Towards Continuous Learned Semantic Representation through a Viewpoint-Dependent Observation Model

Yuri Feldman
Department of Computer Science,
Technion, Israel Institute of Technology

Vadim Indelman
Department of Aerospace Engineering,
Technion, Israel Institute of Technology

THE ISRAELI SMART TRANSPORTATION RESEARCH CENTER

ANPL Autonomous Navigation and Perception Lab

TASP TECHNION AUTONOMOUS SYSTEMS PROGRAM

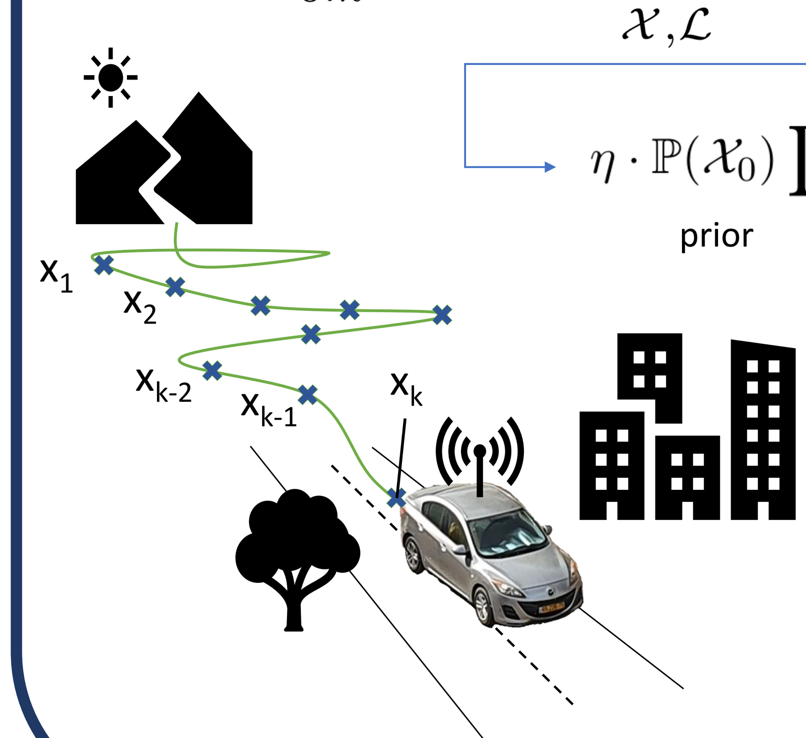
1.a Background - Simultaneous Localization and Mapping (SLAM)

- SLAM is commonly formulated as joint max a-posteriori over agent poses $\mathcal{X}_{0:k}$ and landmarks $\mathcal{L}$ given history of observations and user controls.

$$\mathcal{X}^*_{0:k},\mathcal{L}^* = \arg\max_{\mathcal{X},\mathcal{L}} \mathbb{P}(\mathcal{X}_{0:k},\mathcal{L}\mid\mathcal{H}_k) \qquad \{\mathcal{Z}_{0:k},\mathcal{U}_{0:k-1}\}$$

observations user controls

$$\eta\cdot\underbrace{\mathbb{P}(\mathcal{X}_0)}_{\text{prior}}\prod_i \underbrace{\mathbb{P}(\mathcal{X}_{i+1}\mid\mathcal{U}_i,\mathcal{X}_i)}_{\text{motion model}}\prod_{j\in\mathcal{M}_i}\underbrace{\mathbb{P}(\mathcal{Z}_j\mid\mathcal{L}_j,\mathcal{X}_i)}_{\text{observation model}}$$



© https://morioh.com/p/bd1b6fcgdgeb

See e.g. [1,2].

