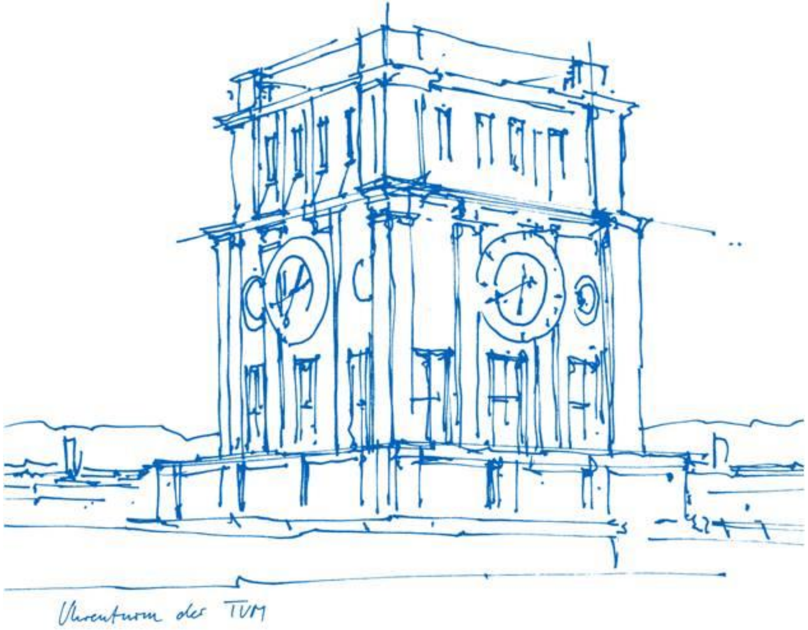
Probabilistic Data Association for Semantic SLAM

Damian Bogunowicz

Robotics, Cognition Intelligence

Technical University of Munich

Munich, 17. January 2018



Probabilistic Data Association for Semantic SLAM

Authors: Sean L.Bowman, Nikolay Atanasov, Kostas Daniilidis, George J. Pappas

Further reading:
A Unifying View of Geometry, Semantics, and Data Association in SLAM (2018)

Outline

- SLAM: A short recap
- Data association – fundamental problem
- Semantic SLAM
- Experiments
- Personal remarks
- Conclusions

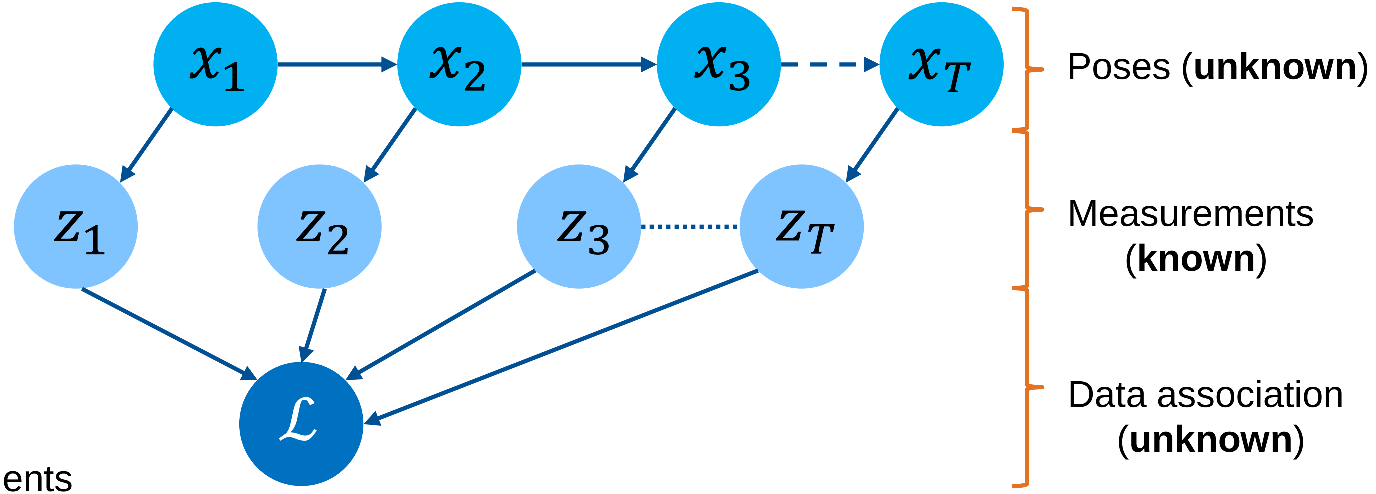
SLAM: A short recap

Chicken and egg problem...

