Motion Tracking

Dr Fangcheng Zhong

UNIVERSITY OF CAMBRIDGE

Last Lecture

- Pose estimation with calibration patterns — unavailable in most XR applications
- Solution: multi-view geometry — correspondence between images of multiple views

Outline

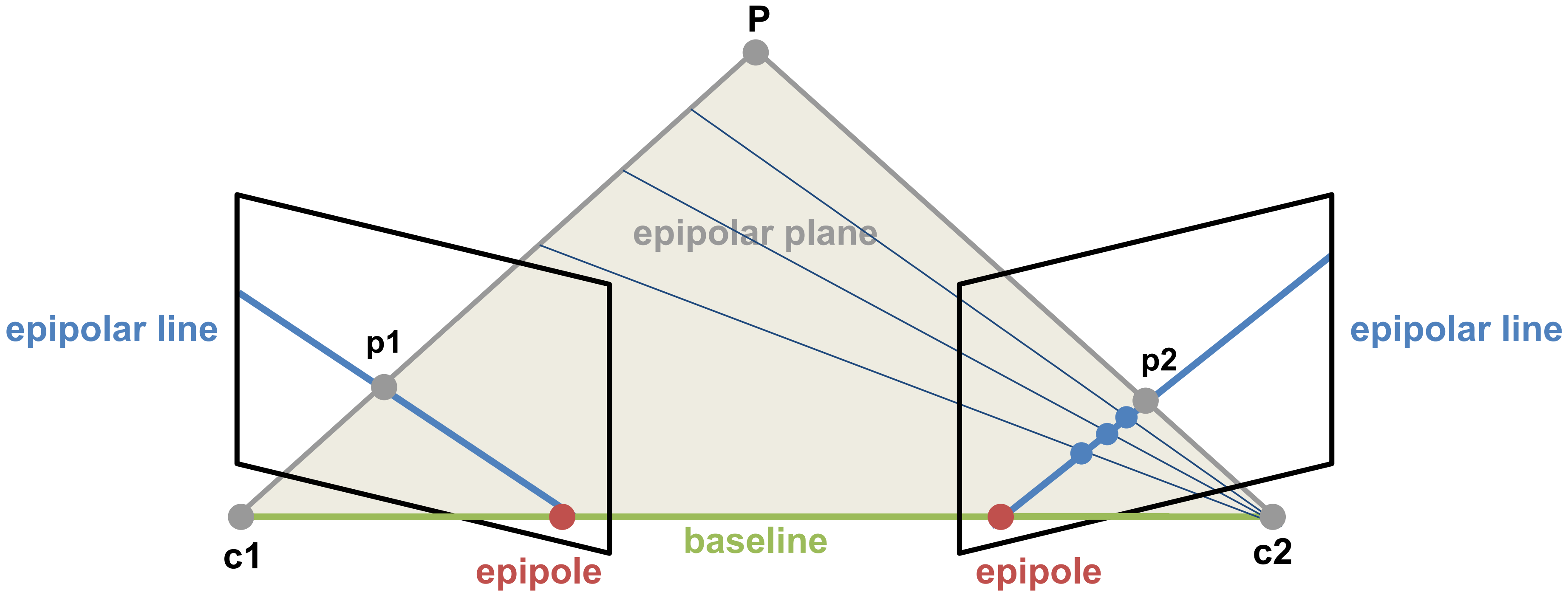
- Multi-view geometry
  - epipolar geometry
  - essential & fundamental matrix
  - stereo calibration
  - triangulation
  - bundle adjustment
- Device tracking
  - simultaneous localization and mapping (SLAM)
  - inertial measurement unit (IMU)

