Robotic 3D Vision

Lecture 11: Visual SLAM 2 – Online SLAM, Indirect EKF-SLAM

WS 2020/21

Dr. Niclas Zeller
Artisense GmbH

What We Will Cover Today

- Online Visual SLAM
- EKF-SLAM
- Mono-SLAM (EKF-SLAM for monocular cameras)

Recap: What is Visual SLAM ?

- SLAM stands for Simultaneous Localization and Mapping
  - Estimate the pose of the camera in a map, and simultaneously
  - Reconstruct the environment map
- Visual SLAM (VSLAM): SLAM with vision sensors
- Loop-closure: Revisiting a place allows for drift compensation

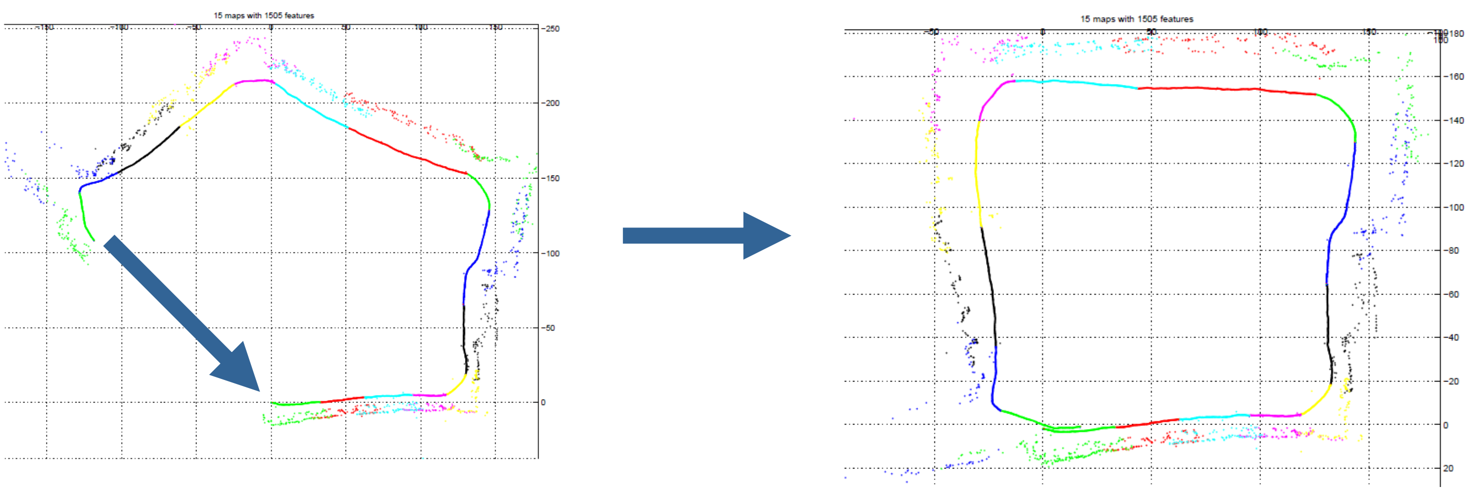
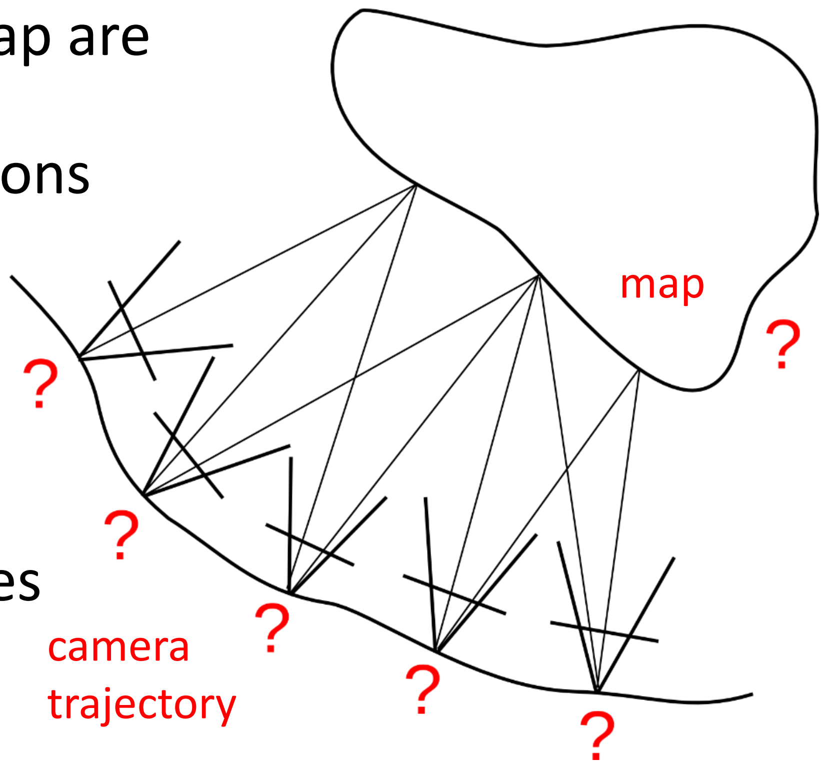


Image from Clemente et al., RSS 2007

Recap: Why is SLAM difficult?

- Chicken-or-egg problem
  - Camera trajectory and map are unknown and need to be estimated from observations
  - Accurate localization requires an accurate map
  - Accurate mapping requires accurate localization
- How can we solve this problem efficiently and robustly?



Recap: Definition of Visual SLAM

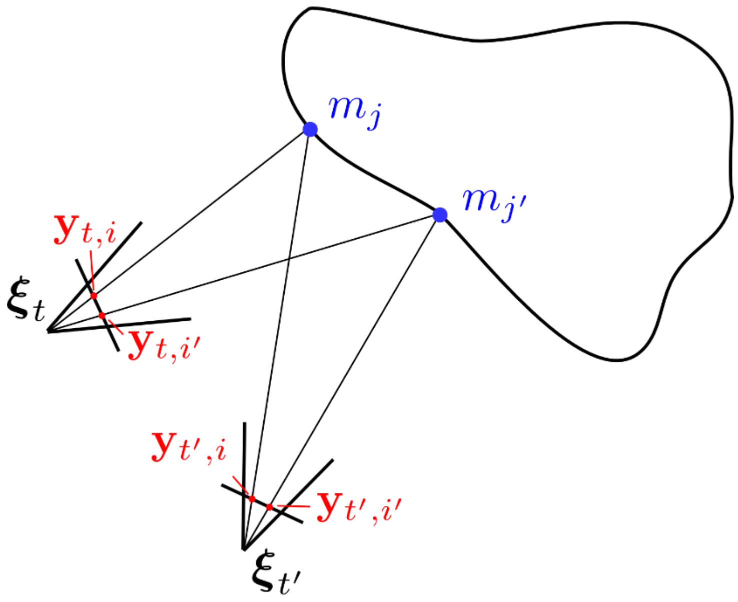
- Visual SLAM is the process of simultaneously estimating the egomotion of an object and the environment map using only inputs from visual sensors on the object and (if available) control inputs

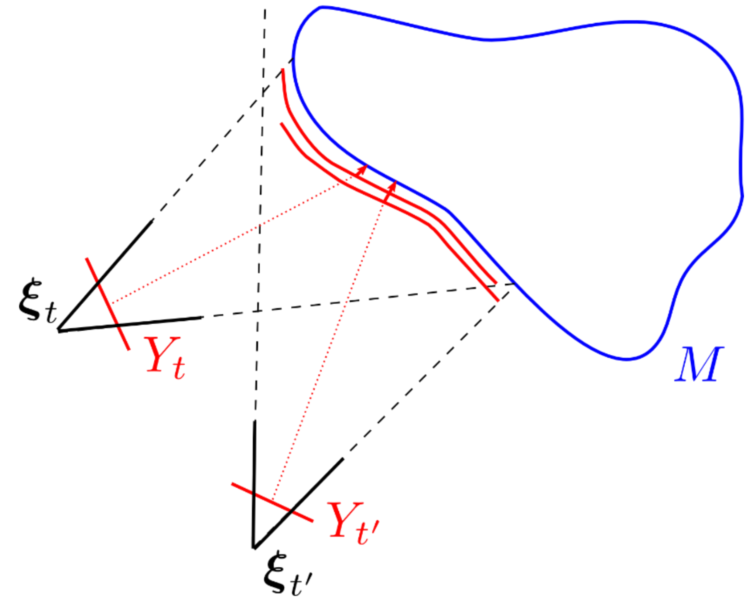
- **Inputs:** images at discrete time steps t,
  - Monocular case: Set of images $I_{0:t} = \{I_0, \ldots, I_t\}$
  - Stereo case: Left/right images $I^l_{0:t} = \{I^l_0, \ldots, I^l_t\}$ $I^r_{0:t} = \{I^r_0, \ldots, I^r_t\}$
  - RGB-D case: Color/depth images $I_{0:t} = \{I_0, \ldots, I_t\}$ $Z_{0:t} = \{Z_0, \ldots, Z_t\}$

  - Robotics: **control inputs** $U_{1:t}$ often not considered/available

- **Output**:
  - Camera pose estimates $\mathbf{T}_t \in \mathbf{SE(3)}$ in world reference frame. For convenience, we also write $\boldsymbol{\xi}_t = \boldsymbol{\xi}\left(\mathbf{T}_t\right)$
    - reference frame is in general anchored at the first frame avalibale
  - Environment map M

