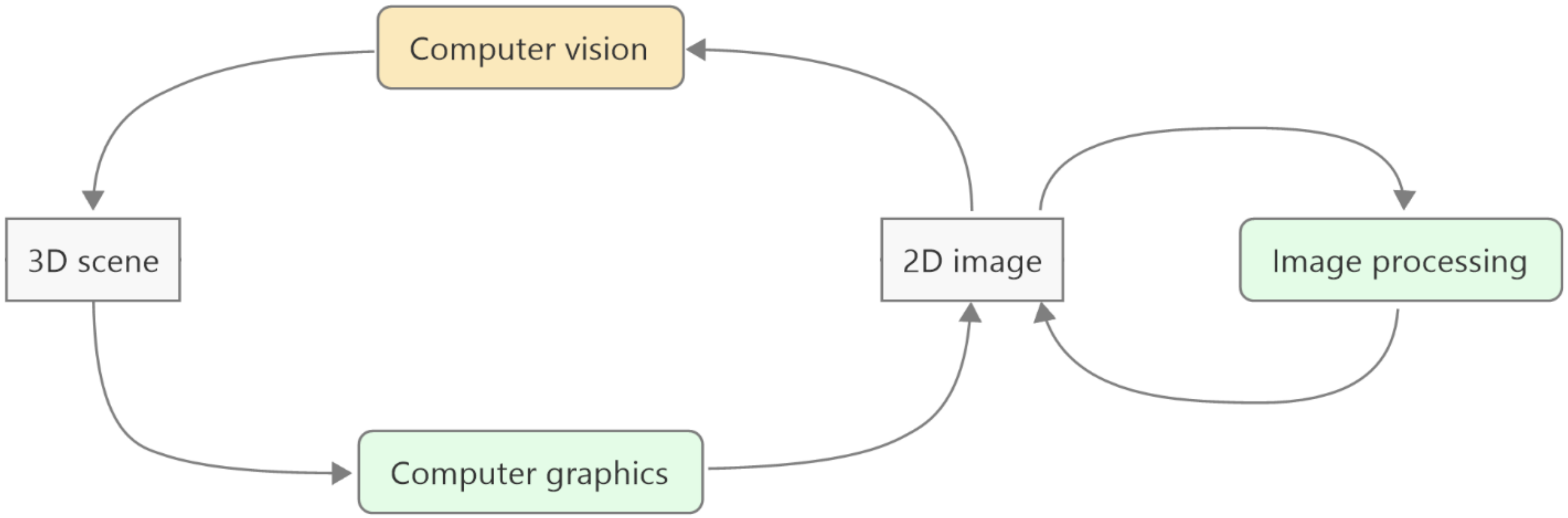
Trym Vegard Haavardsholm

2023

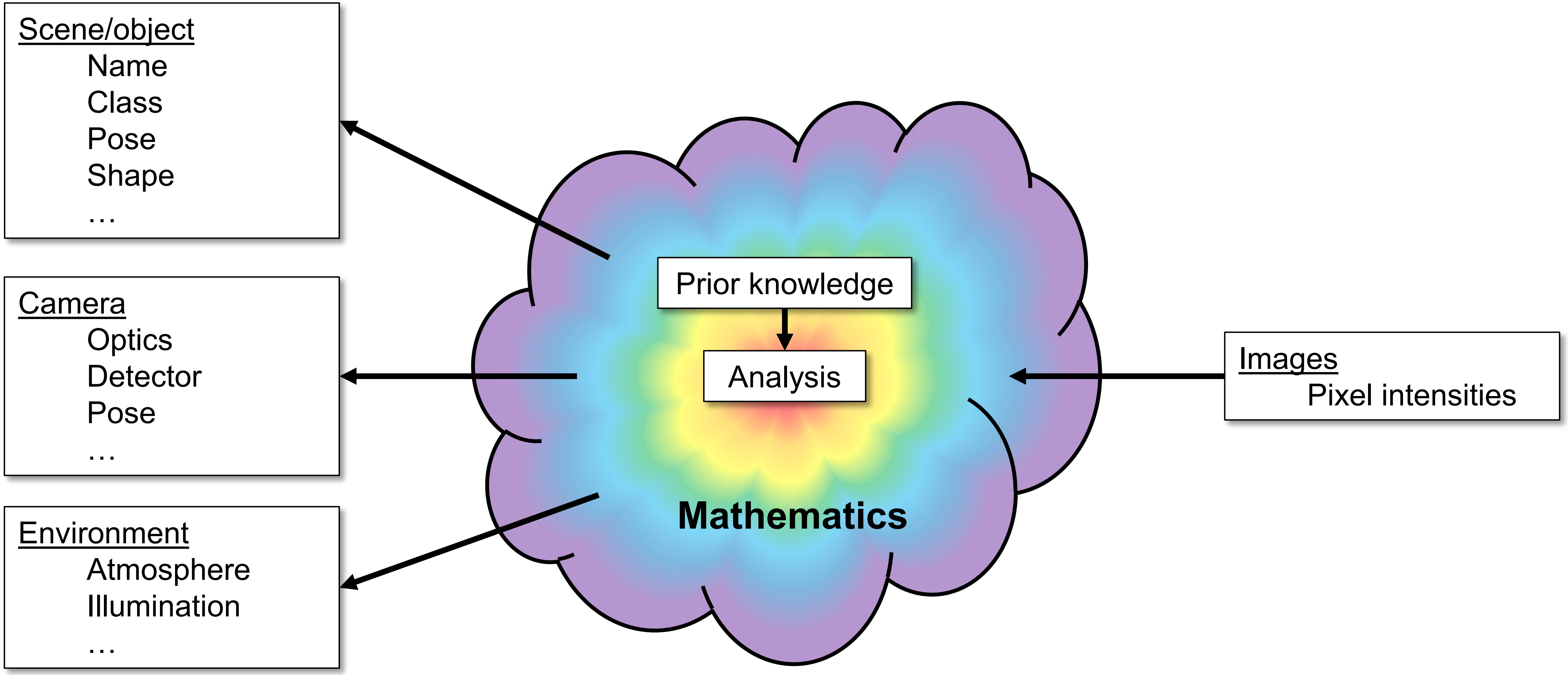


TEK5030

Computer vision is an *inverse problem*!