Estimate the current position of an agent without a map

Build the map of the unknown environment, even though we do not know our current position

SLAM: A short recap



Given:
$\mathcal{Z}$ – set of sensor measurements

Estimate:
$\mathcal{X}$ – sequence of poses
$\mathcal{L}$ – static landmark positions

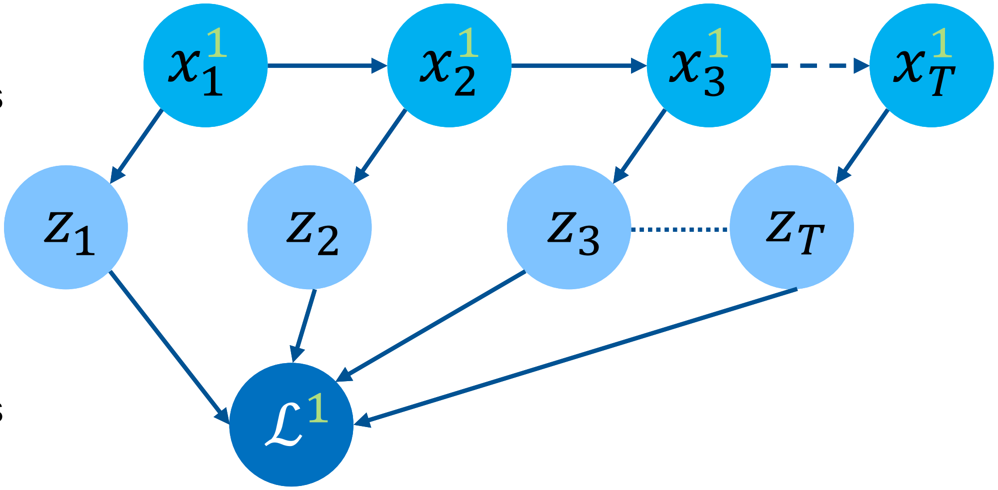
SLAM: A short recap

First step: find the best data association given current measurements and some prior estimates of pose and landmarks

$$\mathcal{D}^{i+1} = arg \max_{\mathcal{D}} p(\mathcal{D}|\ \mathcal{X}^{\mathrm{i}}, \mathcal{L}^{i}, \mathcal{Z})$$

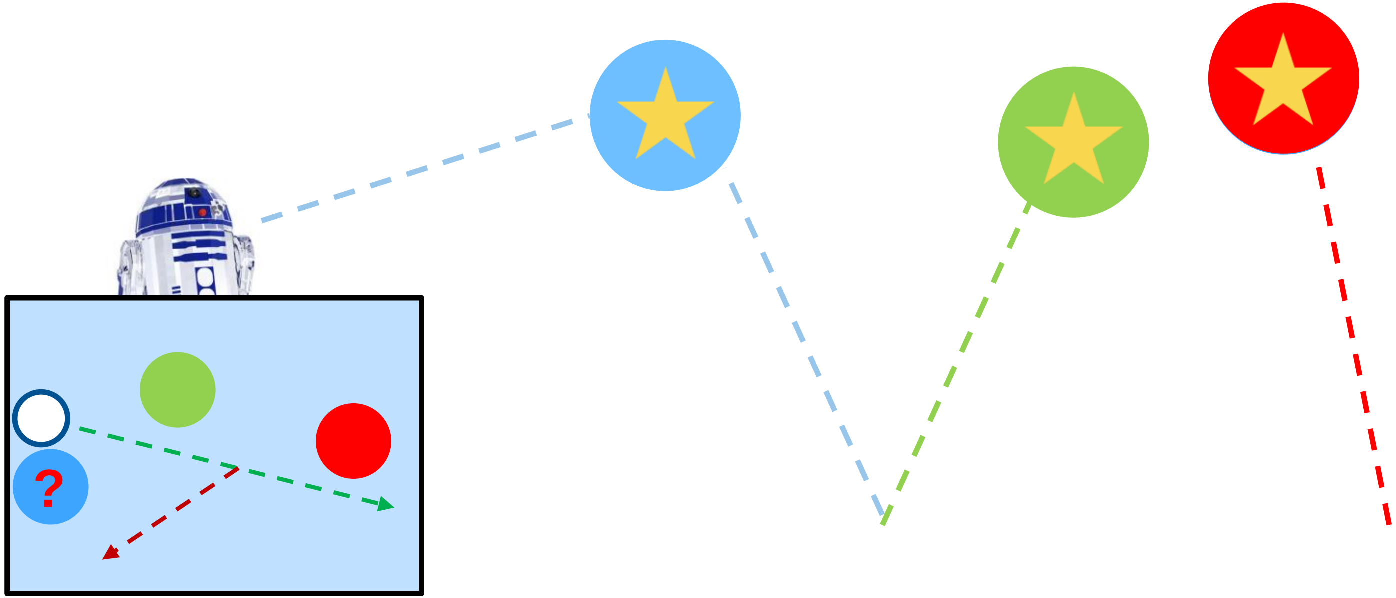
Second step: for the current measurements estimate the most likely landmark and sensor states

$$\mathcal{X}^{i+1}, \mathcal{L}^{i+1} = arg \max_{\mathcal{X},\mathcal{L}} p(\mathcal{Z}|\ \mathcal{X}, \mathcal{L}, \mathcal{D}^{i+1})$$



Data association – fundamental problem

Data association - association between the landmarks and sensor measurements

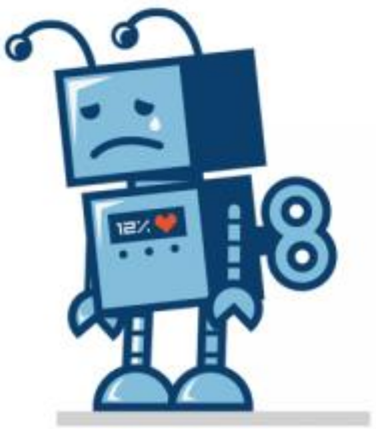


Bad!