UNIVERSITY OF CAMBRIDGE

Epipolar Geometry

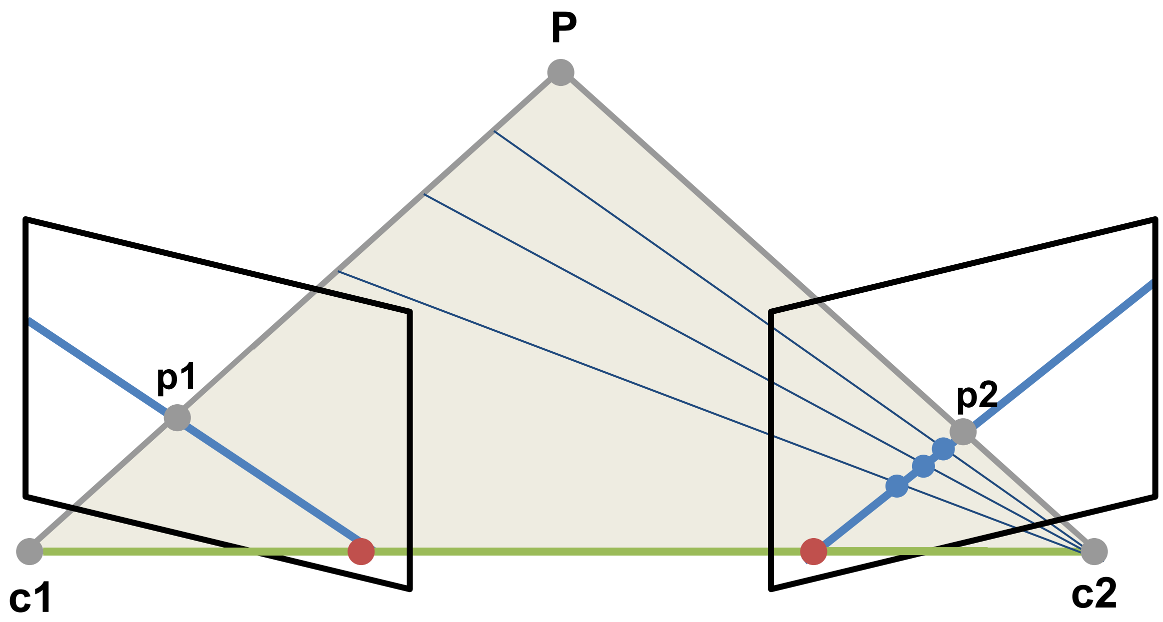
- Geometric relationship between two images of the same scene taken from distinct camera viewpoints

Epipolar Geometry



epipolar constraints

Epipolar Geometry



P' = RP + T

T x P' = T x RP

P'·(T x P') = P'·(T x RP)

0 = P'·(T x RP)