The *inverse* analysis process

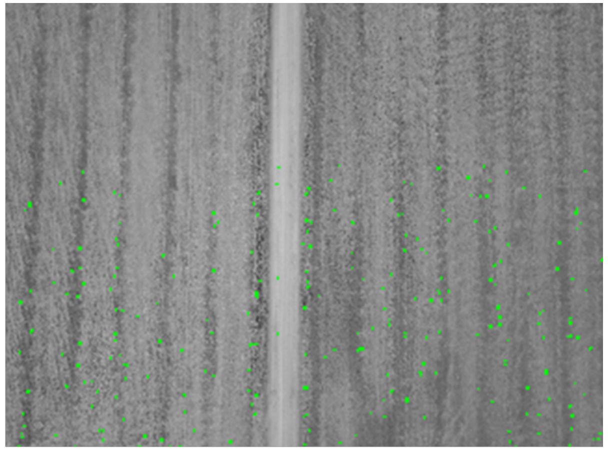


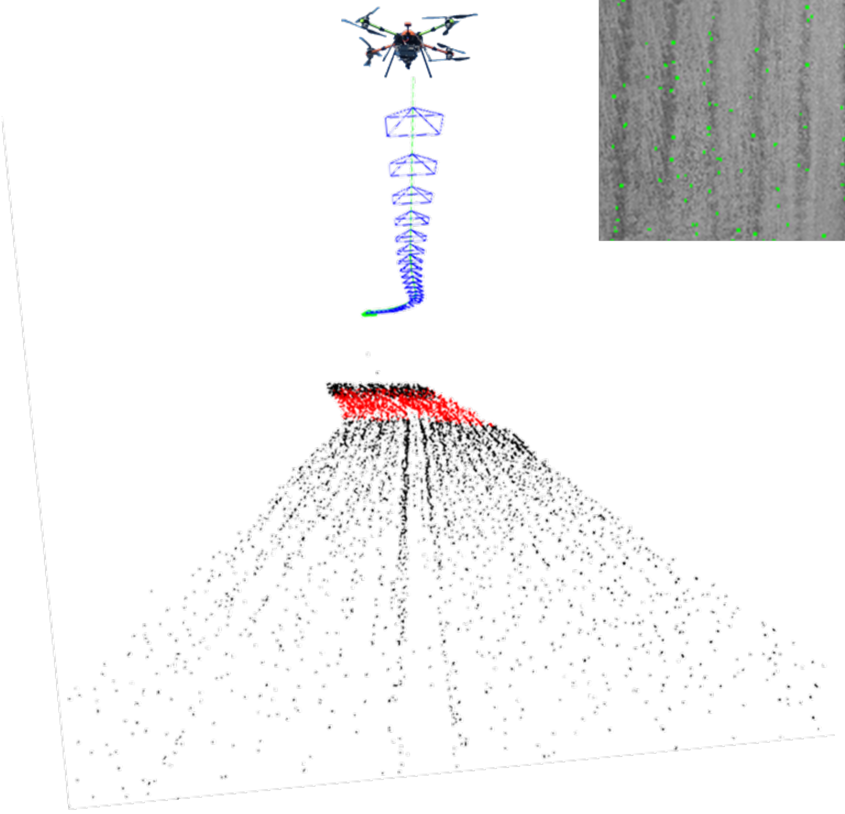
Localisation

Pose estimation based on correspondences with a known map is called **localisation**

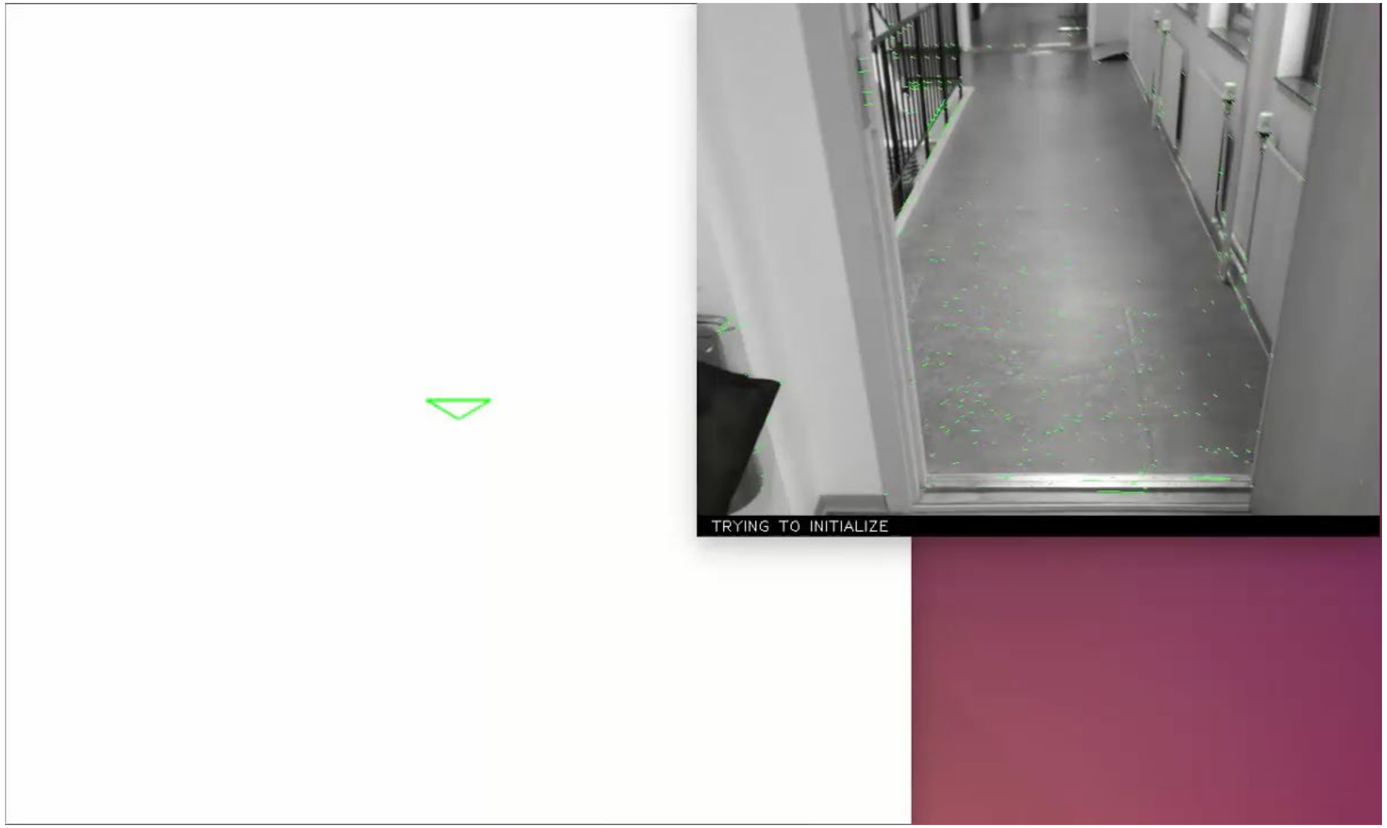
In **visual localisation**,
this is also sometimes called **tracking**