False positive!

Data association – fundamental problem

Problem: Incorrectly chosen data association
has detrimental effect on estimation performance!



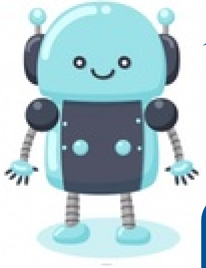
Data association – a simple idea...

Before:

$$\mathcal{X}^{i+1}, \mathcal{L}^{i+1} = arg \max_{\mathcal{X},\mathcal{L}} p(\mathcal{Z} | \mathcal{X}, \mathcal{L}, \mathcal{D}^{i+1})$$

$\mathcal{D}^{i+1}$ as the mode of $p(\mathcal{D} | \mathcal{X}^{i}, \mathcal{L}^{i}, \mathcal{Z})$ - hard decision

Is my measurement related to this landmark?

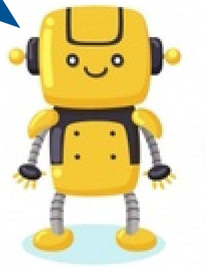


Now:

$$\mathcal{X}^{i+1}, \mathcal{L}^{i+1} = arg \max_{\mathcal{X},\mathcal{L}} \mathbb{E}_{\mathcal{D}}[\log p(\mathcal{Z} | \mathcal{X}, \mathcal{L}, \mathcal{D}) | \mathcal{X}^{i}, \mathcal{L}^{i}, \mathcal{Z}]$$

$$\mathcal{X}^{i+1}, \mathcal{L}^{i+1} = arg \max_{\mathcal{X},\mathcal{L}} \sum_{k=1}^{K} \sum_{j=1}^{M} w_{kj}^{i} \log p(\boldsymbol{z}_k | \boldsymbol{x}_{\alpha_k}, l_j)$$

How sure am I that my measurement is related to that landmark?



Consider entire density of $\mathcal{D}$ – distribution of belief in landmark, given the pose.

Semantic SLAM - landmarks

Before: a landmark is being described using simple, low-level features.

Now: each landmark is described by its state, which constists of its position and class label.

Position $l^p \epsilon \mathbb{R}^3$

Class label l^c, where $\mathbb{C} = \{1, \dots, C\}$

Semantic SLAM - observations

Inertial Information *(IMU + monocular camera)*	Geometric Information *(Keyframes)*	Semantic Information *(Keyframes)*
• **Keyframes** • **Sensor state** (6-D Pose, velocity, IMU bias values)	• **ORB features**	• **Object detections** (DPM) from every keyframe image.