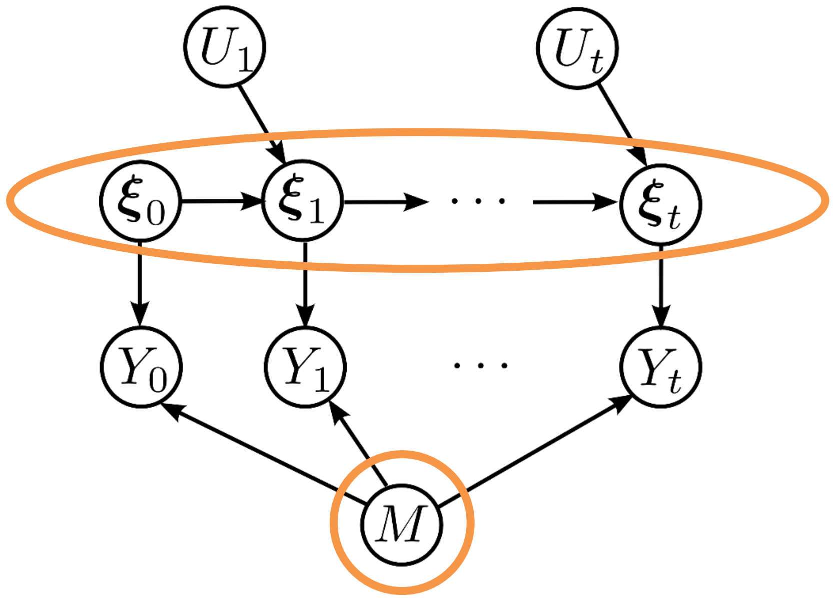
Recap: Map Observations in Visual SLAM





- With Y_t we denote observations of the environment map in image I_t, f.e.
  - Indirect point-based method: $Y_t = \{\mathbf{y}_{t,1}, \ldots, \mathbf{y}_{t,N}\}$ (2D or 3D image points)
  - Direct method (mono): $Y_t = \{I_t\}$ (set of image pixels)
  - Direct RGB-D method: $Y_t = \{I_t, Z_t\}$ (all image pixels)
  - ...
- Involves data association to map elements $M = \{m_1, \ldots, m_S\}$
  - We denote correspondences by $c_{t,i} = j, 1 \leq i \leq N, 1 \leq j \leq S$

Recap: Probabilistic Formulation of Visual SLAM



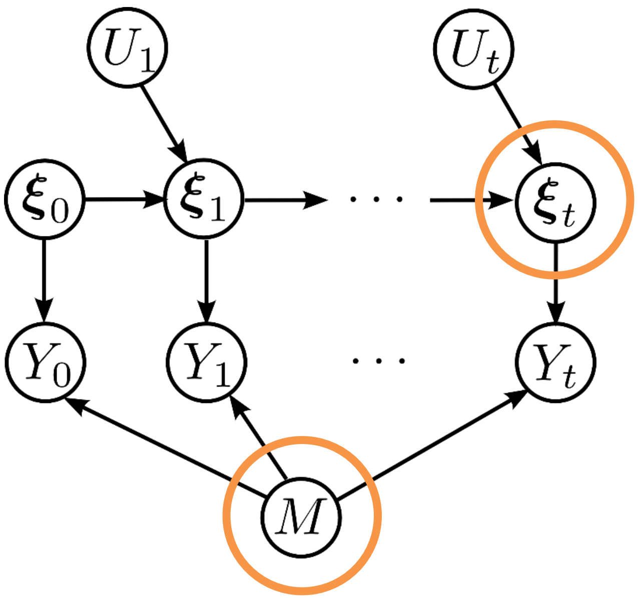
- SLAM posterior probability: $p\left(\boldsymbol{\xi}_{0:t}, M \mid Y_{0:t}, U_{1:t}\right)$
- Observation likelihood: $p\left(Y_t \mid \boldsymbol{\xi}_t, M\right)$
- State-transition probability: $p\left(\boldsymbol{\xi}_t \mid \boldsymbol{\xi}_{t-1}, U_t\right)$

Online SLAM Methods

- Marginalize out previous poses

$$p\left(\boldsymbol{\xi}_t, M \mid Y_{0:t}, U_{1:t}\right) =$$

$$\int \cdots \int p\left(\boldsymbol{\xi}_{0:t}, M \mid Y_{0:t}, U_{1:t}\right) d\boldsymbol{\xi}_{t-1} \ldots d\boldsymbol{\xi}_0$$



- Poses can be marginalized individually in a recursive way

- Variants:
  - Tracking-and-Mapping: Alternating pose and map estimation
  - Probabilistic filters, f.e. EKF-SLAM

Recap: Bayesian Filter

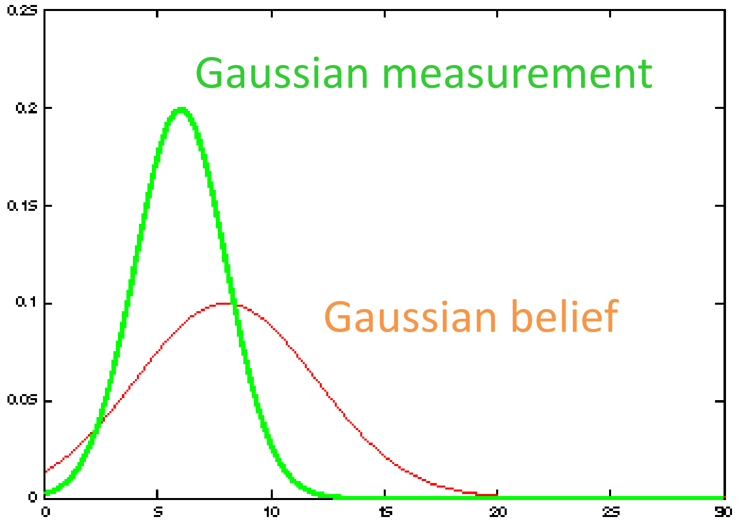
Prediction

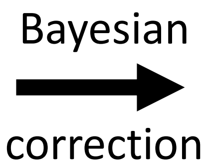
$$p\left(X_t \mid Y_{0:t-1}, U_{1:t}\right) = \int p\left(X_t \mid X_{t-1}, U_t\right)\, p\left(X_{t-1} \mid Y_{0:t-1}, U_{1:t-1}\right) dX_{t-1}$$

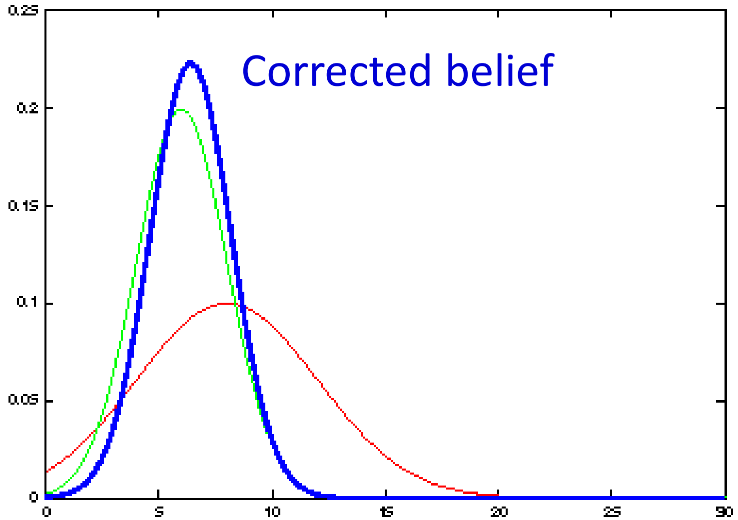
Correction

$$p\left(X_t \mid Y_{0:t}, U_{1:t}\right) = \eta\, p\left(Y_t \mid X_t\right) p\left(X_t \mid Y_{0:t-1}, U_{1:t}\right)$$

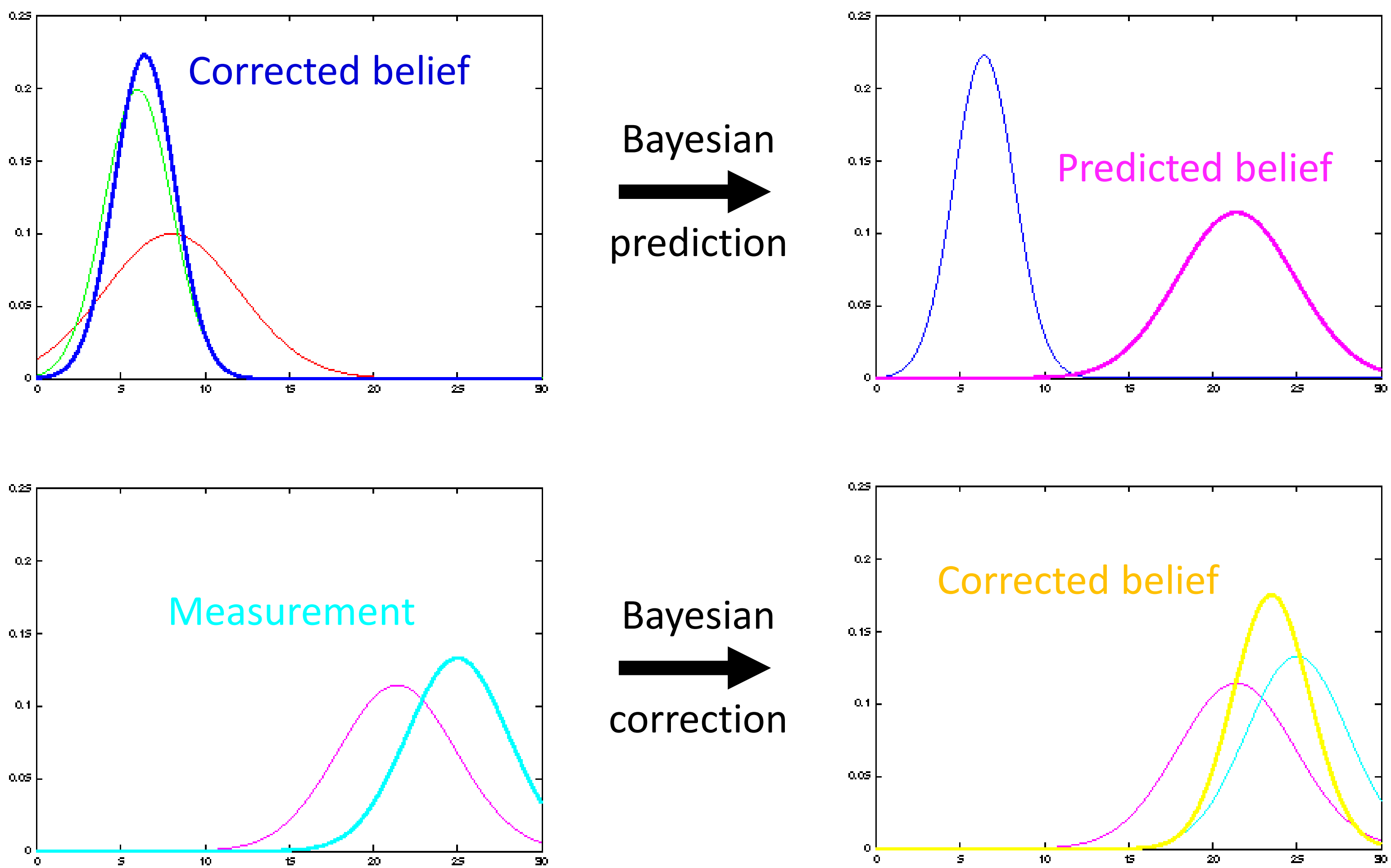
Recap: Kalman Filter







Recap: Kalman Filter



Recap: Extended Kalman Filter

- Non-linear models with Gaussian noise

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}, \mathbf{u}_t) + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{d_t}\right)$$

$$\mathbf{y}_t = h(\mathbf{x}_t) + \boldsymbol{\delta}_t, \boldsymbol{\delta}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{m_t}\right)$$