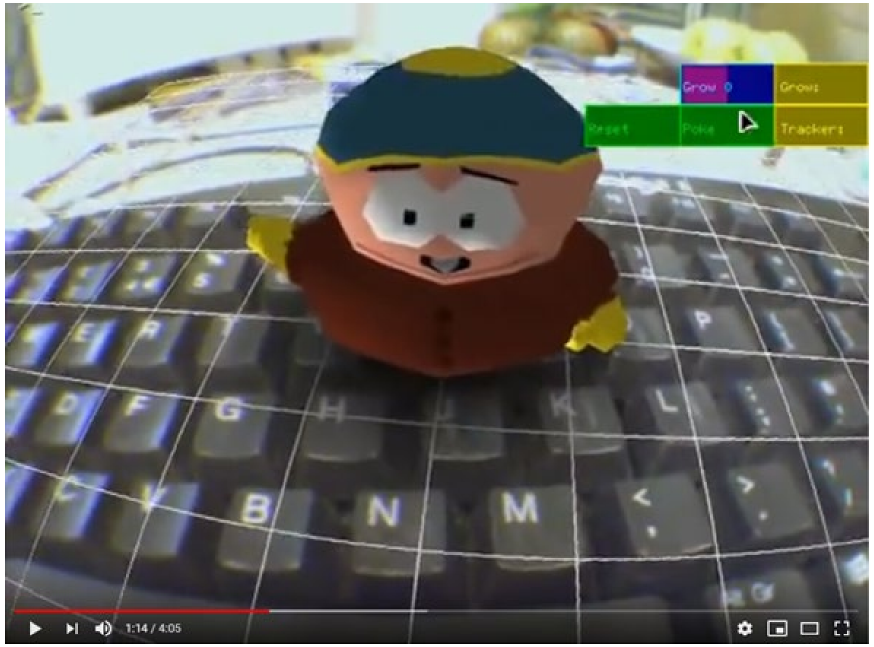
- Tracking the map in the image frames





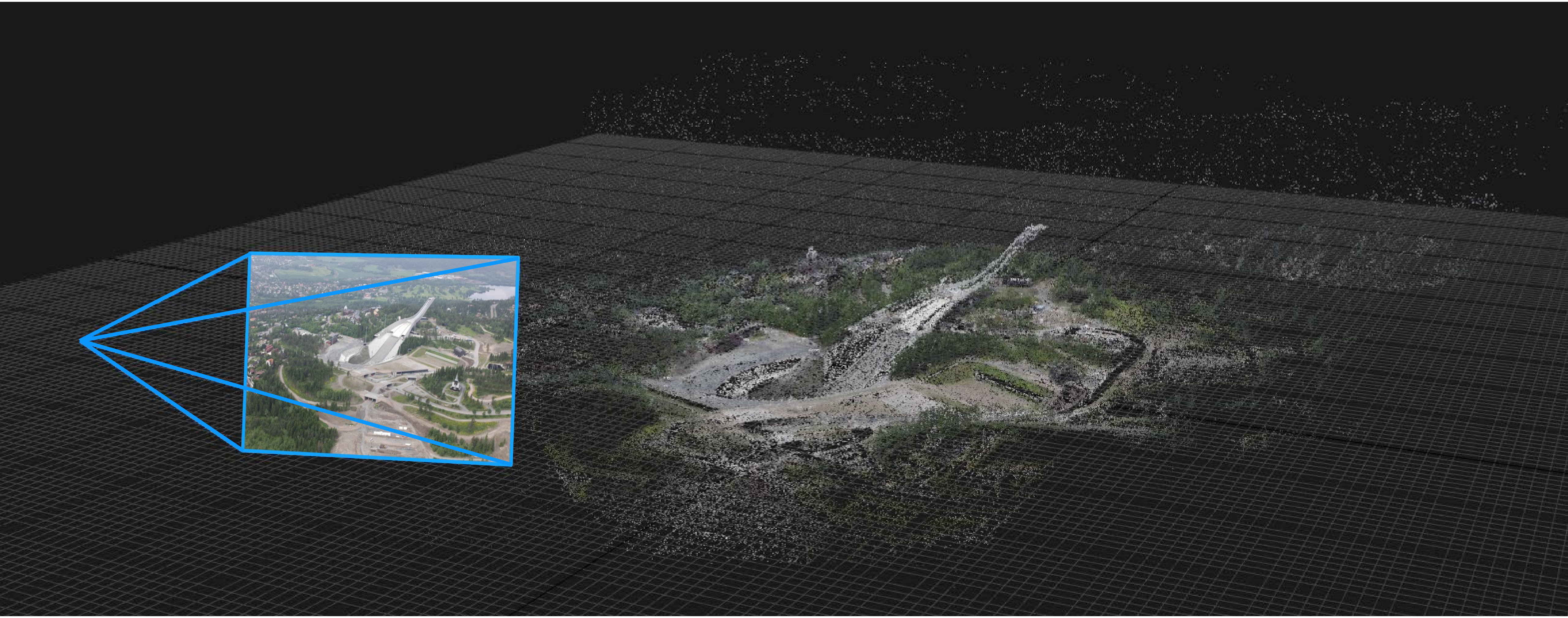
Why learn about localisation?



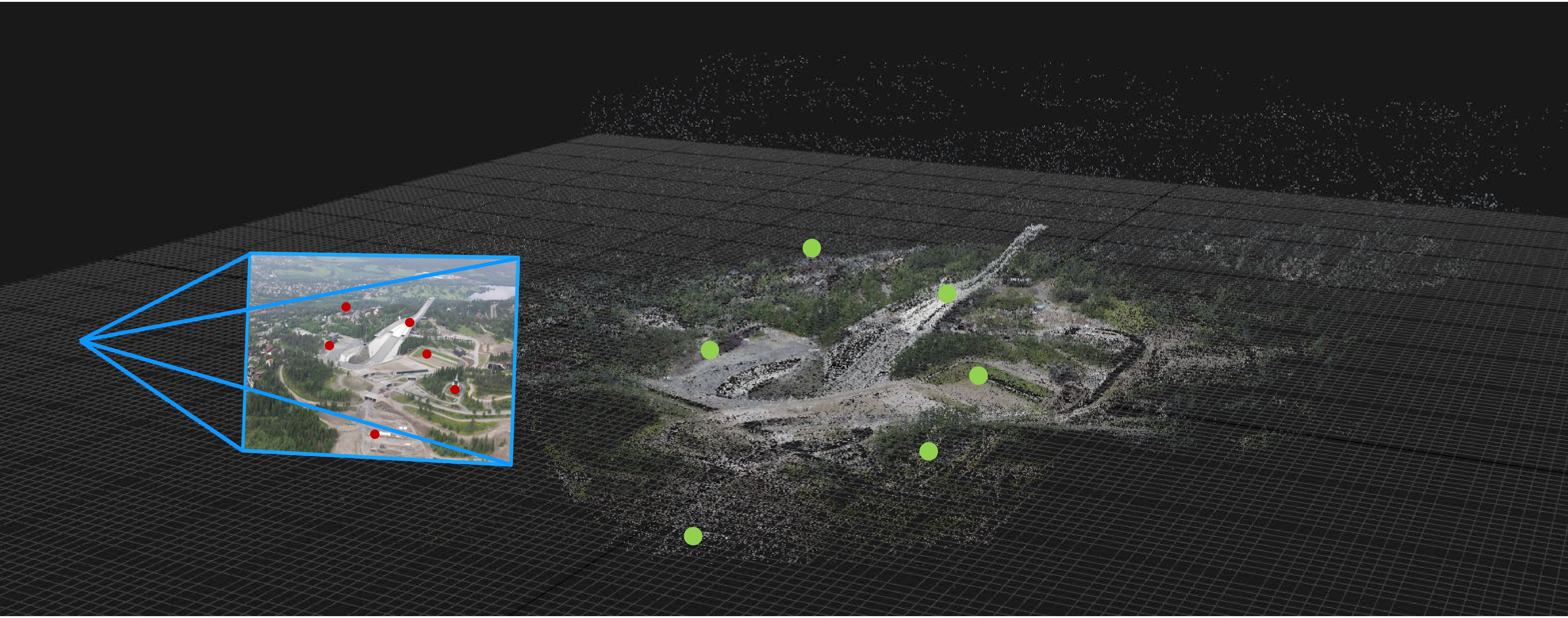


From PTAM by Georg Klein and David Murray (2007)
https://www.youtube.com/watch?v=F3s3M0mokNc