Semantic SLAM - measurements

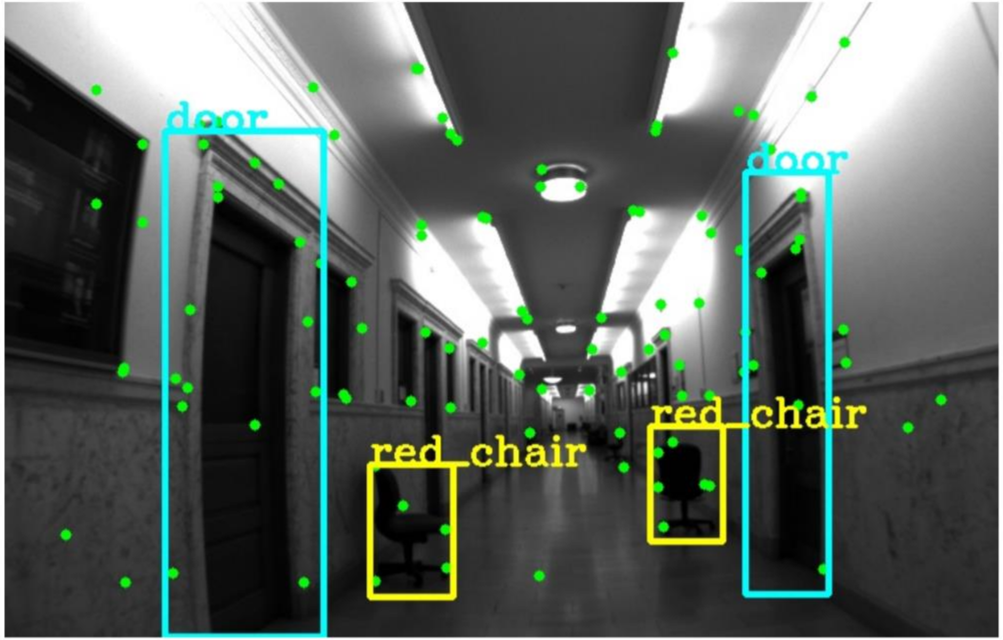
Object detection

$s_k = (s_k^c, s_k^s, s_k^b)$

s_k^c - detected class

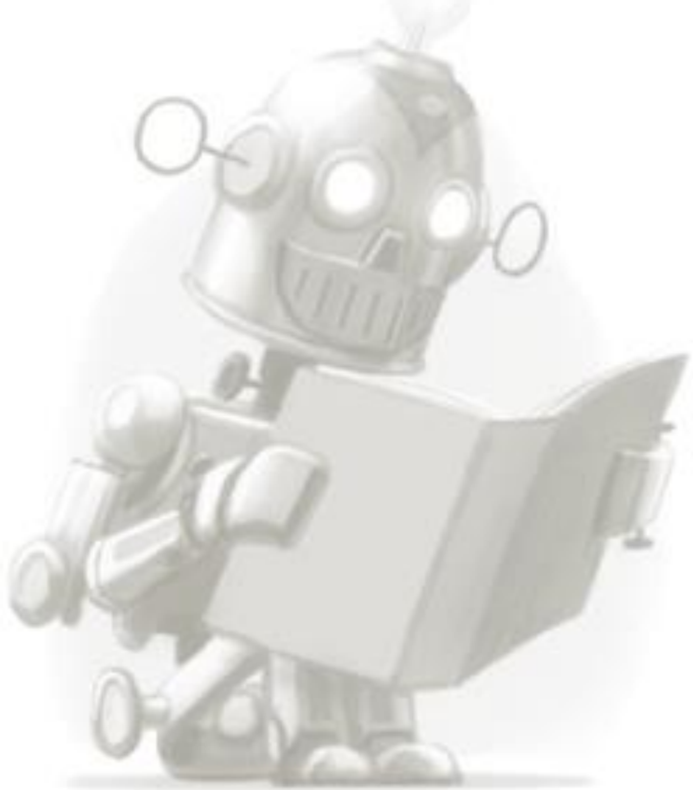
s_k^s - detection confidence

s_k^b - bounding box



Example keyframe image overlaid with ORB features and object detections

Semantic SLAM – EM algorithm



EM – Expectation maximization

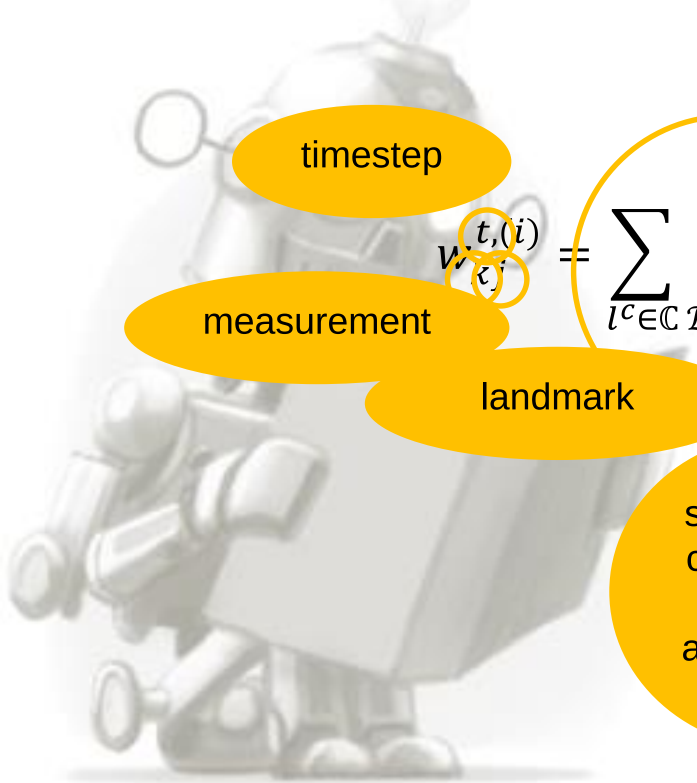
„E” step

Estimate the data association **distribution** (not likelihood) $p(\mathcal{D}|\, \mathcal{X}^{\mathrm{i}}, \mathcal{L}^{i}, \mathcal{Z})$ in form of the weights w_{kj}^{i}

„M” step

Maximize the expected measurement log likelihood over the previously computed distribution.