- Prediction

$$\mathbf{x}_t^- = g(\mathbf{x}_{t-1}^+, \mathbf{u}_t)$$

$$\boldsymbol{\Sigma}_t^- = \mathbf{G}_t \boldsymbol{\Sigma}_{t-1}^+ \mathbf{G}_t^\top + \boldsymbol{\Sigma}_{d_t}$$

$$\mathbf{G}_t = \nabla_{\mathbf{x}} g(\mathbf{x}, \mathbf{u}_t)\big|_{\mathbf{x}=\mathbf{x}_{t-1}^+}$$

- Correction

$$\mathbf{K}_t = \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top \left(\mathbf{H}_t \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top + \boldsymbol{\Sigma}_{m_t}\right)^{-1}$$

$$\mathbf{x}_t^+ = \mathbf{x}_t^- + \mathbf{K}_t \left(\mathbf{y}_t - h\left(\mathbf{x}_t^-\right)\right)$$

$$\boldsymbol{\Sigma}_t^+ = \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}_t\right) \boldsymbol{\Sigma}_t^-$$

$$\mathbf{H}_t = \nabla_{\mathbf{x}} h(\mathbf{x})\big|_{\mathbf{x}=\mathbf{x}_t^-}$$

SLAM with Extended Kalman Filters

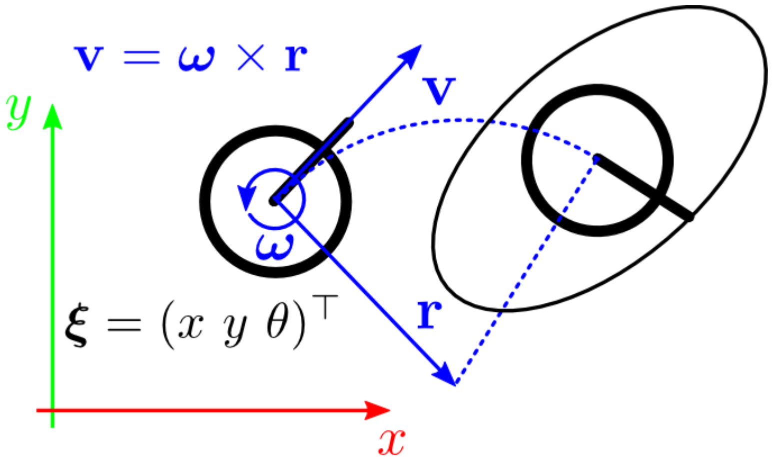
- Detected keypoint $\mathbf{y}_i$ in an image observes „landmark" position $\mathbf{m}_j$ in the map $M = \{\mathbf{m}_1, \ldots, \mathbf{m}_S\}$

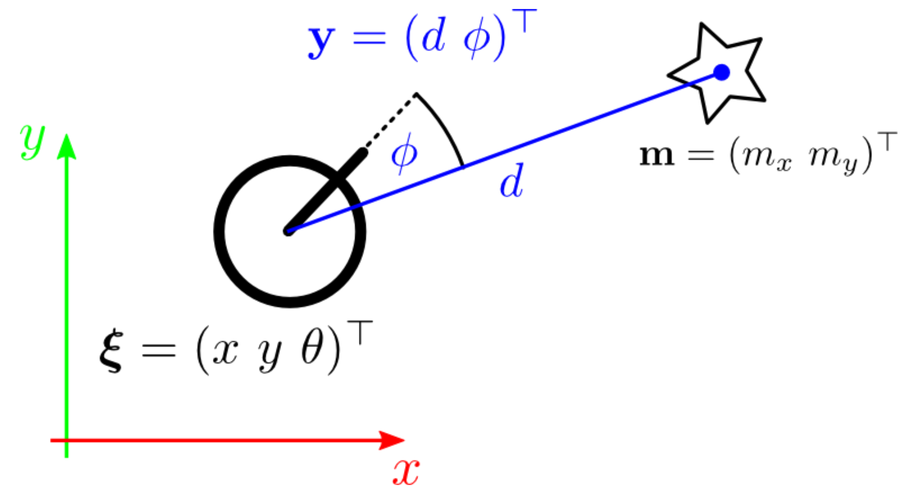
- Idea: Include landmarks into state-variable

$$\mathbf{x}_t = \begin{pmatrix} \boldsymbol{\xi}_t \\ \mathbf{m}_{t,1} \\ \vdots \\ \mathbf{m}_{t,S} \end{pmatrix} \qquad \boldsymbol{\Sigma}_t = \begin{pmatrix} \boldsymbol{\Sigma}_{t,\boldsymbol{\xi}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{t,\boldsymbol{\xi}\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\boldsymbol{\xi}\mathbf{m}_S} \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_1\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_S} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_S\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_S} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\Sigma}_{t,\boldsymbol{\xi}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{t,\boldsymbol{\xi}\mathbf{m}} \\ \boldsymbol{\Sigma}_{t,\mathbf{m}\boldsymbol{\xi}} & \boldsymbol{\Sigma}_{t,\mathbf{m}\mathbf{m}} \end{pmatrix}$$

EKF-SLAM in a 2D World





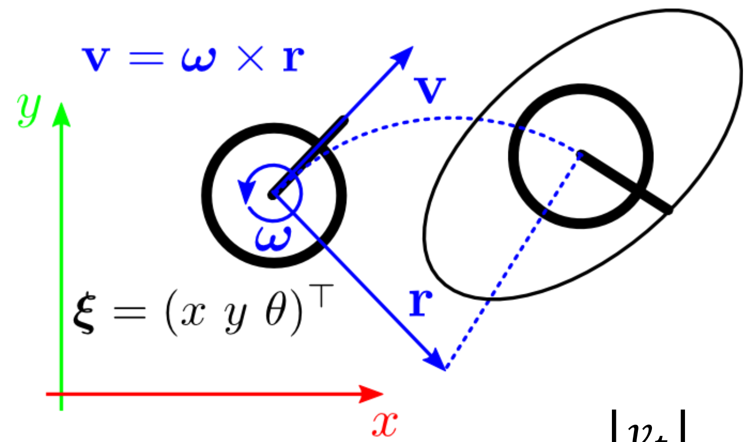
- For simplicity, let‘s assume
  - 3-DoF camera motion on a 2D plane
  - 2D range-and-bearing measurements of 2D landmarks
  - Only one measurement at a time
  - Known data association

2D EKF-SLAM State-Transition Model

- State/control variables

$$\boldsymbol{\xi}_t = (x_t \; y_t \; \theta_t)^\top \qquad \mathbf{m}_{t,j} = (m_{t,j,x} \; m_{t,j,y})^\top$$

$$\mathbf{u}_t = (v_t \; \omega_t)^\top$$



$$r_t = \left|\frac{v_t}{\omega_t}\right|$$

- State-transition model

$$\boldsymbol{\xi}_t = g_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{t-1}, \mathbf{u}_t) + \boldsymbol{\epsilon}_{\boldsymbol{\xi},t} \qquad \boldsymbol{\epsilon}_{\boldsymbol{\xi},t} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{d_t,\boldsymbol{\xi}}\right)$$

  - Pose:

$$g_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{t-1}, \mathbf{u}_t) = \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ \theta_{t-1} \end{pmatrix} + \begin{pmatrix} -\frac{v_t}{\omega_t} \sin\theta_{t-1} + \frac{v_t}{\omega_t} \sin(\theta_{t-1} + \omega_t \Delta t) \\ \frac{v_t}{\omega_t} \cos\theta_{t-1} - \frac{v_t}{\omega_t} \cos(\theta_{t-1} + \omega_t \Delta t) \\ \omega_t \Delta t \end{pmatrix}$$

  - Landmarks: $\mathbf{m}_t = g_{\mathbf{m}}(\mathbf{m}_{t-1}) = \mathbf{m}_{t-1}$
  - Combined:

$$\mathbf{x}_t = g(\mathbf{x}_{t-1}, \mathbf{u}_t) + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{d_t}\right) \qquad g(\mathbf{x}_{t-1}, \mathbf{u}_t) = \begin{pmatrix} g_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{t-1}, \mathbf{u}_t) \\ g_{\mathbf{m}}(\mathbf{m}_{t-1}) \end{pmatrix} \quad \boldsymbol{\Sigma}_{d_t} = \begin{pmatrix} \boldsymbol{\Sigma}_{d_t,\boldsymbol{\xi}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

2D EKF-SLAM Observation Model

- State/measurement variables

$$\mathbf{y}_t = (d_t \;\; \phi_t)^\top \qquad \mathbf{m}_{t,j} = (m_{t,j,x} \;\; m_{t,j,y})^\top$$