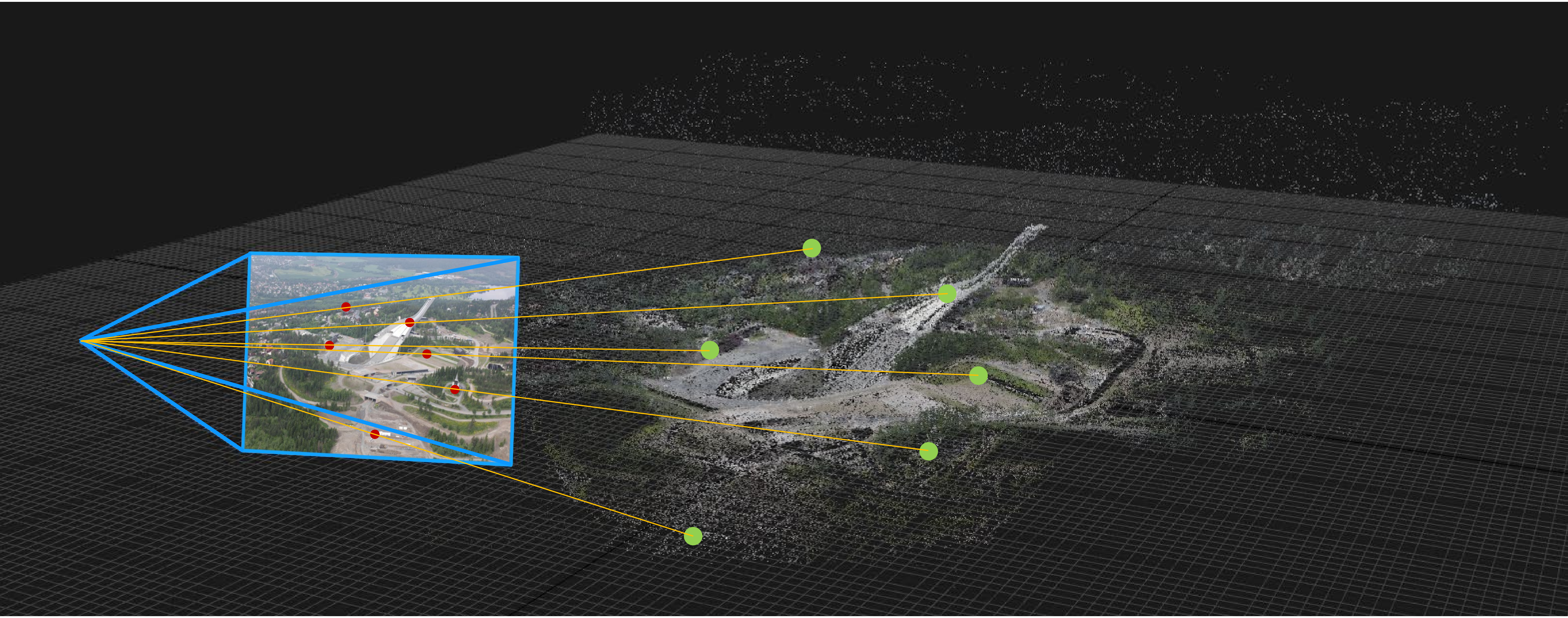
How can we track a map with a camera?



Pose from 2D correspondences with known 3D points



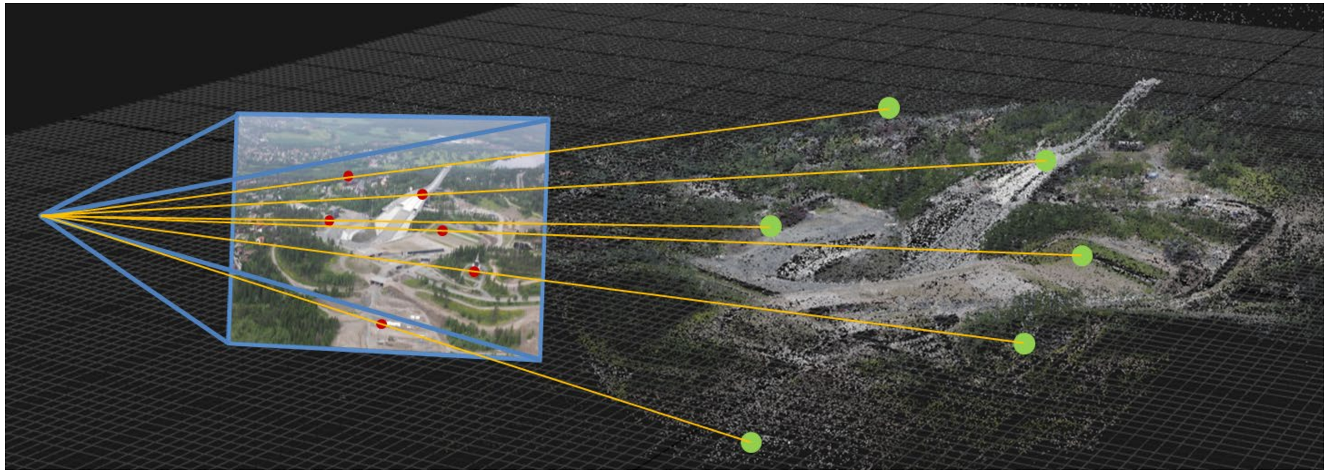
Pose from 2D correspondences with known 3D points



Pose from 2D correspondences with known 3D points

Minimise **geometric error**

$$\mathbf{T}_{wc}^{*} = \underset{\mathbf{T}_{wc}}{\operatorname{argmin}} \sum_{i} \left\| \pi(\mathbf{T}_{wc}^{-1} \cdot \mathbf{x}_{i}^{w}) - \mathbf{u}_{i} \right\|^{2}$$

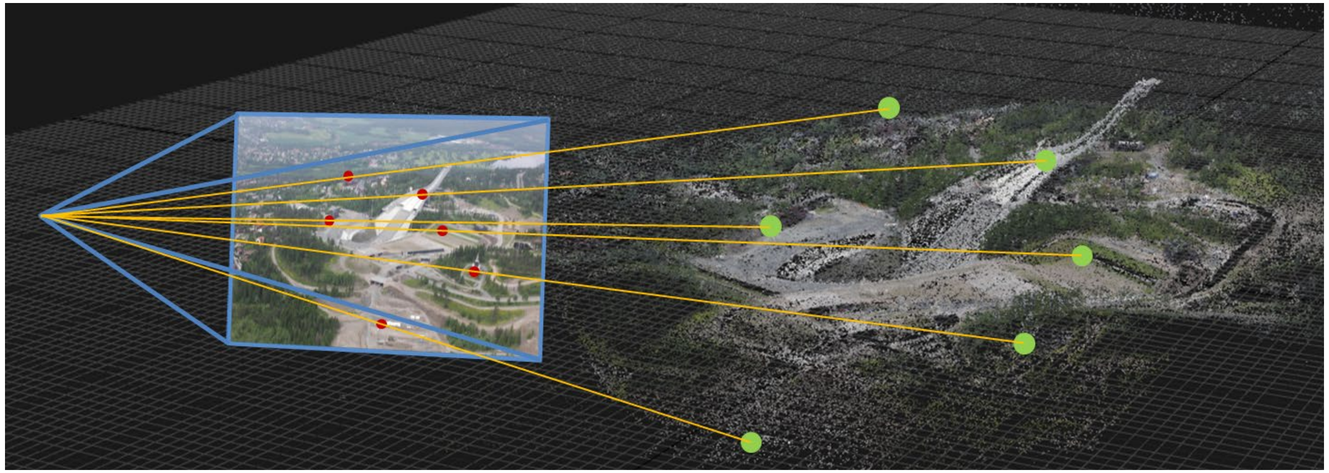


Pose from 2D correspondences with known 3D points

Minimise **geometric error**

$$\mathbf{T}_{wc}^{*} = \underset{\mathbf{T}_{wc}}{\operatorname{argmin}} \sum_{i} \left\| \pi(\mathbf{T}_{wc}^{-1} \cdot \mathbf{x}_i^{w}) - \mathbf{u}_i \right\|^2$$

also called **reprojection error**



Pose estimation

We will solve the indirect tracking problem

$$\mathbf{T}_{wc}^{*} = \underset{\mathbf{T}_{wc}}{\operatorname{argmin}} \sum_{i} \left\| \pi(\mathbf{T}_{wc}^{-1} \cdot \mathbf{x}_{i}^{w}) - \mathbf{u}_{i} \right\|^{2}$$