Semantic SLAM – Expectation step



$$w_{k,j}^{t,(i)} = \sum_{l^c \in \mathbb{C}} \sum_{\mathcal{D}_t \in \mathbb{D}_t(k,j)} \frac{p(S_t \mid \mathcal{X}^{(i)}, \mathcal{L}^{(i)}, \mathcal{D}_t)}{\sum_{l^c} \sum_{\mathcal{D}_t \in \mathbb{D}_t} p(S_t \mid \mathcal{X}^{(i)}, \mathcal{L}^{(i)}, \mathcal{D}_t)}$$

timestep

measurement

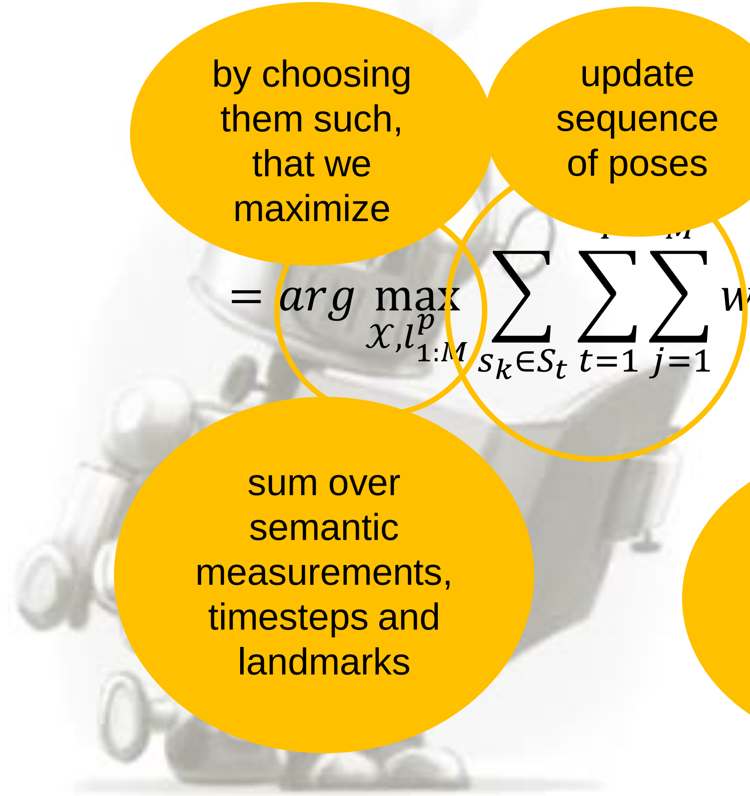
landmark

sum over all classes and all data associations

object detection

normalization over classes and data associations in given timestep

Semantic SLAM – Maximization st



$$\mathcal{X}^{i+1}, l^{p,(i+1)}_{1:M} =$$

$$= \arg\max_{\mathcal{X}, l^{p}_{1:M}} \sum_{s_k \in S_t} \sum_{t=1}^{T} \sum_{j=1}^{M} w_{kj}^{t,(i)} \log p(\boldsymbol{s}_k | \boldsymbol{x}_t, l_j) + \log p(\mathcal{Y}|\mathcal{X}) + \log p(\mathcal{I}|\mathcal{X})$$

by choosing them such, that we maximize

update sequence of poses

update position of landmarks

sum over semantic measurements, timesteps and landmarks

probability of the current semantic measurement

probability of the current geometric measurement

probability of the current inerial measurement

Semantic SLAM – maximization step (semantic factors)

$$\arg\max_{\mathcal{X}, l^p_{1:M}} \sum_{s_k \in S_t} \sum_{t=1}^{T} \sum_{j=1}^{M} \boxed{w_{kj}^{t,(i)} \log p(\boldsymbol{s}_k | \boldsymbol{x}_t, l_j)} + \log p(\mathcal{Y}|\mathcal{X}) + \log p(\mathcal{I}|\mathcal{X})$$

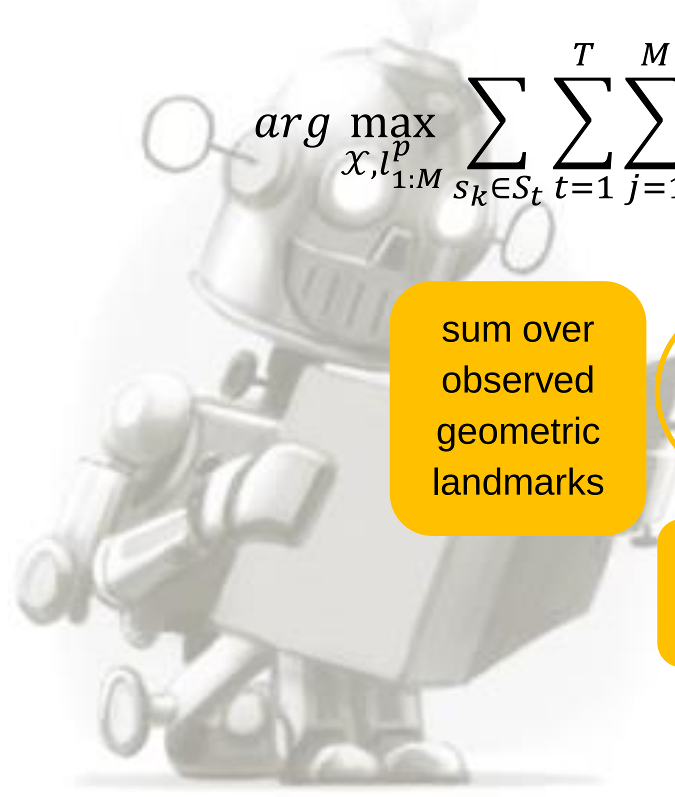
$$\left\| h_\pi\left(\boldsymbol{x}_t, l_j^p\right) - s_k^b \right\|^2_{\boldsymbol{R}_s / w_{kj}^{t,(i)}}$$