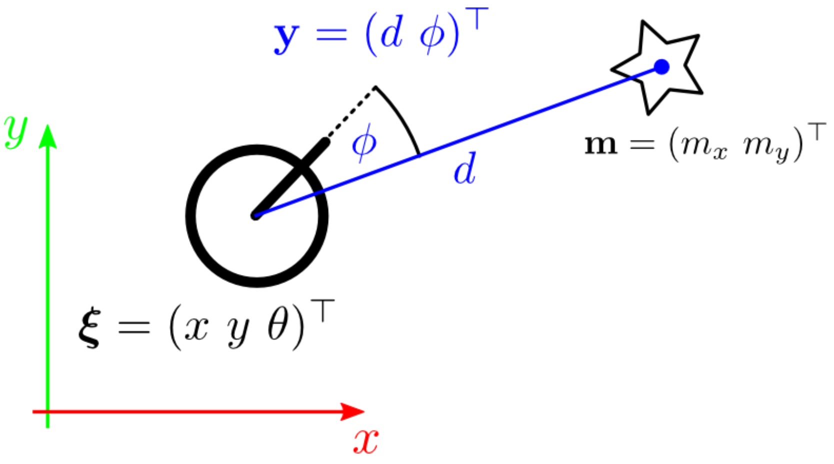
- Observation model:

$$\mathbf{y}_t = h(\boldsymbol{\xi}_t, \mathbf{m}_{t,c_t}) + \boldsymbol{\delta}_t \qquad \boldsymbol{\delta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{m_t})$$

$$h(\boldsymbol{\xi}_t, \mathbf{m}_{t,c_t}) = \begin{pmatrix} \left\|\mathbf{m}^{\text{rel}}_{t,c_t}\right\|_2 \\ \operatorname{atan2}\left(m^{\text{rel}}_{t,c_t,y}, m^{\text{rel}}_{t,c_t,x}\right) \end{pmatrix}$$

angle with respect to the heading of the robot

$$\mathbf{m}^{\text{rel}}_{t,c_t} := \mathbf{R}(-\theta_t)\left(\mathbf{m}_{t,c_t} - (x_t \;\; y_t)^\top\right)$$



State Initialization

- First frame:
  - Initialize reference frame at zero $\mathbf{x}_0^- = \mathbf{0}$
  - Set pose covariance to zero (anchoring) $\boldsymbol{\Sigma}_{0,\xi\xi}^- = \mathbf{0}$
- New landmark:
  - Initial position unknown
  - Initialize mean at zero $\boldsymbol{\Sigma}_{0,\xi\mathbf{m}}^- = {\boldsymbol{\Sigma}_{0,\mathbf{m}\xi}^-}^\top = \mathbf{0}$
  - Initialize covariance to infinity (large value) $\boldsymbol{\Sigma}_{0,\mathbf{mm}}^- = \infty\mathbf{I}$

Evolution of State Estimate on Prediction

- How is the state estimate modified on a state-transition?
- Recap: Prediction

$$\mathbf{G}_t = \nabla_{\mathbf{x}} g(\mathbf{x}, \mathbf{u}_t)\big|_{\mathbf{x}=\mathbf{x}_{t-1}^+}$$

$$\mathbf{x}_t^- = g(\mathbf{x}_{t-1}^+, \mathbf{u}_t) \qquad \mathbf{\Sigma}_t^- = \mathbf{G}_t \mathbf{\Sigma}_{t-1}^+ \mathbf{G}_t^\top + \mathbf{\Sigma}_{d_t}$$

$$\mathbf{G}_{t,\xi} := \nabla_{\xi} g_{\xi}(\xi, \mathbf{u}_t)\big|_{\xi=\xi_{t-1}^+}$$

↓

$$\mathbf{x}_t^- = \begin{pmatrix} g_{\xi}(\xi_{t-1}^+, \mathbf{u}_t) \\ \mathbf{m}_{t-1}^+ \end{pmatrix}$$

only the mean pose is updated!

$$\mathbf{x}_t = \begin{pmatrix} \xi_t \\ \mathbf{m}_{t,1} \\ \vdots \\ \mathbf{m}_{t,S} \end{pmatrix}$$

Evolution of State Estimate on Prediction

- How is the state estimate modified on a state-transition?
- Recap: Prediction

$$\mathbf{G}_t = \nabla_{\mathbf{x}} g(\mathbf{x}, \mathbf{u}_t)\big|_{\mathbf{x}=\mathbf{x}_{t-1}^+}$$

$$\mathbf{x}_t^- = g(\mathbf{x}_{t-1}^+, \mathbf{u}_t) \qquad \boldsymbol{\Sigma}_t^- = \mathbf{G}_t \boldsymbol{\Sigma}_{t-1}^+ \mathbf{G}_t^\top + \boldsymbol{\Sigma}_{d_t} \qquad \mathbf{G}_{t,\xi} := \nabla_{\boldsymbol{\xi}} g_{\boldsymbol{\xi}}(\boldsymbol{\xi}, \mathbf{u}_t)\big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{t-1}^+}$$

$$\begin{pmatrix} \mathbf{G}_{t,\xi} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{t-1,\xi\xi}^+ & \boldsymbol{\Sigma}_{t-1,\xi\mathbf{m}}^+ \\ \boldsymbol{\Sigma}_{t-1,\mathbf{m}\xi}^+ & \boldsymbol{\Sigma}_{t-1,\mathbf{mm}}^+ \end{pmatrix} \begin{pmatrix} \mathbf{G}_{t,\xi}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Sigma}_{d_t,\xi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} =$$

$$\begin{pmatrix} \mathbf{G}_{t,\xi} \boldsymbol{\Sigma}_{t-1,\xi\xi}^+ \mathbf{G}_{t,\xi}^\top + \boldsymbol{\Sigma}_{d_t,\xi} & \mathbf{G}_{t,\xi} \boldsymbol{\Sigma}_{t-1,\xi\mathbf{m}}^+ \\ \boldsymbol{\Sigma}_{t-1,\mathbf{m}\xi}^+ \mathbf{G}_{t,\xi}^\top & \boldsymbol{\Sigma}_{t-1,\mathbf{mm}}^+ \end{pmatrix}$$

covariances are transformed to the new pose!

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \boldsymbol{\Sigma}_{t,\xi\xi} & \boldsymbol{\Sigma}_{t,\xi\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\xi\mathbf{m}_S} \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_1\xi} & \boldsymbol{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_S} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_S\xi} & \boldsymbol{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_S} \end{pmatrix}$$

Evolution of State Estimate on Correction

- How is the state estimate modified on a landmark measurement?
- Recap: Correction

$$\mathbf{K}_t = \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top \left(\mathbf{H}_t \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top + \boldsymbol{\Sigma}_{m_t} \right)^{-1}$$

$$\mathbf{H}_t = \nabla_{\mathbf{x}} h(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_t^-}$$

$$\mathbf{x}_t^+ = \mathbf{x}_t^- + \mathbf{K}_t \left(\mathbf{y}_t - h\left(\mathbf{x}_t^-\right)\right)$$

$$\boldsymbol{\Sigma}_t^+ = \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}_t\right) \boldsymbol{\Sigma}_t^-$$

Evolution of State Estimate on Correction

- Let's have a closer look at the Kalman gain

$$\mathbf{K}_t = \mathbf{\Sigma}_t^- \mathbf{H}_t^\top \left(\mathbf{H}_t \mathbf{\Sigma}_t^- \mathbf{H}_t^\top + \mathbf{\Sigma}_{m_t} \right)^{-1} \qquad \mathbf{H}_t = \nabla_{\mathbf{x}} h(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_t^-}$$

- The Jacobian of the observation function is only non-zero for the pose and the measured landmark:

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{H}_{t,\boldsymbol{\xi}} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_{t,\mathbf{m}_{c_t}} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}$$

$$\mathbf{H}_{t,\boldsymbol{\xi}} = \nabla_{\boldsymbol{\xi}} h\left(\boldsymbol{\xi}, \mathbf{m}_{t,c_t}\right)\big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_t^-} \qquad \mathbf{H}_{t,\mathbf{m}_{c_t}} = \nabla_{\mathbf{m}_{c_t}} h\left(\boldsymbol{\xi}_t, \mathbf{m}_{c_t}\right)\Big|_{\mathbf{m}_{c_t}=\mathbf{m}_{t,c_t}^-}$$

Evolution of State Estimate on Correction

- Let's have a closer look at the Kalman gain

$$\mathbf{K}_t = \mathbf{\Sigma}_t^- \mathbf{H}_t^\top \left(\boxed{\mathbf{H}_t \mathbf{\Sigma}_t^- \mathbf{H}_t^\top} + \mathbf{\Sigma}_{m_t} \right)^{-1} \qquad \mathbf{H}_t = \nabla_\mathbf{x} h(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_t^-}$$

- The matrix $\mathbf{H}_t \mathbf{\Sigma}_t^- \mathbf{H}_t^\top$ only involves covariances between pose and the measured landmark:

$$\mathbf{H}_t \mathbf{\Sigma}_t^- \mathbf{H}_t^\top = \mathbf{H}_{t,\xi} \mathbf{\Sigma}_{t,\xi\xi}^- \mathbf{H}_{t,\xi}^\top + \mathbf{H}_{t,\mathbf{m}_{c_t}} \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\xi}^- \mathbf{H}_{t,\xi}^\top + \mathbf{H}_{t,\xi} \mathbf{\Sigma}_{t,\xi\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top + \mathbf{H}_{t,\mathbf{m}_{c_t}} \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top$$

$$\mathbf{\Sigma}_t^- = \begin{pmatrix} \mathbf{\Sigma}_{t,\xi\xi}^- & \mathbf{\Sigma}_{t,\xi\mathbf{m}_1}^- & \cdots & \mathbf{\Sigma}_{t,\xi\mathbf{m}_{c_t}}^- & \cdots & \mathbf{\Sigma}_{t,\xi\mathbf{m}_S}^- \\ \mathbf{\Sigma}_{t,\mathbf{m}_1\xi}^- & \mathbf{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_1}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_{c_t}}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_S}^- \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\xi}^- & \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\mathbf{m}_1}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\mathbf{m}_{c_t}}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_{c_t}\mathbf{m}_S}^- \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{t,\mathbf{m}_S\xi}^- & \mathbf{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_1}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_{c_t}}^- & \cdots & \mathbf{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_S}^- \end{pmatrix}$$

Evolution of State Estimate on Correction

$$\mathbf{K}_t = \boxed{\Sigma_t^- \mathbf{H}_t^\top} (\mathbf{H}_t \Sigma_t^- \mathbf{H}_t^\top + \Sigma_{m_t})^{-1} \qquad \mathbf{H}_t = \nabla_\mathbf{x} h(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_t^-}$$