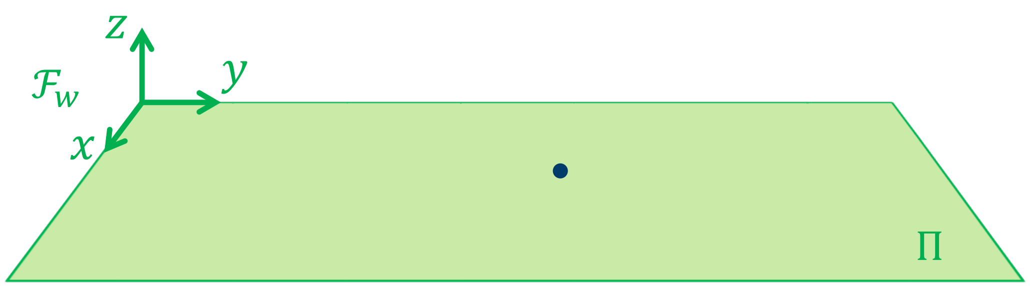
in the next few videos.

But lets first solve a simpler problem,
when we can assume that the world is planar!

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

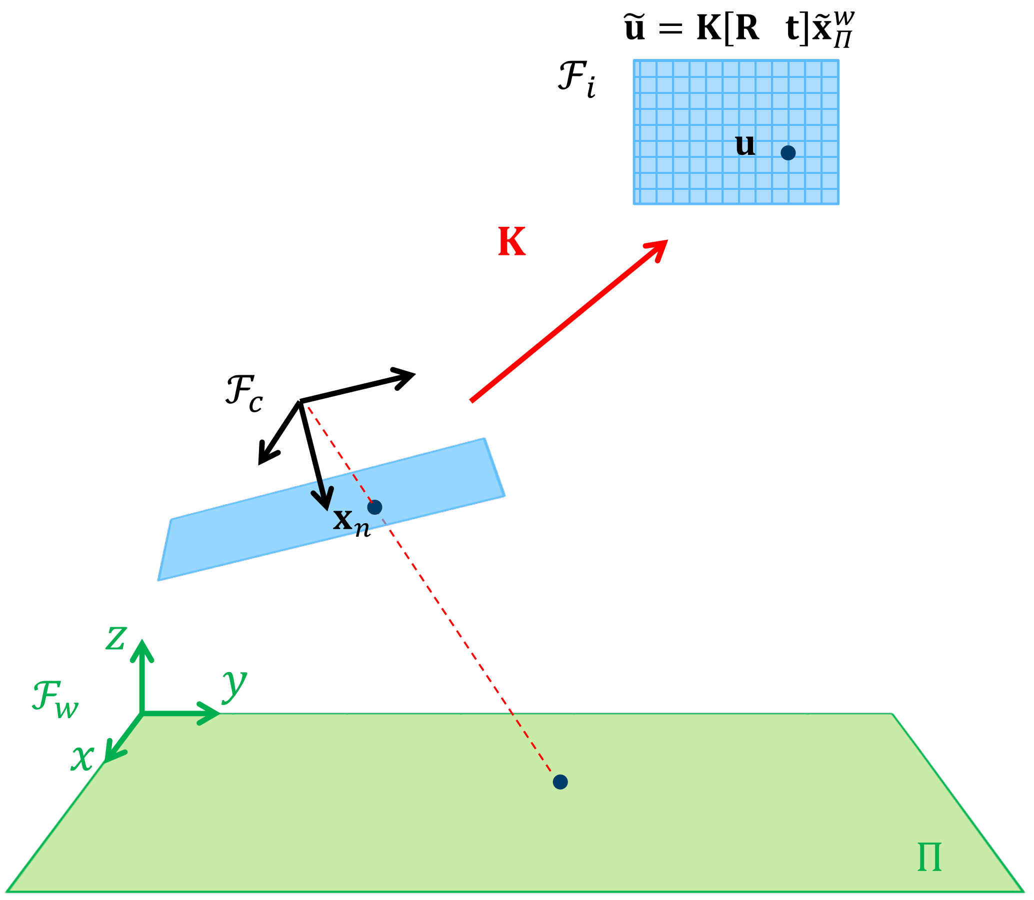
$$\mathbf{x}_{\Pi}^{w} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \qquad \mathbf{x}^{\Pi} = \begin{bmatrix} x \\ y \end{bmatrix}$$



Pose estimation relative to a world plane

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

$$\tilde{\mathbf{u}} = \mathbf{K}\left[\mathbf{R}\ \mathbf{t}\right]\tilde{\mathbf{x}}_{\Pi}^{w} \qquad \mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$



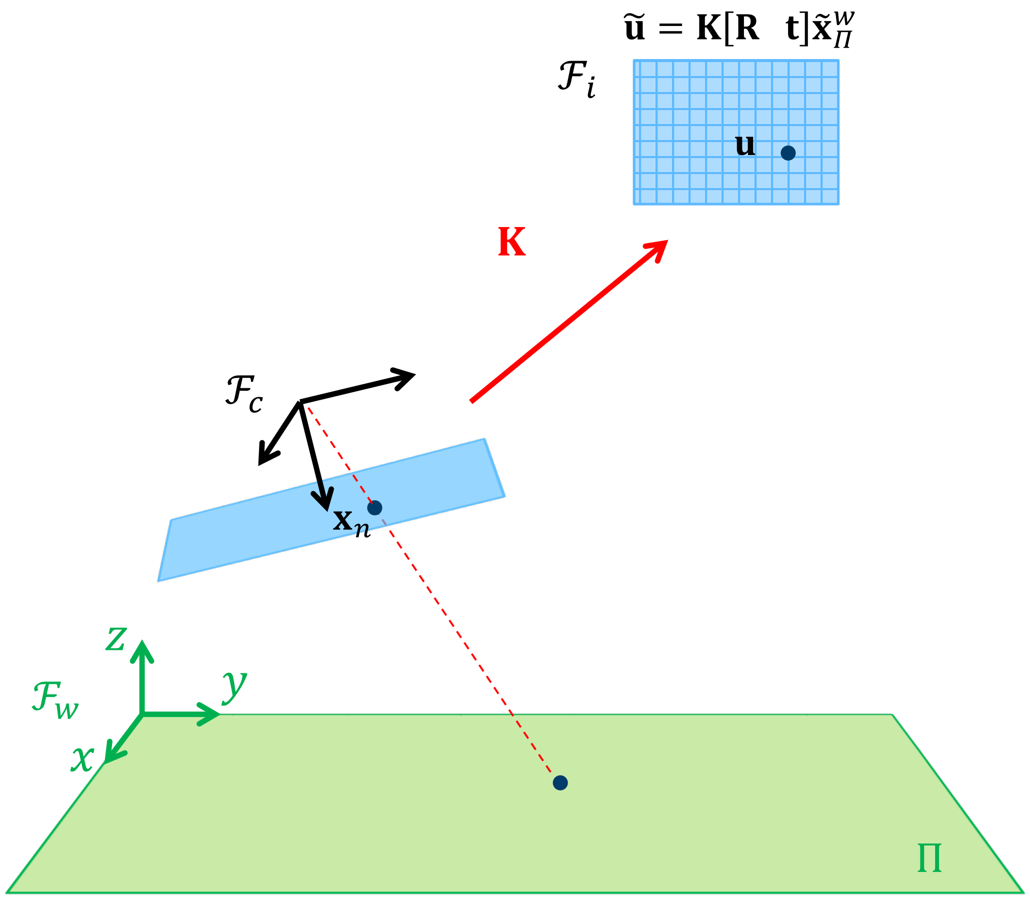
Pose estimation relative to a world plane