bounding box inferred from the measurement

projection of the visible landmark onto a camera pose

covariance of camera noise scaled by the weight

Semantic SLAM – maximization step (geometric factors)



$$\arg\max_{\mathcal{X}, l^p_{1:M}} \sum_{s_k\in S_t}\sum_{t=1}^{T}\sum_{j=1}^{M} w_{kj}^{t,(i)} \log p(\boldsymbol{s}_k|\boldsymbol{x}_t, l_j) + \log p(\mathcal{Y}|\mathcal{X}) + \log p(\mathcal{I}|\mathcal{X})$$

geometric measurement

sum over observed geometric landmarks

$$\sum_{i=1}^{N_y}\sum_{k:B_k^y=i} \left\| h_\pi\left(\boldsymbol{x}_{\alpha_k^y}, \rho_i\right) - \boldsymbol{y}_k\right\|^2_{\boldsymbol{R}_y}$$

covariance

sum over measurements

3D position of ORB features corresponding to the landmark that generated the measurement

Semantic SLAM – maximization step (inertial factors)

$$\underset{\mathcal{X},\, l^{p}_{1:M}}{arg\,\max} \sum_{s_k \in S_t} \sum_{t=1}^{T} \sum_{j=1}^{M} w_{kj}^{t,(i)} \log p\left(\boldsymbol{s}_k \middle| \boldsymbol{x}_t, l_j\right) + \log p(\mathcal{Y}|\mathcal{X}) + \log p(\mathcal{I}|\mathcal{X})$$

$$-\left\| \boldsymbol{r}_{\mathcal{I}_{ij}} \right\|^2_{\Sigma_{ij}}$$

rotation, velocity and position differences between consecutive keyframes

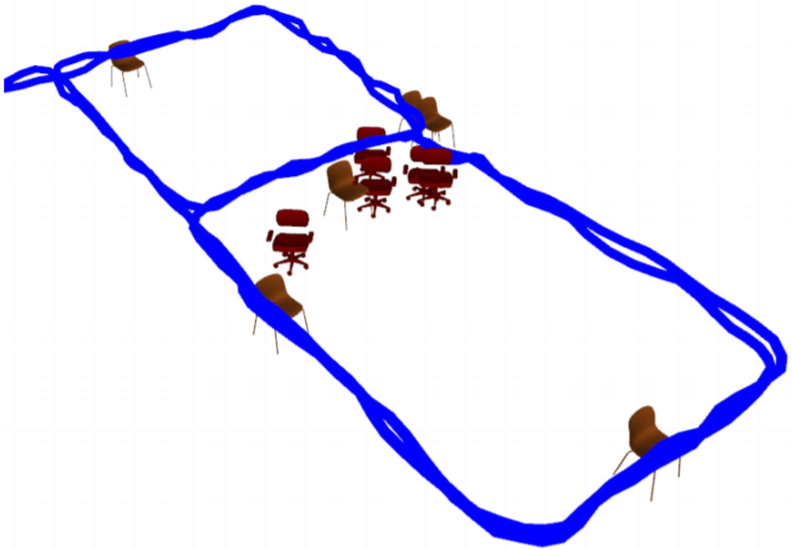
noise covariance

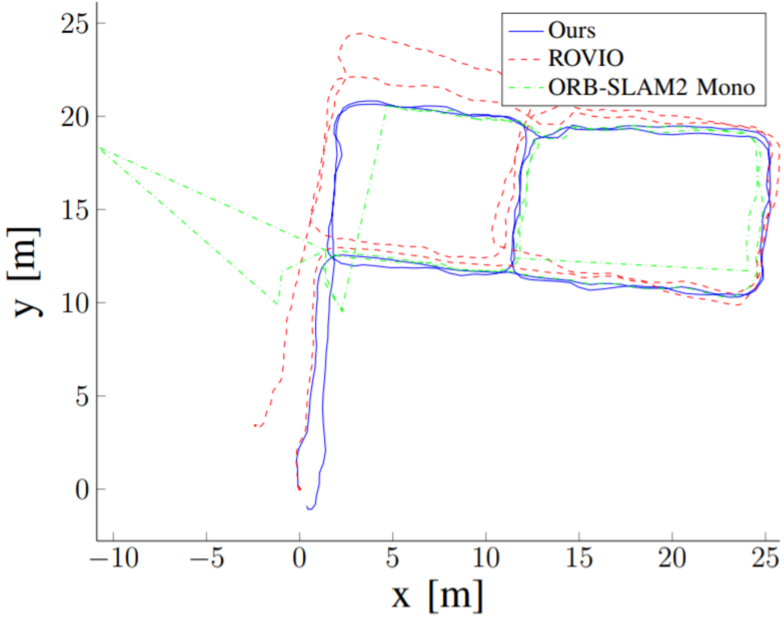
Experiments– key points

- Real-time computation
- Naive selection of keyframes
- Algorithm tested in three experiments
- Interesting initialization of new landmarks