- The matrix $\Sigma_t^- \mathbf{H}_t^\top$ stacks the covariances between the pose/the measured landmark and all state variables (pose+landmarks)

$$\Sigma_t^- \mathbf{H}_t^\top = \begin{pmatrix} \Sigma_{t,\xi\xi}^- \mathbf{H}_{t,\xi}^\top + \Sigma_{t,\xi\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \\ \Sigma_{t,\mathbf{m}_1\xi}^- \mathbf{H}_{t,\xi}^\top + \Sigma_{t,\mathbf{m}_1\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \\ \vdots \\ \Sigma_{t,\mathbf{m}_S\xi}^- \mathbf{H}_{t,\xi}^\top + \Sigma_{t,\mathbf{m}_S\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \end{pmatrix}$$

$$\Sigma_t^- = \begin{pmatrix} \Sigma_{t,\xi\xi}^- & \Sigma_{t,\xi\mathbf{m}_1}^- & \cdots & \Sigma_{t,\xi\mathbf{m}_{c_t}}^- & \cdots & \Sigma_{t,\xi\mathbf{m}_S}^- \\ \Sigma_{t,\mathbf{m}_1\xi}^- & \Sigma_{t,\mathbf{m}_1\mathbf{m}_1}^- & \cdots & \Sigma_{t,\mathbf{m}_1\mathbf{m}_{c_t}}^- & \cdots & \Sigma_{t,\mathbf{m}_1\mathbf{m}_S}^- \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{t,\mathbf{m}_{c_t}\xi}^- & \Sigma_{t,\mathbf{m}_{c_t}\mathbf{m}_1}^- & \cdots & \Sigma_{t,\mathbf{m}_{c_t}\mathbf{m}_{c_t}}^- & \cdots & \Sigma_{t,\mathbf{m}_{c_t}\mathbf{m}_S}^- \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{t,\mathbf{m}_S\xi}^- & \Sigma_{t,\mathbf{m}_S\mathbf{m}_1}^- & \cdots & \Sigma_{t,\mathbf{m}_S\mathbf{m}_{c_t}}^- & \cdots & \Sigma_{t,\mathbf{m}_S\mathbf{m}_S}^- \end{pmatrix}$$

Evolution of State Estimate on Correction

- Hence, the Kalman gain distributes information onto all state dimensions that are correlated with the pose or the measured landmark

$$\mathbf{K}_t = \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top \left(\mathbf{H}_t \boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top + \boldsymbol{\Sigma}_{m_t} \right)^{-1}$$

$$\boldsymbol{\Sigma}_t^- \mathbf{H}_t^\top = \begin{pmatrix} \boldsymbol{\Sigma}_{t,\xi\xi}^- \mathbf{H}_{t,\xi}^\top + \boldsymbol{\Sigma}_{t,\xi\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_1\xi}^- \mathbf{H}_{t,\xi}^\top + \boldsymbol{\Sigma}_{t,\mathbf{m}_1\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \\ \vdots \\ \boldsymbol{\Sigma}_{t,\mathbf{m}_S\xi}^- \mathbf{H}_{t,\xi}^\top + \boldsymbol{\Sigma}_{t,\mathbf{m}_S\mathbf{m}_{c_t}}^- \mathbf{H}_{t,\mathbf{m}_{c_t}}^\top \end{pmatrix}$$

- The correction step updates all state dimensions in the mean that are correlated with the pose or measured landmark

$$\mathbf{x}_t^+ = \mathbf{x}_t^- + \mathbf{K}_t \left(\mathbf{y}_t - h\left(\mathbf{x}_t^- \right) \right)$$

Evolution of State Estimate on Correction

- How is the state covariance updated in the correction step?

$$\boldsymbol{\Sigma}_t^+ = \left(\mathbf{I} - \boxed{\mathbf{K}_t \mathbf{H}_t\right) \boldsymbol{\Sigma}_t^-}$$

$$\begin{pmatrix} \mathbf{K}_{t,\xi}\mathbf{H}_{t,\xi} & 0 & \dots & 0 & \mathbf{K}_{t,\xi}\mathbf{H}_{t,\mathbf{m}_{c_t}} & 0 & \dots & 0 \\ \mathbf{K}_{t,\mathbf{m}_1}\mathbf{H}_{t,\xi} & 0 & \dots & 0 & \mathbf{K}_{t,\mathbf{m}_1}\mathbf{H}_{t,\mathbf{m}_{c_t}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{t,\mathbf{m}_S}\mathbf{H}_{t,\xi} & 0 & \dots & 0 & \mathbf{K}_{t,\mathbf{m}_S}\mathbf{H}_{t,\mathbf{m}_{c_t}} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}^-_{t,\xi\xi} & \boldsymbol{\Sigma}^-_{t,\xi\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}^-_{t,\xi\mathbf{m}_{c_t}} & \cdots & \boldsymbol{\Sigma}^-_{t,\xi\mathbf{m}_S} \\ \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\xi} & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\mathbf{m}_{c_t}} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\mathbf{m}_S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^-_{t,\mathbf{m}_{c_t}\xi} & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_{c_t}\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_{c_t}\mathbf{m}_{c_t}} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_{c_t}\mathbf{m}_S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^-_{t,\mathbf{m}_S\xi} & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_S\mathbf{m}_1} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_S\mathbf{m}_{c_t}} & \cdots & \boldsymbol{\Sigma}^-_{t,\mathbf{m}_S\mathbf{m}_S} \end{pmatrix}$$

$$\boldsymbol{\Sigma}^+_{t,\mathbf{m}_1\mathbf{m}_{c_t}} = \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\mathbf{m}_{c_t}} - \mathbf{K}_{t,\mathbf{m}_1}\mathbf{H}_{t,\xi}\boldsymbol{\Sigma}^-_{t,\xi\mathbf{m}_{c_t}} - \mathbf{K}_{t,\mathbf{m}_1}\mathbf{H}_{t,\mathbf{m}_{c_t}}\boldsymbol{\Sigma}^-_{t,\mathbf{m}_{c_t}\mathbf{m}_{c_t}}$$

- Covariance change for a landmark that is not the measured landmark:

$$\mathbf{K}_{t,\mathbf{m}_1} = \left(\boxed{\boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\xi}}\mathbf{H}^\top_{t,\xi} + \boldsymbol{\Sigma}^-_{t,\mathbf{m}_1\mathbf{m}_{c_t}}\mathbf{H}^\top_{t,\mathbf{m}_{c_t}}\right)\left(\mathbf{H}_t\boldsymbol{\Sigma}^-_t\mathbf{H}^\top_t + \boldsymbol{\Sigma}_{m_t}\right)^{-1}$$

non-zero if landmark not new!

Evolution of State Estimate on Correction

- The correction step updates all state dimensions in the state covariance that correlate with the pose or measured landmark

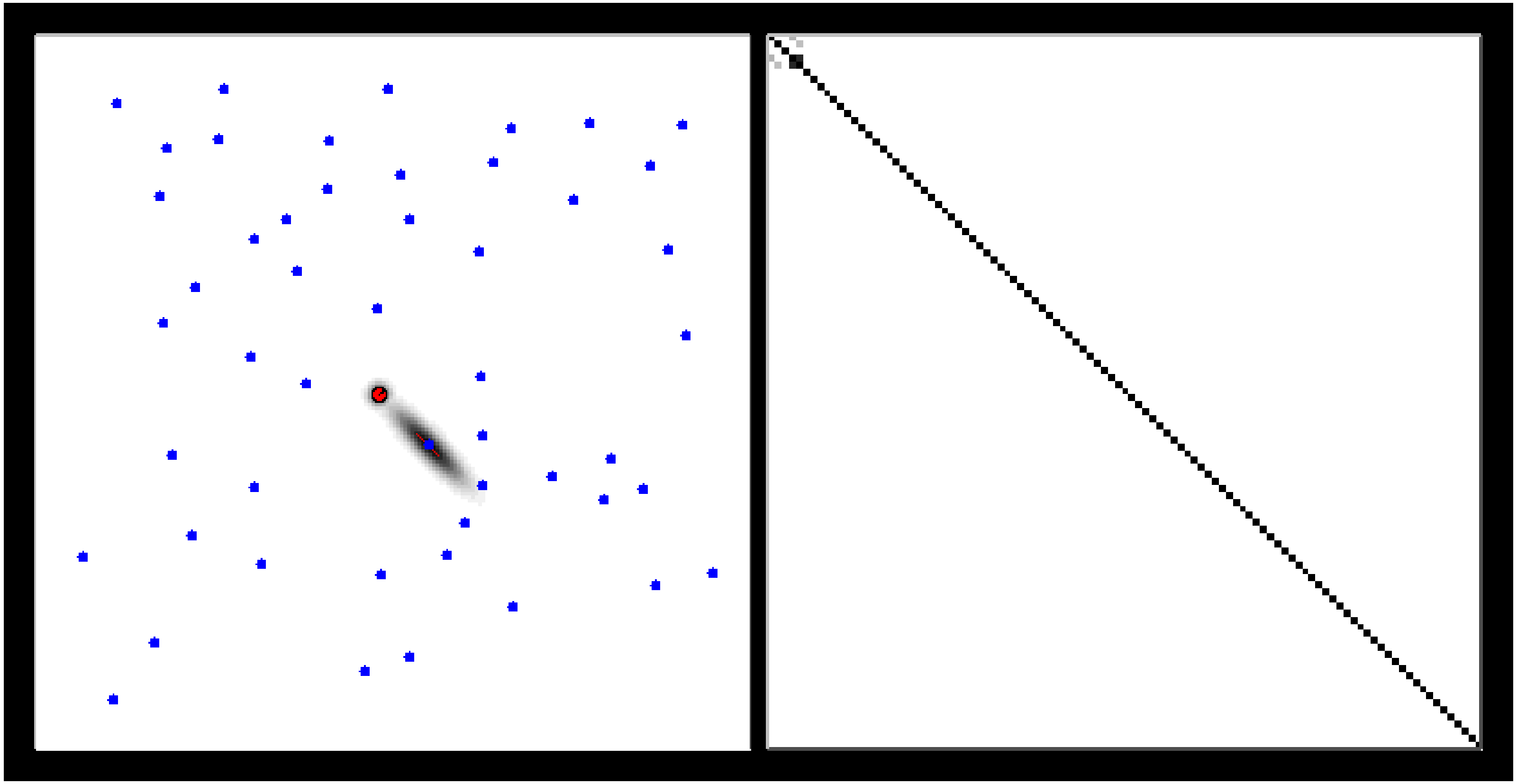
$$\boldsymbol{\Sigma}_t^+ = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)\, \boldsymbol{\Sigma}_t^-$$

- Since all (not newly initialized) landmarks are correlated with pose, the landmark correlations with the measured landmark get updated

- Hence, all state variables become correlated: The state covariance is dense!