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

$$\tilde{\mathbf{u}} = \mathbf{K}\left[\mathbf{R}\ \mathbf{t}\right]\tilde{\mathbf{x}}_{\Pi}^{w}$$

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

$$\mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$



Pose estimation relative to a world plane

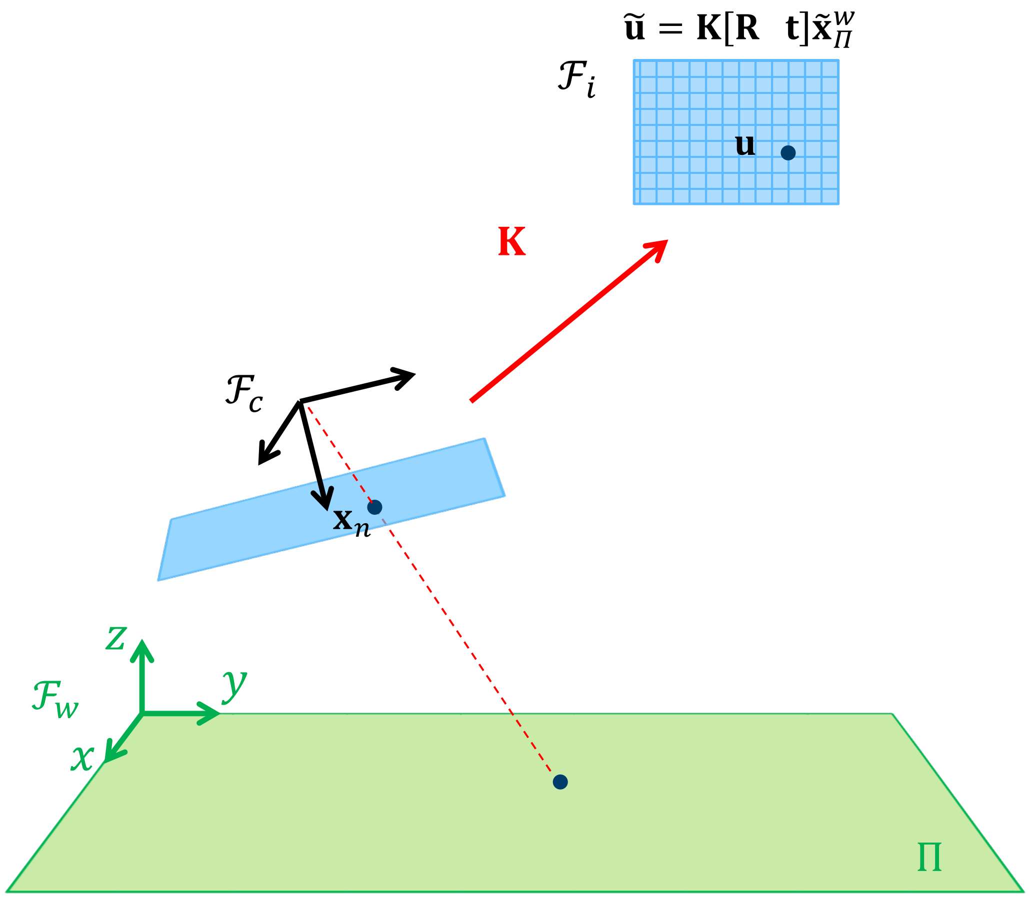
We can map points on the world plane into image coordinates by using the perspective camera model

$$\tilde{\mathbf{u}} = \mathbf{K}\left[\mathbf{R}\ \mathbf{t}\right]\tilde{\mathbf{x}}_{\Pi}^{w}$$

$$\mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$

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

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



Pose estimation relative to a world plane

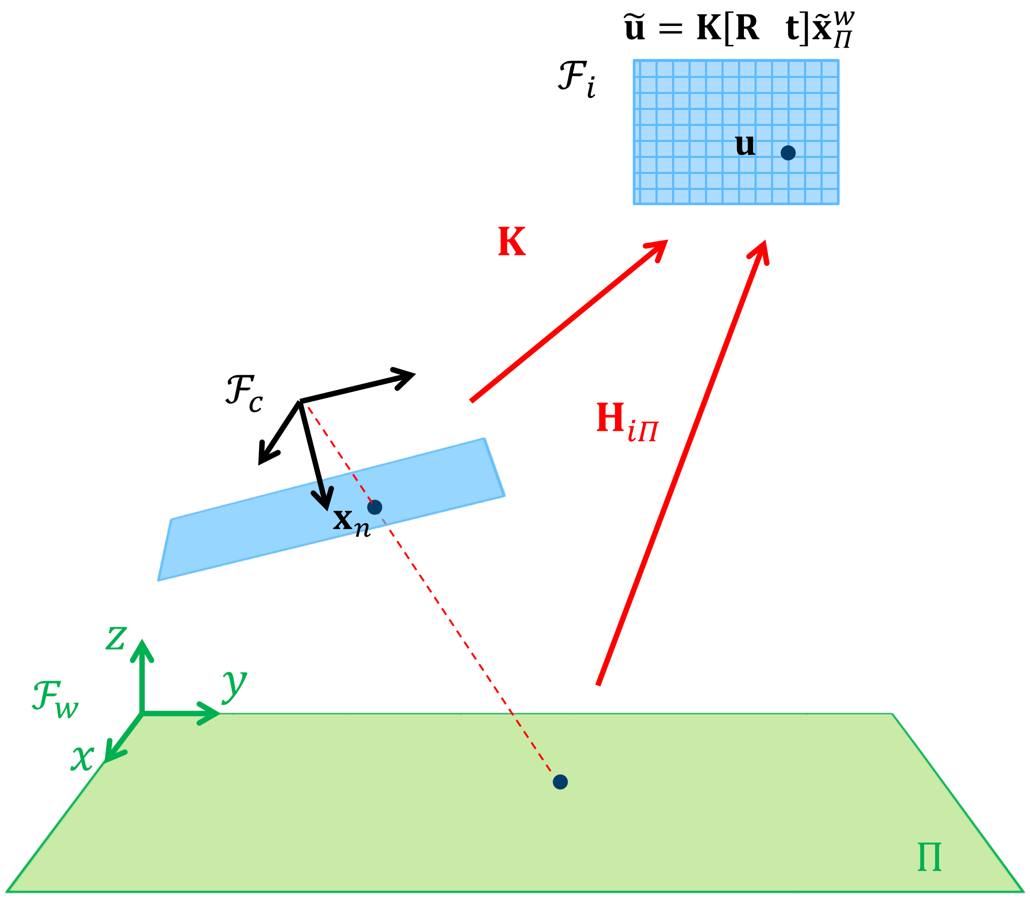
We can map points on the world plane into image coordinates by using the perspective camera model

$$\tilde{\mathbf{u}} = \mathbf{K}\left[\mathbf{R}\ \mathbf{t}\right]\tilde{\mathbf{x}}_{\Pi}^{w} \qquad \mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$