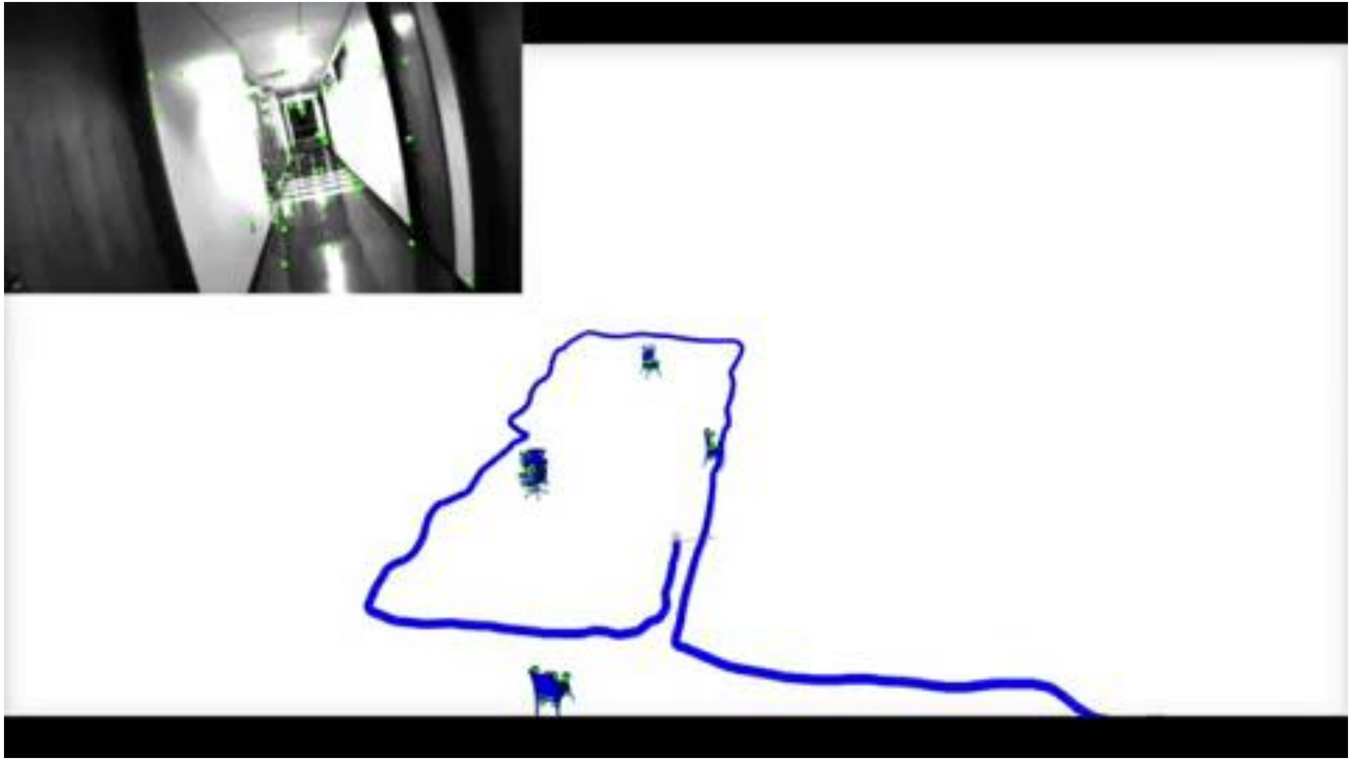
Experiments

Experiment 1: 175m run around the office floor



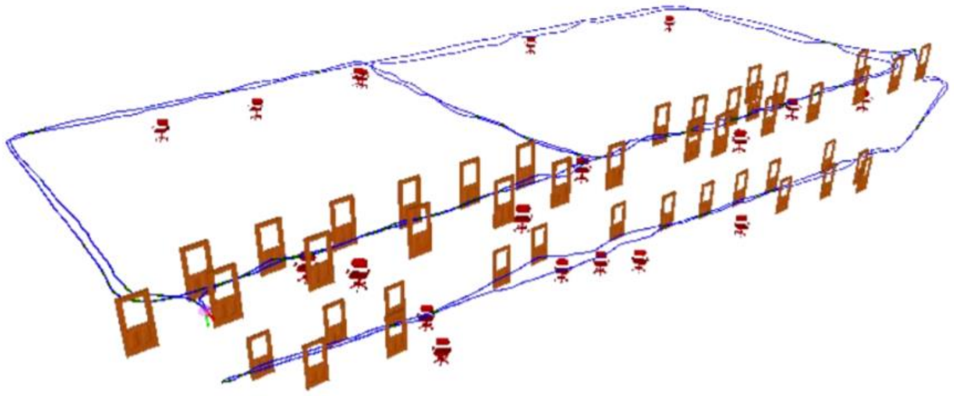


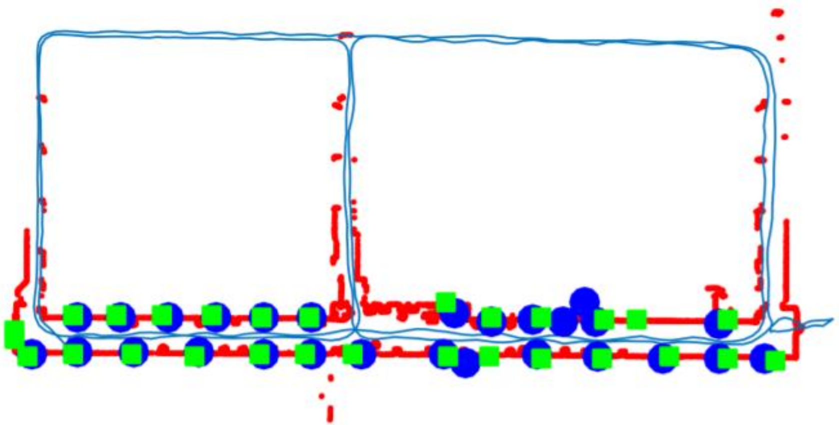
Experiments



Experiments

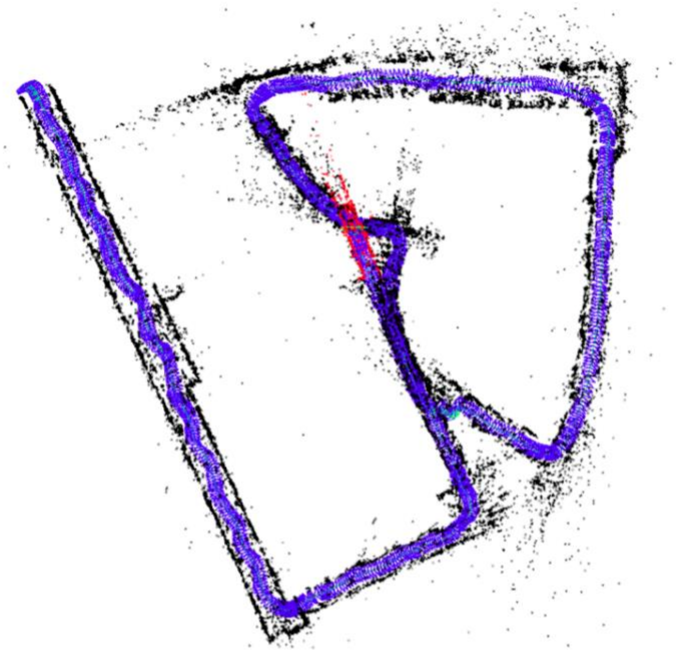
Experiment 2: 625m run around two different offices





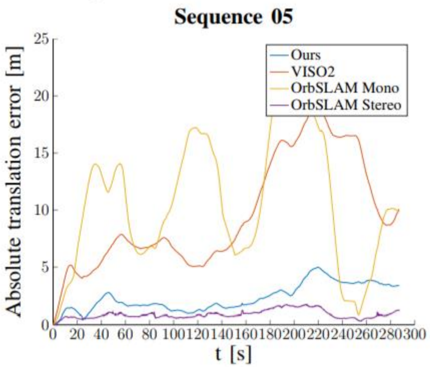
Experiments

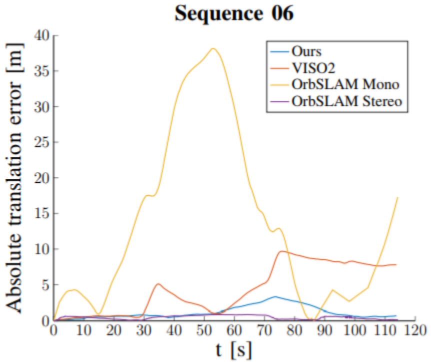
Experiment 2: 625m run around two different office floor



Experiments

Experiment 3: KITTI outdoor dataset (sequences 05 and 06)



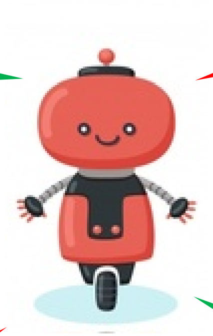


Experiments

Personal remarks

Very thorough (sometimes hairy) mathematical description!

No code available online.



Experiments could have been documented better

Simple (albeit brilliant) idea

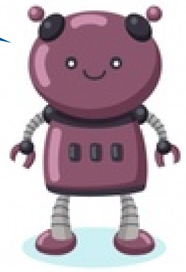
Conclusion

Take home message!

Semantic features improve localization performance and ability to close loops

Semantic information great in aiding autonomous operation of robots

Room for improvement!



Estimate full pose of the semantic objects

Consider systems with multiple sensors and non-stationary objects