- Measurement information propagates on all landmarks along the trajectory

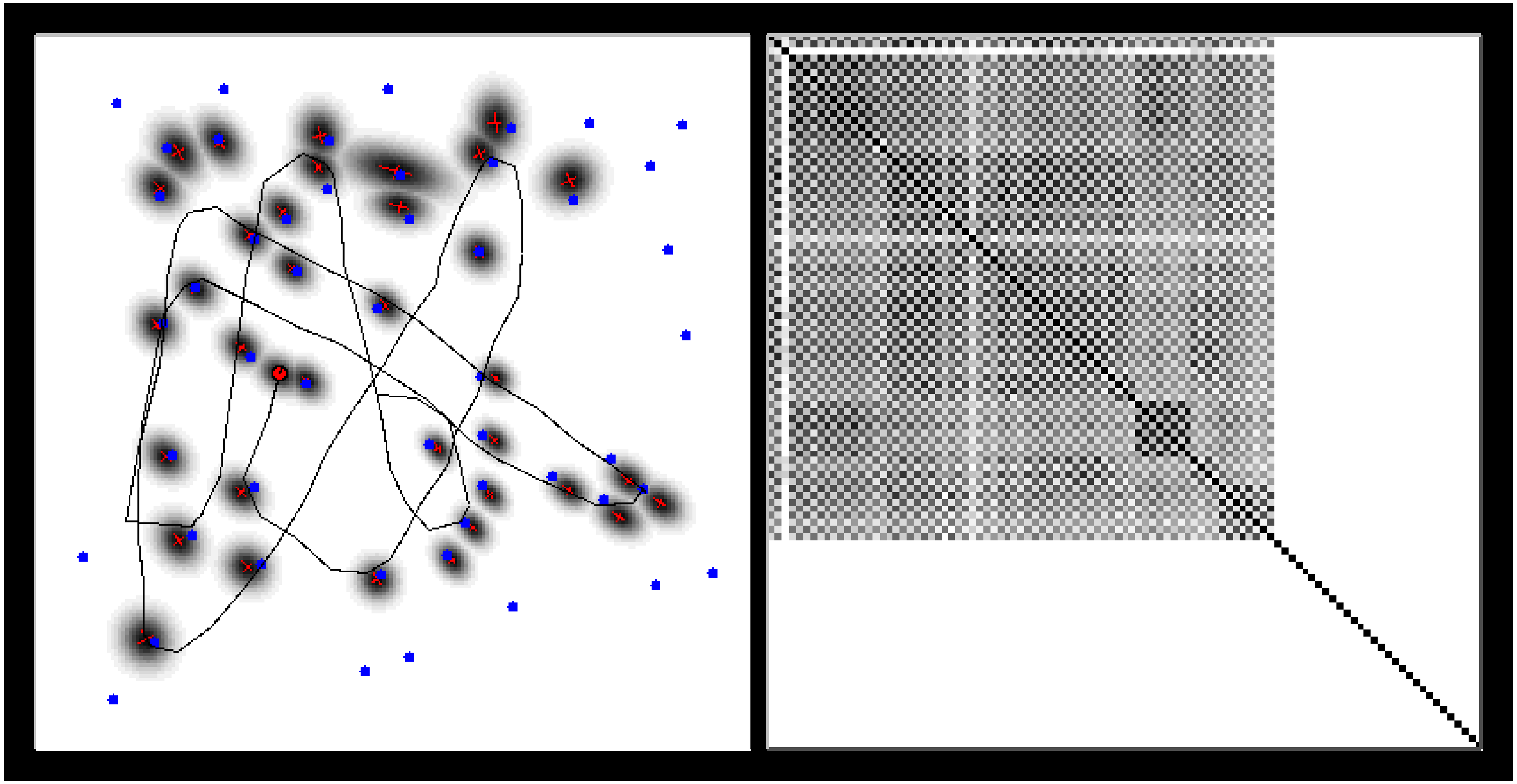
Example Evolution of the Covariance



Pose and map

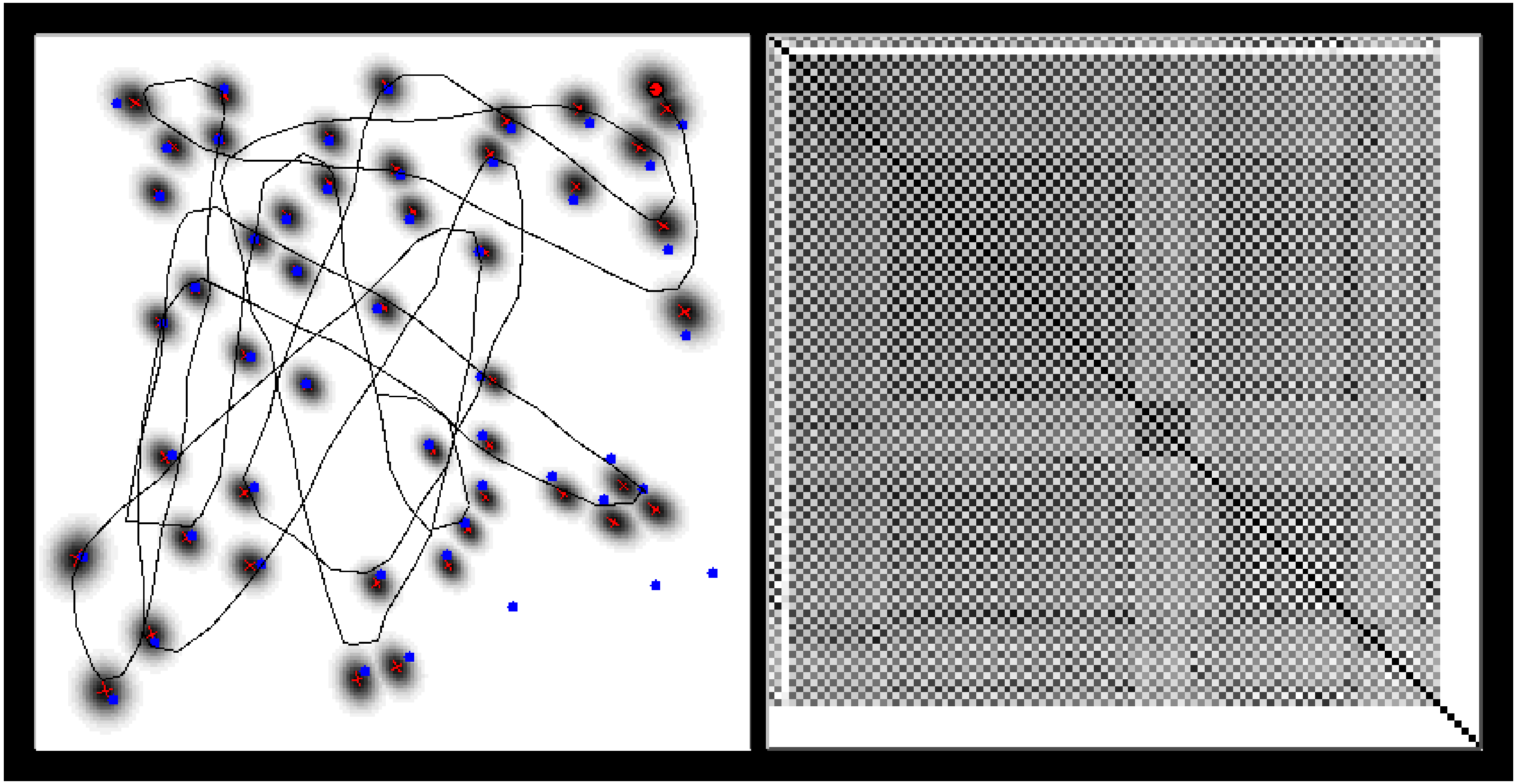
Correlation matrix

Example Evolution of the Covariance



Pose and map Correlation matrix

Example Evolution of the Covariance



Pose and map   Correlation matrix

Closing a Loop

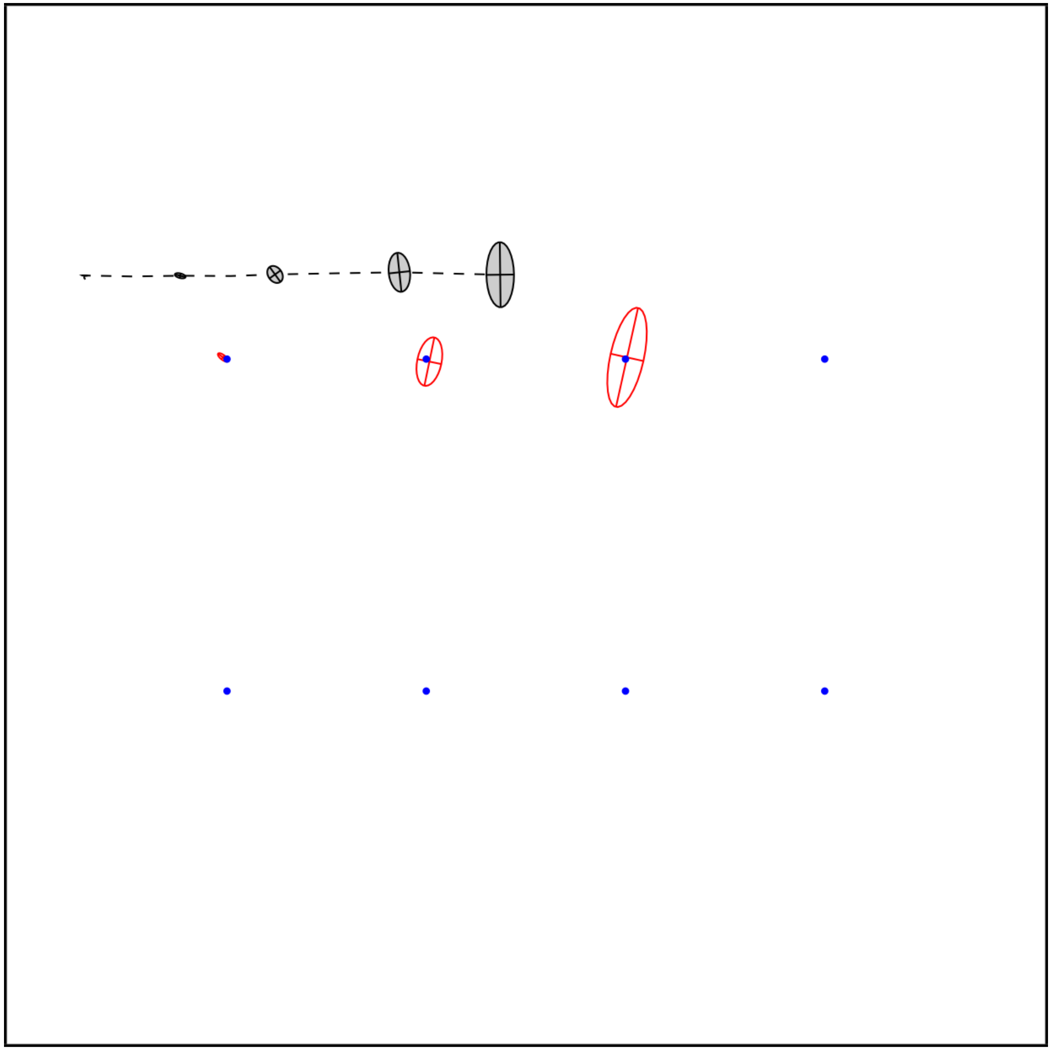


Image: Michael Montemerlo

Closing a Loop

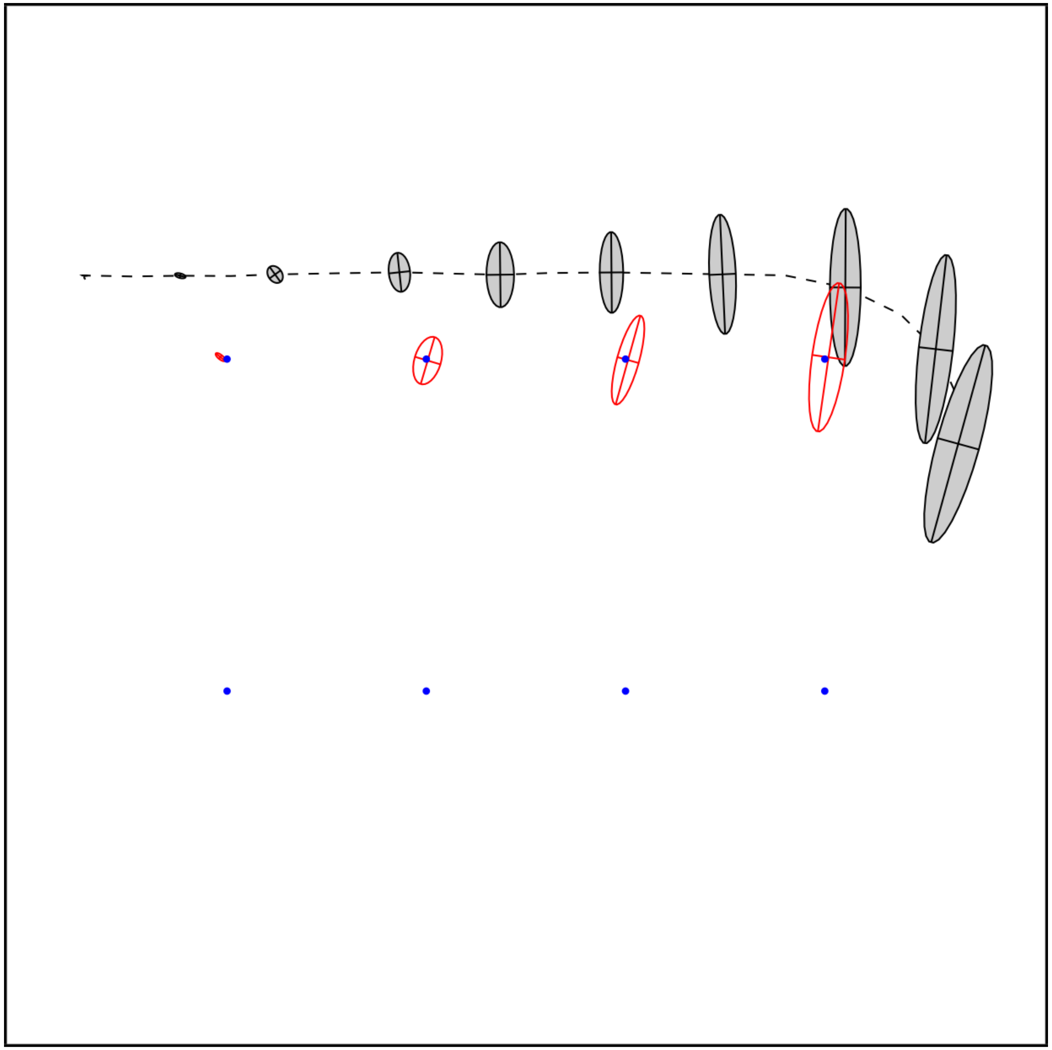


Image: Michael Montemerlo

Closing a Loop

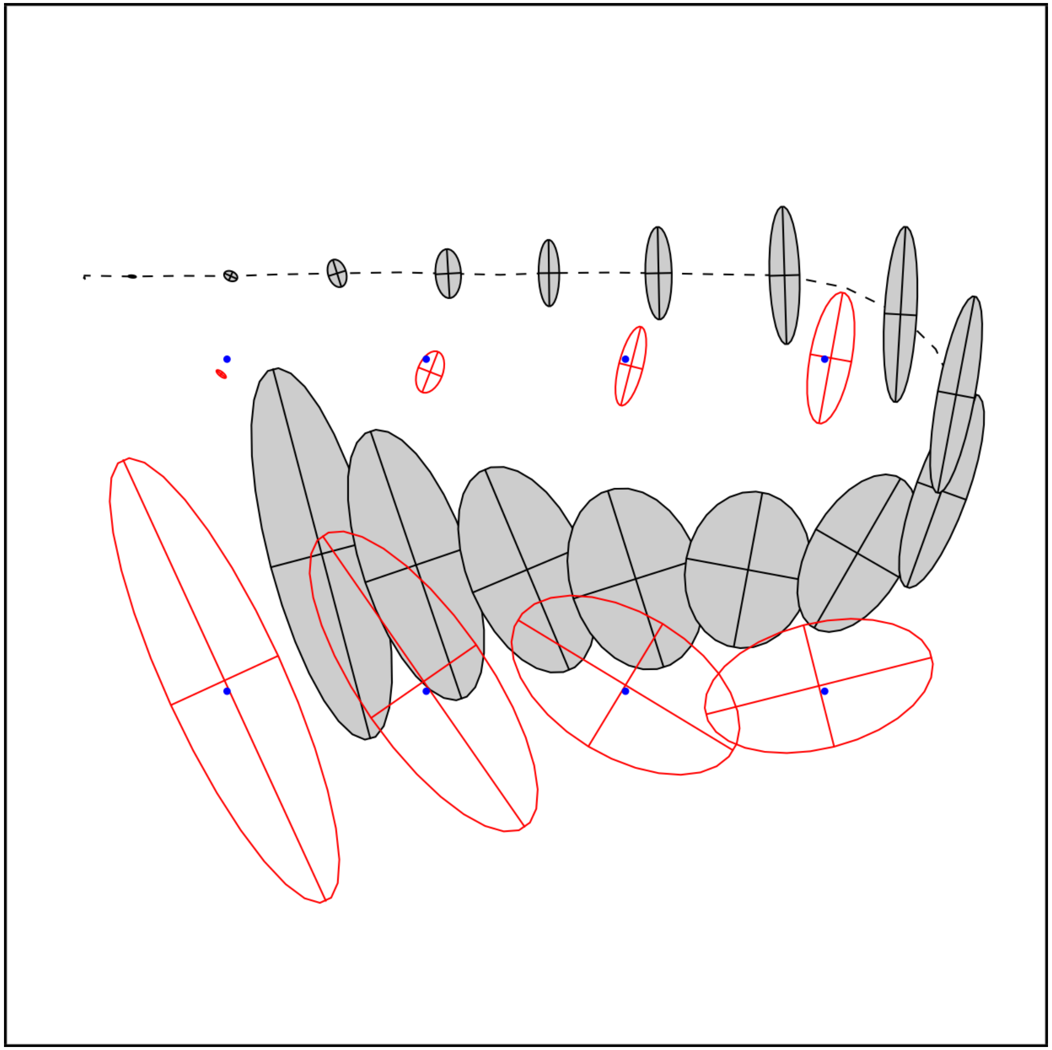


Image: Michael Montemerlo

Closing a Loop

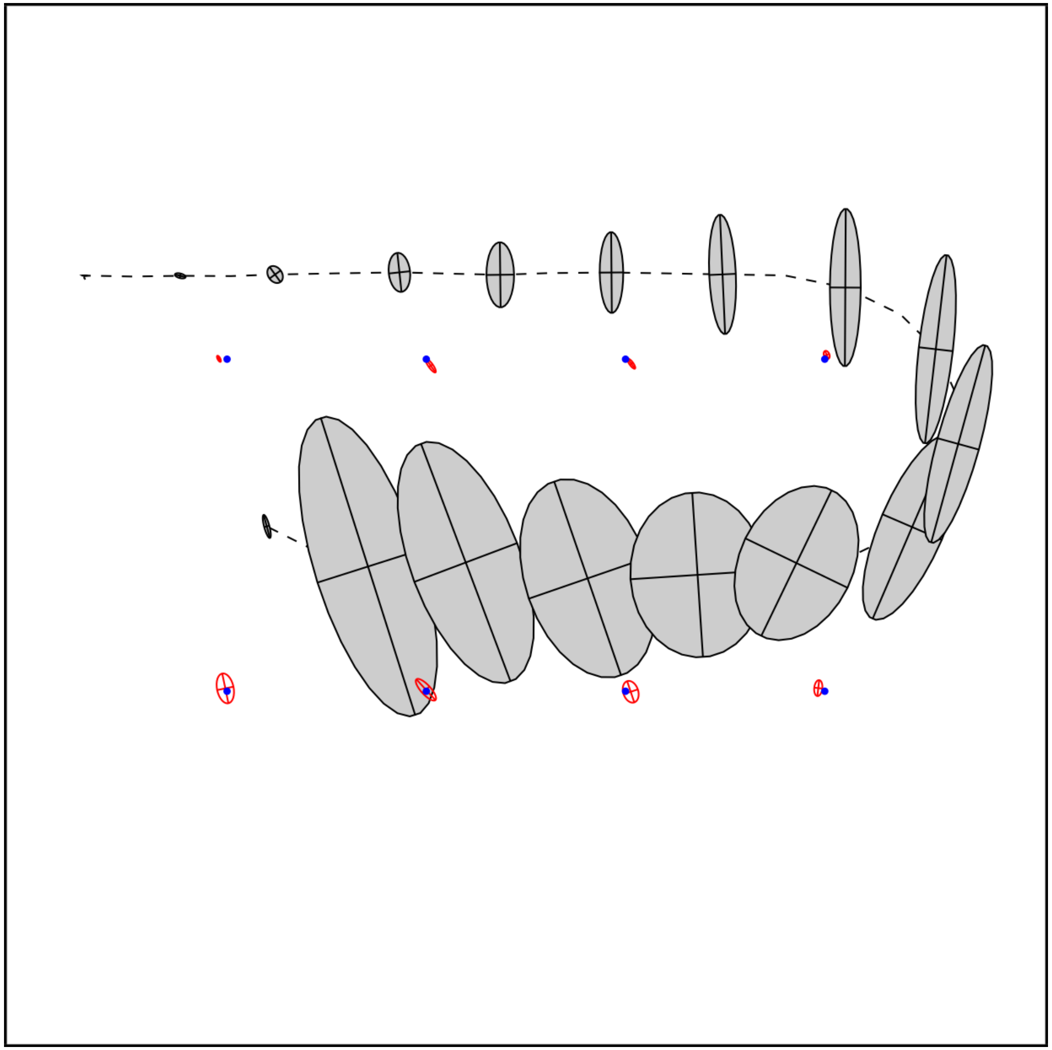


Image: Michael Montemerlo

Closing a Loop

- On loop closure, old landmarks in the map get reobserved
- Strong correlations are added between older parts of the map that were not observed for some time and the current pose / recently observed landmarks
- Pose and all landmarks are corrected to make the estimate more consistent with the reobservation
- Loop closure reduces uncertainty in pose and landmark estimates
  - High certainty in the old part of the map propagates to current pose and recent landmark estimates
  - But: Wrong correspondences can lead to divergence towards a wrong estimate!

MonoSLAM: Monocular EKF-SLAM

(Davison et al. PAMI, 2007)

https://www.youtube.com/watch?v=mimAWVm-0qA

State Parametrization

- Camera motion

$$\boldsymbol{\xi}_t = \begin{pmatrix} \mathbf{p}_t \\ \mathbf{q}_t \\ \mathbf{v}_t \\ \boldsymbol{\omega}_t \end{pmatrix}$$

3D position in world frame

Quaternion for rotation from camera to world frame

Linear velocity in world frame

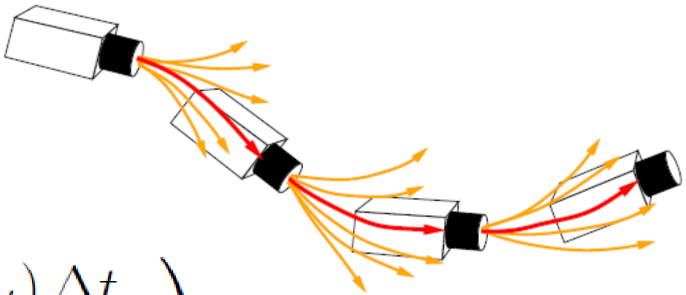