P'·([Tx]RP) = 0

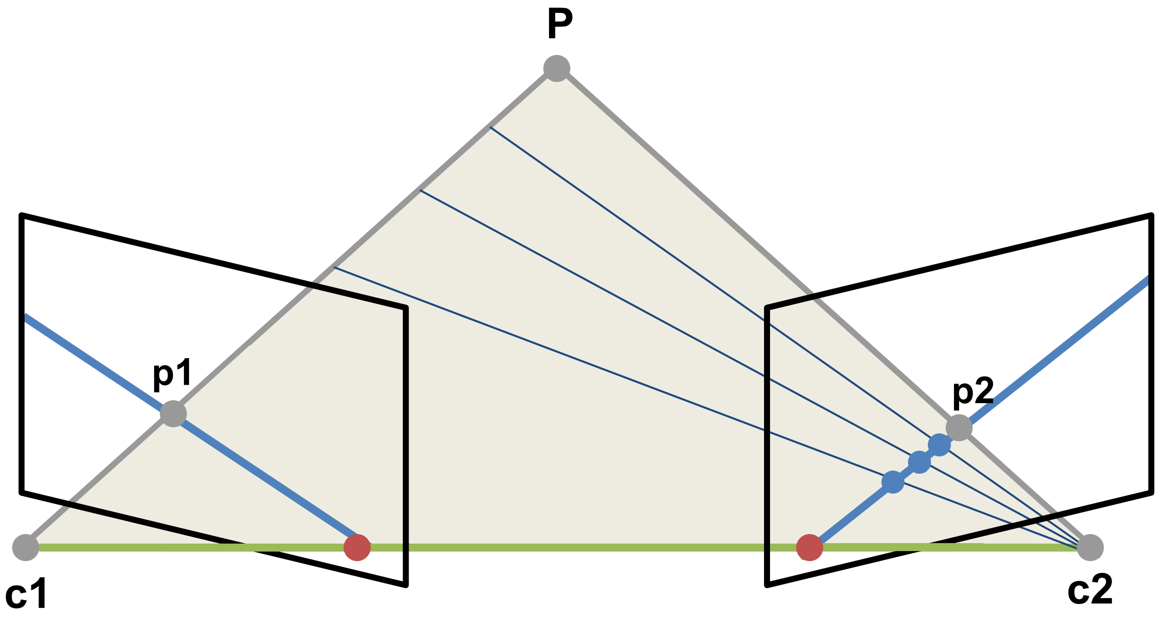
Let **E = [Tx]R**

$P'^{T}EP = 0$

E: essential matrix

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = [\mathbf{a}_{\times}]\,\mathbf{b}$$

Epipolar Geometry

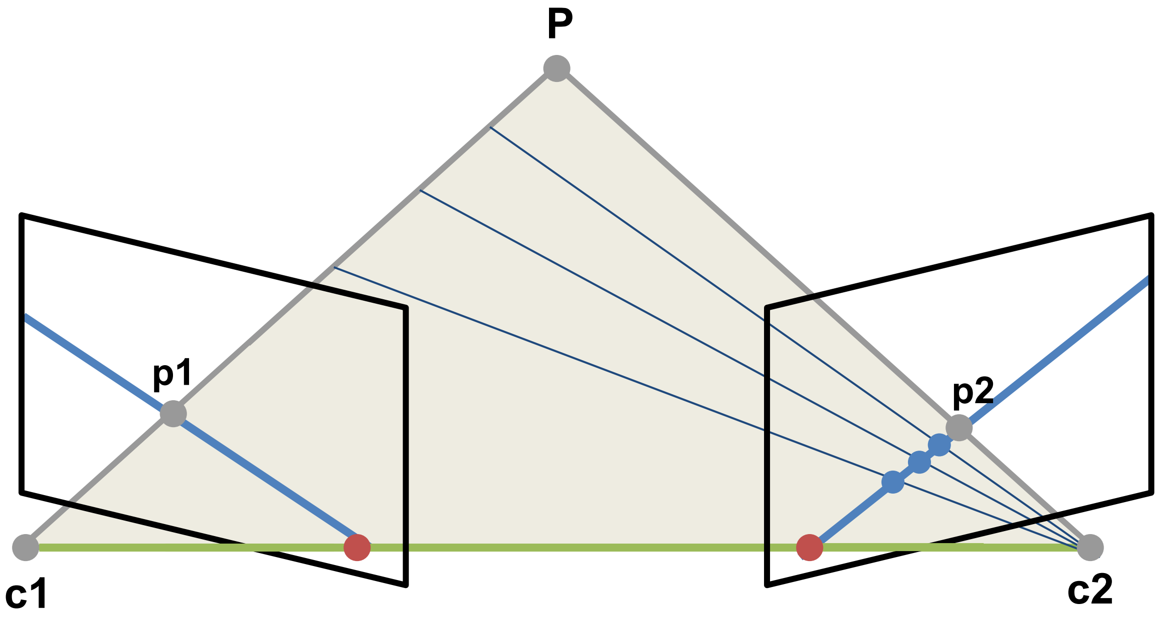


Can we establish a relationship between two image points $\mathbf{p}_1$ and $\mathbf{p}_2$?

Recall the intrinsic matrix

$$\begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & s & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsic matrix } \mathbf{K}} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

Epipolar Geometry



$$\mathbf{P}'^T \mathbf{E} \mathbf{P} = \mathbf{0}$$

$$\left(\mathbf{K}_1^{-1}\mathbf{p}_1\right)^T \mathbf{E} \left(\mathbf{K}_2^{-1}\mathbf{p}_2\right) = \mathbf{0}$$

$$\mathbf{p}_1^T \underbrace{\left(\mathbf{K}_1^{-1}\right)^T \mathbf{E}\,\mathbf{K}_2^{-1}}_{\text{fundamental matrix } \mathbf{F}} \mathbf{p}_2 = \mathbf{0}$$

$$\mathbf{p}_1^T \mathbf{F}\, \mathbf{p}_2 = \mathbf{0}$$

Q: does it bother you that the intrinsic matrix is invertible?
If $\mathbf{p}_1^T\mathbf{F} = \mathbf{0}$ or $\mathbf{F}\mathbf{p}_2 = \mathbf{0}$, what can you tell about $\mathbf{p}_1$ and $\mathbf{p}_2$?