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

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

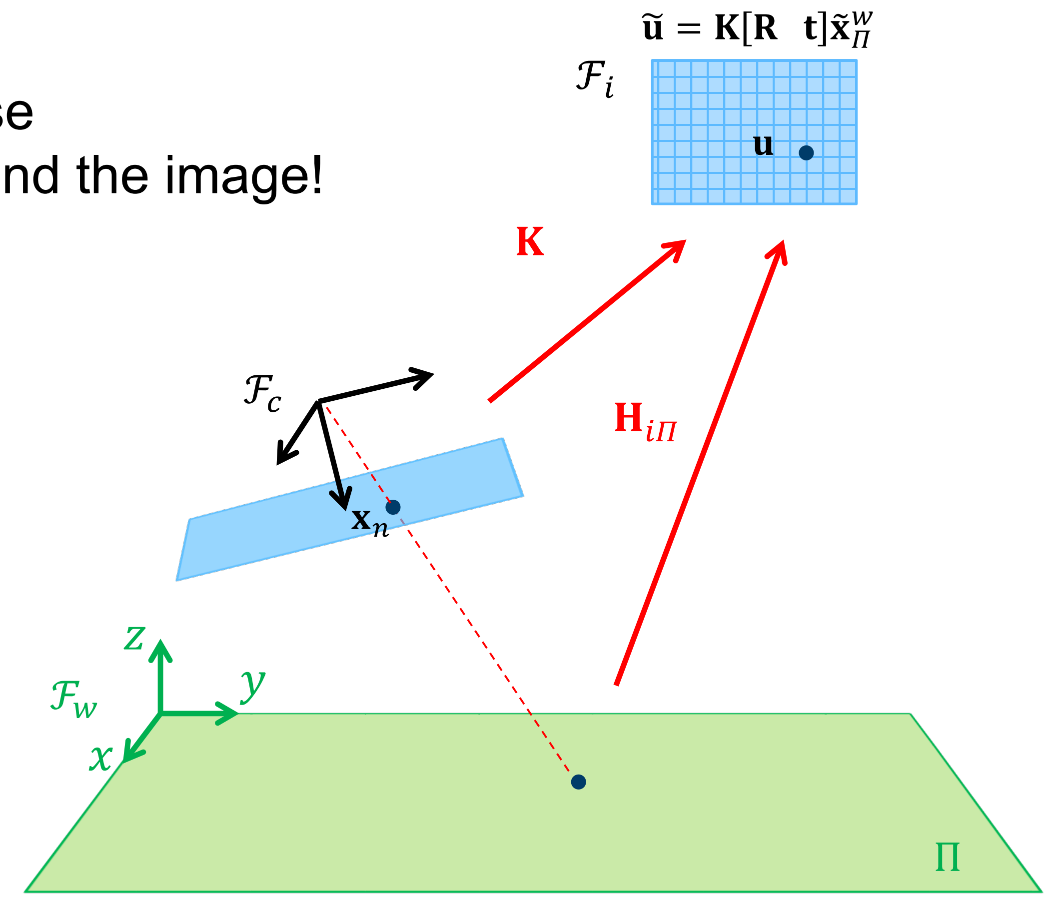
$$= \mathbf{H}_{i\Pi}\tilde{\mathbf{x}}^{\Pi}$$



Pose estimation relative to a world plane

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

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right] \qquad \mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$

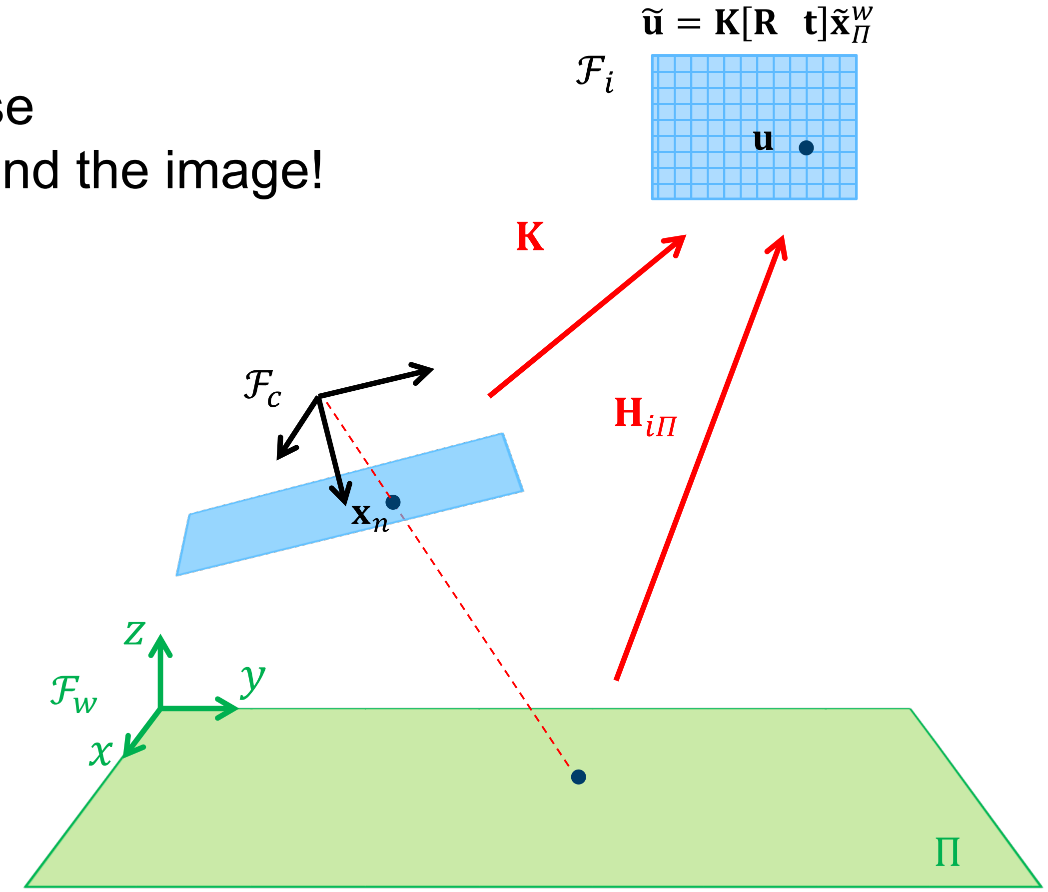


Pose estimation relative to a world plane

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

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right] \qquad \mathbf{T}_{cw} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$$

How can we use this to estimate 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:

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right]$$

Pose estimation relative to a world plane

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

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right]$$

Then, because of scale ambiguity:

$$\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right] \sim \mathbf{K}^{-1}\mathbf{H}_{i\Pi} = \mathbf{M}$$