Angular velocity in world frame

- Landmarks

$$\mathbf{m}_{t,j} = \begin{pmatrix} m_{t,j,x} \\ m_{t,j,y} \\ m_{t,j,z} \end{pmatrix}$$

3D position in world frame

State-Transition Model



- 6-DoF camera dynamics model (constant-velocity)

$$\boldsymbol{\xi}_t = g_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{t-1}) = \begin{pmatrix} \mathbf{p}_{t-1} + (\mathbf{v}_{t-1} + \boldsymbol{\epsilon}_{\mathbf{v},t})\,\Delta t \\ \mathbf{q}_{t-1}\mathbf{q}\left((\boldsymbol{\omega}_{t-1} + \boldsymbol{\epsilon}_{\boldsymbol{\omega},t})\,\Delta t\right) \\ \mathbf{v}_{t-1} + \boldsymbol{\epsilon}_{\mathbf{v},t} \\ \boldsymbol{\omega}_{t-1} + \boldsymbol{\epsilon}_{\boldsymbol{\omega},t} \end{pmatrix}$$

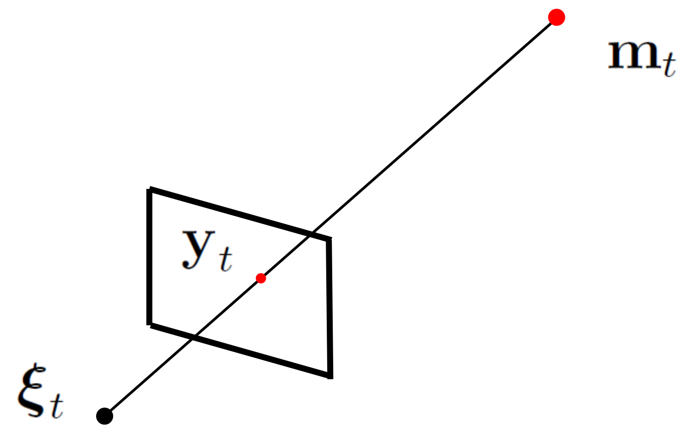
Gaussian noise

- Map remains static, $\mathbf{m}_t = g_{\mathbf{m}}(\mathbf{m}_{t-1}) = \mathbf{m}_{t-1}$

Images: Davison et al, 2007

Bearing-Only Observation Model

- Depth is not measured in a monocular image



- Landmark observation model

$$\overline{\mathbf{y}}_t = h(\boldsymbol{\xi}_t, \mathbf{m}_{t,c_t}) + \boldsymbol{\delta}_t = \mathbf{C}\pi\left(\mathbf{R}(\mathbf{q}_t)^\top (\mathbf{m}_{t,c_t} - \mathbf{p}_t)\right) + \boldsymbol{\delta}_t \qquad \boldsymbol{\delta}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{m_t}\right)$$

- MonoSLAM additionally considers the radial distortion in a wide-angle camera image using an analytically invertible model

Data Association

- Active search: likely region of measurement from innovation covariance

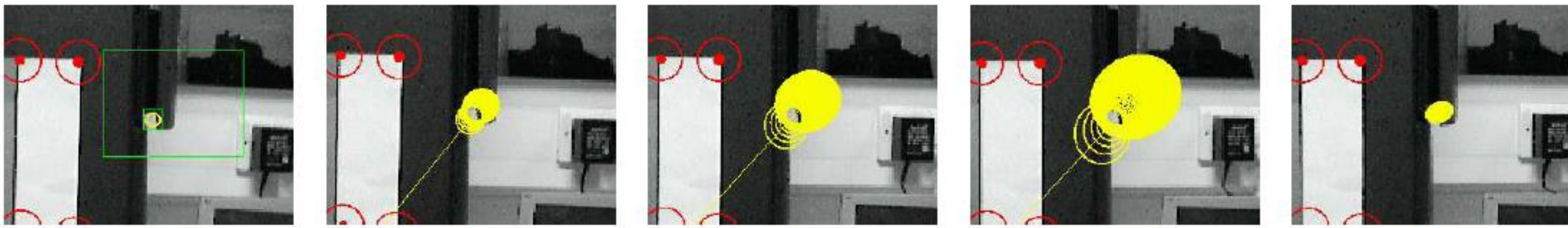
$$\mathbf{H}_t \mathbf{\Sigma}_t^- \mathbf{H}_t^\top + \mathbf{\Sigma}_{m_t}$$

- Correspondence measure
  - Matching of small image patches (f.e. 9x9 to 15x15)