Epipolar Geometry

- 8-point algorithm: find the fundamental matrix with correspondence between stereo images

$$\boldsymbol{p}_i' F \boldsymbol{p}_i = 0$$

$$\begin{bmatrix} x_i' & y_i' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

$$x_i x_i' f_{11} + x_i y_i' f_{21} + x_i f_{31} + y_i x_i' f_{12} + y_i y_i' f_{22} + y_i f_{32} + x_i' f_{13} + y_i' f_{23} + f_{33} = 0$$

Epipolar Geometry

$$\underbrace{\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx'_m & x_my'_m & x_m & y_mx'_m & y_my'_m & y_m & x'_m & y'_m & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix}}_{\mathbf{f}} = \mathbf{0}$$

- The solution vector **f** holds for an arbitrary scale
- Can apply the same DLT trick to find **f** that minimises ||**Af**|| subject to a unit vector constraint ||**f**||=1
- But the fundamental matrix is not full rank! why?

Epipolar Geometry

- The fundamental matrix and the essential matrix are both singular because the cross product matrix is singular

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = [\mathbf{a}_\times]\,\mathbf{b}$$

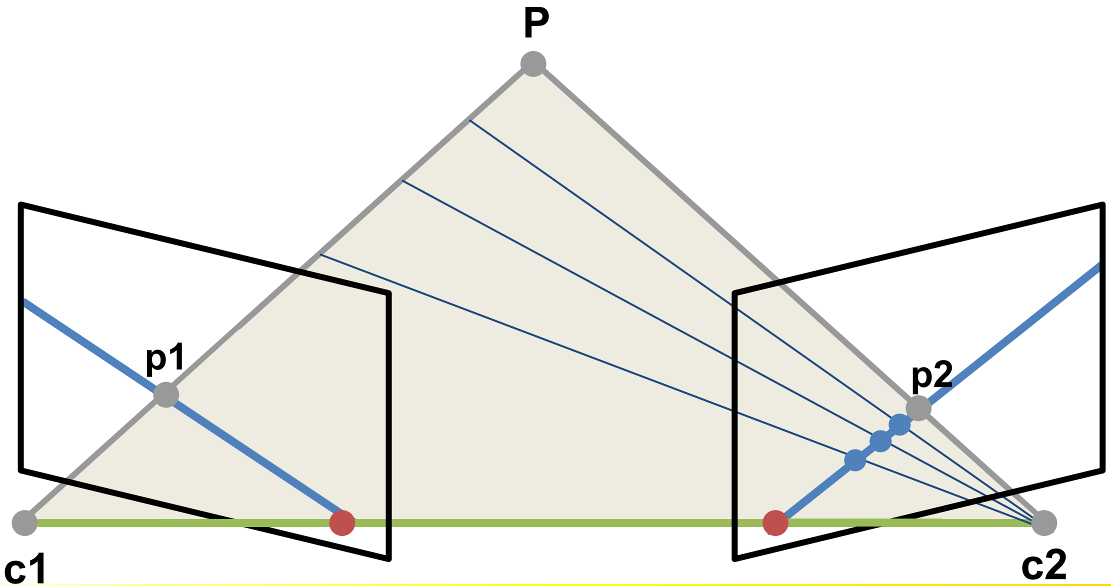
- Enforce the singularity of F by enforcing the smallest singular value of F to be zero
- Exercise: find the essential matrix from the fundamental matrix and extract the relative camera pose

Epipolar Geometry

- Single camera calibration
  - Correspondence between known world points and image points
  - Estimate the camera pose relative to the world frame
- Multi-view calibration
  - Correspondence between image points from different views
  - Estimate the camera pose relative to the first/latest camera frame