Pose estimation relative to a world plane

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

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right]$$

Then, because of scale ambiguity:

$$\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right] \sim \mathbf{K}^{-1}\mathbf{H}_{i\Pi} = \mathbf{M}$$

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

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

Pose estimation relative to a world plane

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

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right]$$

Then, because of scale ambiguity:

$$\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right] \sim \mathbf{K}^{-1}\mathbf{H}_{i\Pi} = \mathbf{M}$$

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

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

The last column in $\hat{\mathbf{R}}$ is given by the cross product of the two first columns:

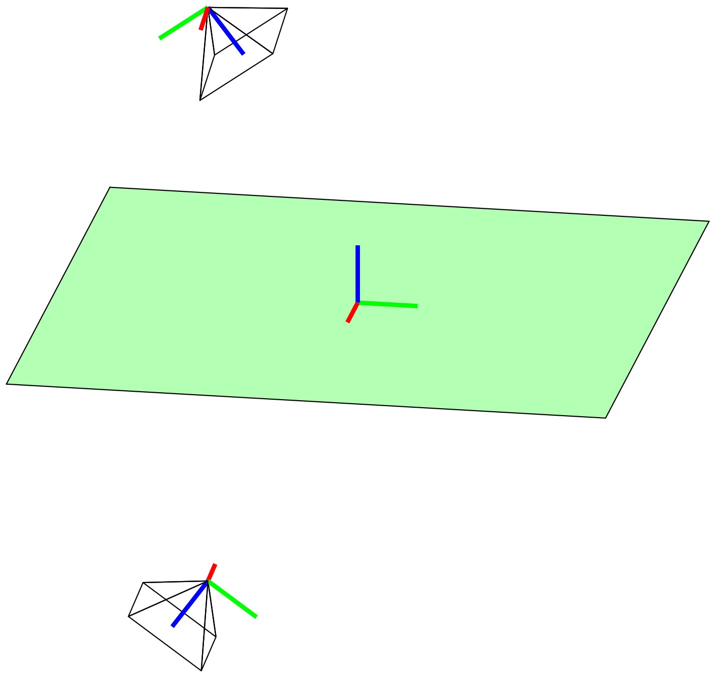
$\hat{\mathbf{r}}_3 = \pm\left(\hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2\right)$, where the sign is chosen so that $\det\left(\hat{\mathbf{R}}\right) = 1$

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:

$$\hat{\mathbf{T}}_{wc} = \hat{\mathbf{T}}_{cw}^{-1} = \begin{bmatrix} \hat{\mathbf{R}}^T & -\hat{\mathbf{R}}^T\hat{\mathbf{t}} \\ \mathbf{0} & 1 \end{bmatrix} \qquad \text{where} \qquad \hat{\mathbf{R}} = \left[\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \hat{\mathbf{r}}_3\right]$$

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



Pose estimation with planar correspondences

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

$$\hat{\mathbf{R}} \notin SO(3)$$

Pose estimation with planar correspondences

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

$$\hat{\mathbf{R}} \notin SO(3)$$

But it is possible to find the closest rotation matrix with SVD!

$$\hat{\mathbf{R}} \rightarrow \hat{\mathbf{R}}^* \in SO(3)$$

Pose estimation with planar correspondences

Let $\bar{\mathbf{M}}$ be the matrix with the two first columns of $\mathbf{M}$:

$$\bar{\mathbf{M}} = [\mathbf{m}_1, \mathbf{m}_2]$$

Pose estimation with planar correspondences

Let $\bar{\mathbf{M}}$ be the matrix with the two first columns of $\mathbf{M}$:

$$\bar{\mathbf{M}} = [\mathbf{m}_1, \mathbf{m}_2]$$

With SVD we can get the decomposition $\bar{\mathbf{M}} = \mathbf{U}_{3\times2}\mathbf{\Sigma}_{2\times2}\mathbf{V}_{2\times2}^T$.

Pose estimation with planar correspondences

Let $\bar{\mathbf{M}}$ be the matrix with the two first columns of $\mathbf{M}$:

$$\bar{\mathbf{M}} = [\mathbf{m}_1, \mathbf{m}_2]$$

With SVD we can get the decomposition $\bar{\mathbf{M}} = \mathbf{U}_{3\times2}\mathbf{\Sigma}_{2\times2}\mathbf{V}_{2\times2}^T$.
The first two columns $\bar{\mathbf{R}}^*$ of a proper $\hat{\mathbf{R}}^*$ is then

$$\bar{\mathbf{R}}^* = \mathbf{U}\mathbf{V}^T$$

Pose estimation with planar correspondences

Let $\bar{\mathbf{M}}$ be the matrix with the two first columns of $\mathbf{M}$:

$$\bar{\mathbf{M}} = [\mathbf{m}_1, \mathbf{m}_2]$$

With SVD we can get the decomposition $\bar{\mathbf{M}} = \mathbf{U}_{3\times 2}\mathbf{\Sigma}_{2\times 2}\mathbf{V}_{2\times 2}^T$.
The first two columns $\bar{\mathbf{R}}^*$ of a proper $\hat{\mathbf{R}}^*$ is then

$$\bar{\mathbf{R}}^* = \mathbf{U}\mathbf{V}^T$$

The corresponding scale λ can be computed as:

$$\lambda = \frac{\operatorname{trace}\left(\bar{\mathbf{R}}^{*T}\bar{\mathbf{M}}\right)}{\operatorname{trace}\left(\bar{\mathbf{M}}^T\bar{\mathbf{M}}\right)} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{2} r_{ij}^* m_{ij}}{\sum_{i=1}^{3}\sum_{j=1}^{2} m_{ij}^2}$$

Pose estimation with planar correspondences

Let $\bar{\mathbf{M}}$ be the matrix with the two first columns of $\mathbf{M}$:

$$\bar{\mathbf{M}} = [\mathbf{m}_1, \mathbf{m}_2]$$

With SVD we can get the decomposition $\bar{\mathbf{M}} = \mathbf{U}_{3\times2}\mathbf{\Sigma}_{2\times2}\mathbf{V}_{2\times2}^T$.
The first two columns $\bar{\mathbf{R}}^*$ of a proper $\hat{\mathbf{R}}^*$ is then

$$\bar{\mathbf{R}}^* = \mathbf{U}\mathbf{V}^T$$

The corresponding scale λ can be computed as:

$$\lambda = \frac{\operatorname{trace}\left(\bar{\mathbf{R}}^{*T}\bar{\mathbf{M}}\right)}{\operatorname{trace}\left(\bar{\mathbf{M}}^T\bar{\mathbf{M}}\right)} = \frac{\sum_{i=1}^{3}\sum_{j=1}^{2} r_{ij}^* m_{ij}}{\sum_{i=1}^{3}\sum_{j=1}^{2} m_{ij}^2}$$

With $\bar{\mathbf{R}}^*$ and λ, we can now compute the pose with ambiguity as we did in the error-free case

Summary

2D-3D pose estimation:

- Homography-based method

$$\mathbf{H}_{i\Pi} = \mathbf{K}\left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}\right]$$

- Minimising geometric/reprojection error

$$\mathbf{T}_{wc}^{*} = \underset{\mathbf{T}_{wc}}{\operatorname{argmin}} \sum_{i} \left\| \pi(\mathbf{T}_{wc}^{-1} \cdot \mathbf{x}_i^{w}) - \mathbf{u}_i \right\|^2$$