  - Projective warping using a patch normal estimate
  - Sum of squared intensity differences

Images: Davison et al, 2007

Map Maintanance

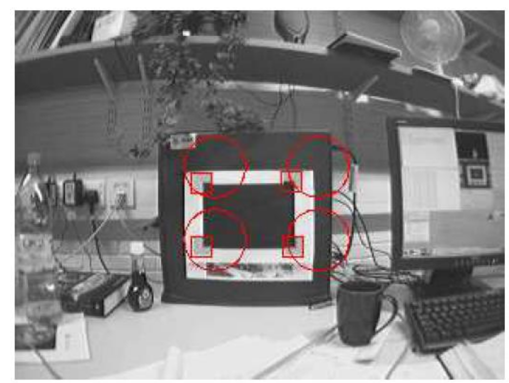
- Heuristics to keep number of visible landmarks from any camera view point small (~12 landmarks)



0.000s 0.033s 0.066s 0.100s 0.133s

- Depth initialization for new landmark using multiple hypothesis

- Map initialized with landmarks on a known 3D pattern
  - Sets metric scale
  - Good initial state for tracking
  - Stable pose for adding new landmarks



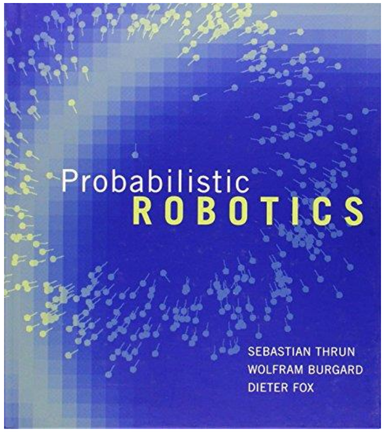
Images: Davison et al, 2007

Lessons Learned

- Online SLAM methods marginalize out past trajectory
- Extended Kalman Filters can be used for online SLAM
  - Maintains correlations between camera pose and all landmarks
  - Quadratic update run-time complexity limits map size
- MonoSLAM:
  - Implements Visual EKF-SLAM for monocular cameras
  - Data association via active search and patch correlation

Further Reading

- Probabilistic Robotics textbook



Probabilistic Robotics, S. Thrun, W. Burgard, D. Fox, MIT Press, 2005

- A.J. Davison et al., MonoSLAM: Real-Time Single Camera SLAM. IEEE Transaction on Patterm Analysis and Machine Intelligence, 2007

Thanks for your attention!

Slides Information

- These slides have been initially created by Jörg Stückler as part of the lecture “Robotic 3D Vision” in winter term 2017/18 at Technical University of Munich.
- The slides have been revised by myself (Niclas Zeller) for the same lecture held in winter term 2020/21
- Acknowledgement of all people that contributed images or video material has been tried (please kindly inform me if such an acknowledgement is missing so it can be added).