Triangulation

- Compute the 3D location of a matching pair of image points from calibrated stereo images



Bundle Adjustment

- Simultaneous refinement of the **3D coordinates** and the **camera parameters**, minimizing the sum of the squared **reprojection errors** between the observed image points and the projections of the 3D points across multiple camera views
- Can be solved as a nonlinear least square problem (e.g. LM algorithm)

$$\arg\min_{\mathbf{C},\mathbf{p}} \sum_{i,j} \| v_{ij} \left(\mathbf{C}_i(\mathbf{p}_j) - \mathbf{x}_{i,j} \right) \|^2$$

$\mathbf{C}_i$ — camera parameters of the i-th view
$\mathbf{p}_j$ — j-th 3D point in world coordinates
$\mathbf{x}_{ij}$ — ground-truth image coordinates of the i-th 3D point in the j-th view
v_{ij} — visibility of the i-th 3D point in the j-th view

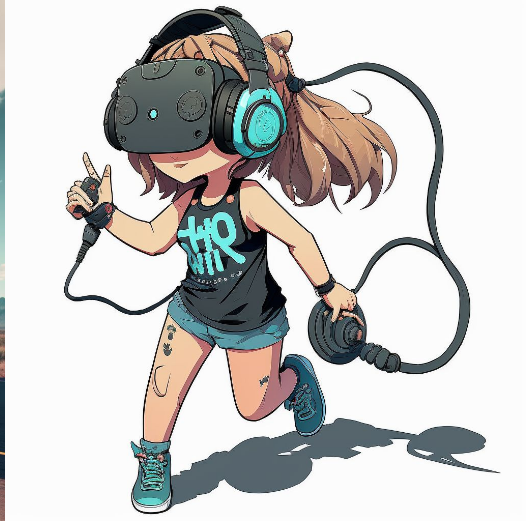
Simultaneous Localization and Mapping (SLAM)

- The problem of constructing a **map** (featured point cloud) of an unknown environment while simultaneously determining the **pose** (position + orientation) of a device/agent coupled with a (or multiple) **sensor**(s)

Simultaneous Localization and Mapping (SLAM)

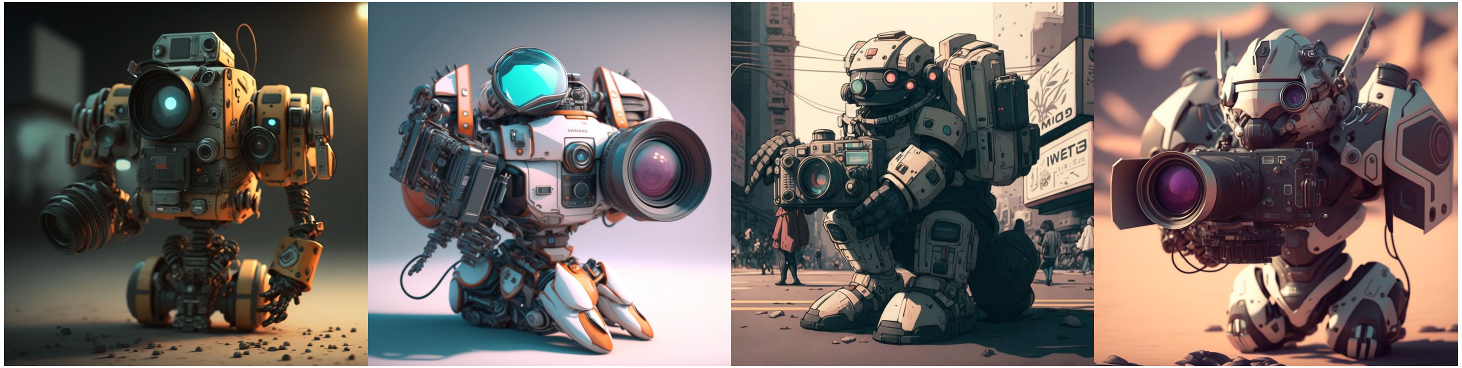
- Applications in robotics, navigation, extended reality, autonomous driving, object recognition, etc.





V-SLAM

- Visual-SLAM (V-SLAM) when the process takes only **visual inputs** from the camera
- Key steps — initialisation, tracking, mapping, loop closure



V-SLAM

- Initialisation
  - purpose
    - initialising the camera pose and the map of the environment
  - steps
    - set the initial pose of the first camera to be [**I** |**0**]
    - estimate the essential matrices and poses for the starting frames from matching features in the first few images
    - establish the initial map using triangulation and feature matching

V-SLAM

- Tracking
  - purpose
    - updating the camera pose
  - steps
    - estimate the essential matrix by matching features in the current frame to features in the previous frame (or keyframe)
    - Kalman filtering to smooth out noise or uncertainty in the estimation

V-SLAM

- Mapping
  - purpose
    - adding new features to the map and updating the positions of existing features
  - steps
    - feature extraction and matching
    - bundle adjustment to minimise the reprojection error by adjusting the camera pose and the 3D point cloud

V-SLAM

- Loop closure
  - purpose
    - correcting the accumulated drift in the camera pose and map if the the camera has revisited a location it has previously visited
  - steps
    - loop closure detection (image matching, feature matching)
    - pose alignment (perspective-n-point, iterative closest point)
    - map update (merging features)

Inertial Measurement Unit (IMU)

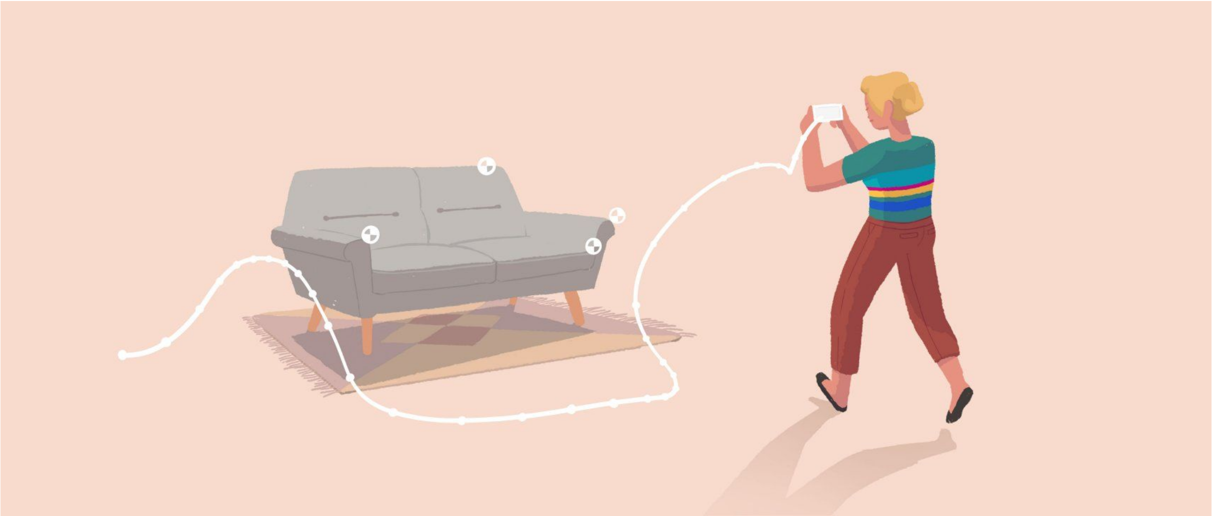
- An IMU is a device for measuring acceleration, both linear and angular
- The **accelerometer** measures linear acceleration in m/s^2
- The **gyroscope** measures angular velocity in degrees/s
- Integration to recover the accumulated changes in position and orientation

Inertial Measurement Unit (IMU)

- Faster pose measurement than V-SLAM but prone to accumulated error
- Correcting error overtime via V-SLAM

Inside-Out Tracking

- Device tracking via cameras/sensors located on the device, no need for other external equipments



Outside-In Tracking

- Device tracking via external sensors, cameras, or markers (i.e. tracking constrained to specific area)