1.b Background – Object-Level SLAM

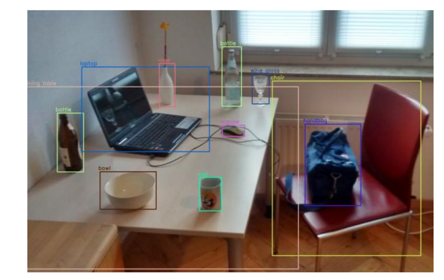
- In *Object SLAM* (e.g. [3-8]) mapping is done on the level of objects.

robot track | object categories (discrete!) | measurement and control history

$$\arg\max_{\mathcal{X},\mathcal{C},\mathcal{O}} \mathbb{P}(\mathcal{X}_{0:k},\mathcal{C},\mathcal{O}\mid\mathcal{H}_k)$$

object geometry

$$=\arg\max_{\mathcal{X},\mathcal{C},\mathcal{O}} \underbrace{\mathbb{P}(\mathcal{X}_{0:k},\mathcal{O}\mid\mathcal{C},\mathcal{H}_k)}_{\text{Continuous hypothesis}}\cdot\underbrace{\mathbb{P}(\mathcal{C}\mid\mathcal{H}_k)}_{\text{Hypothesis weight}}$$



© Wikipedia

$$\eta\cdot\mathbb{P}(\mathcal{X}_0)\prod_i \underbrace{\mathbb{P}(\mathcal{X}_{i+1}\mid\mathcal{U}_i,\mathcal{X}_i)}_{\text{motion model}}\prod_{j\in\mathcal{M}_i}\underbrace{\mathbb{P}(\mathcal{Z}_j\mid\mathcal{O}_j,\mathcal{X}_i,\mathcal{C}_j)}_{\text{observation model}}$$

(one per class!)

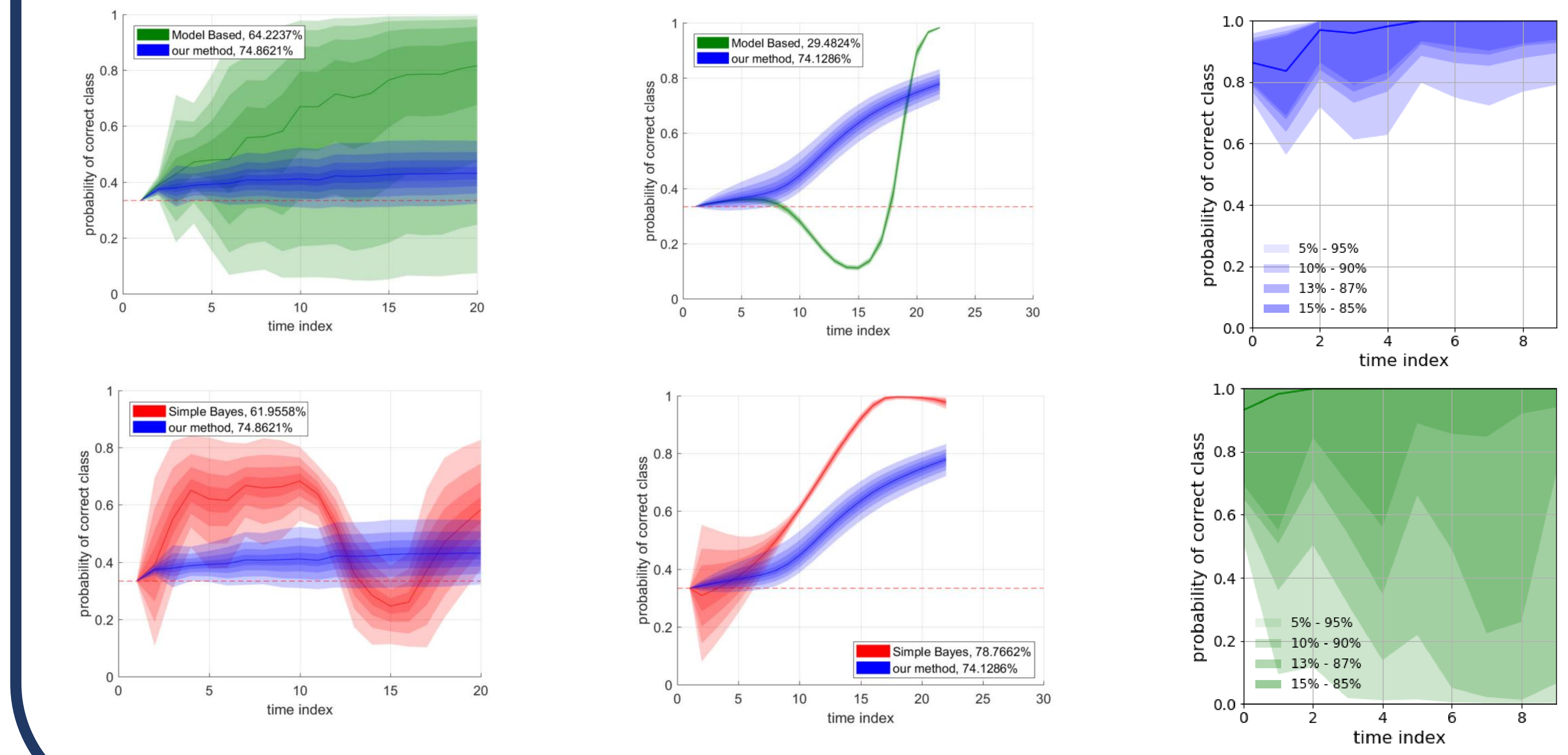
Drawbacks:

- Per-class models required
- Mixed inference, exponential number of hypotheses:

$$c\in\{1,\dots,N\} \qquad |\{(c_1,\dots,c_m)\}| = N^m$$

A. Bayesian Viewpoint-Dependent Classification [9]

Model uncertainty - simulation | Localization uncertainty - simulation | Localization uncertainty - Active Vision Dataset



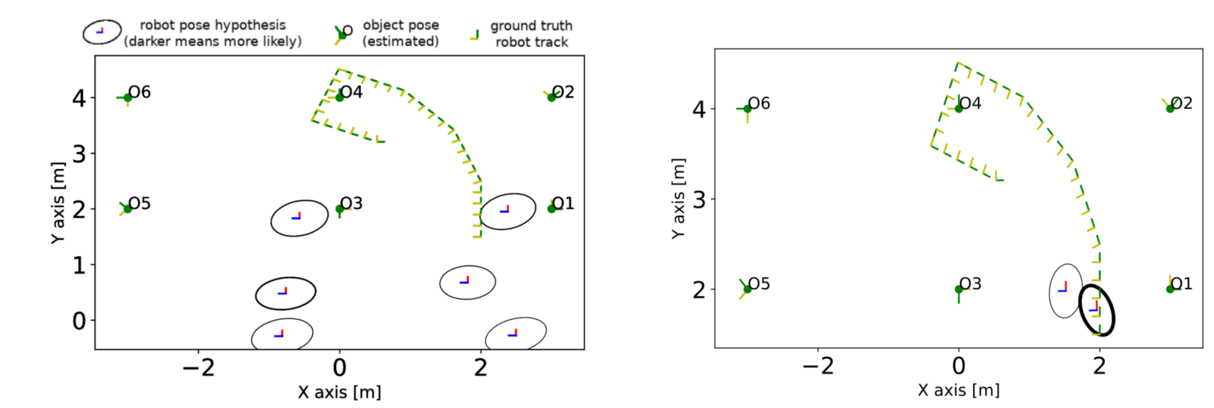
B. Data-Association Aware Semantic Mapping [10]

data association

$$\mathbb{P}(\mathcal{X}_k,C,\beta_{1:k}\mid\mathcal{H}_k)=\underbrace{\mathbb{P}(\mathcal{X}_k\mid C,\beta_{1:k},\mathcal{H}_k)}_{b^C_{\beta_{1:k}}[\mathcal{X}_k]}\underbrace{\mathbb{P}(C,\beta_{1:k}\mid\mathcal{H}_k)}_{w^C_{\beta_{1:k}}}$$

(continuous) hypothesis | hypothesis weight

- Viewpoint-dependent model modulates belief propagation, aids disambiguation: $b^C_{\beta_{1:k}}[\mathcal{X}_k]\propto b^C_{\beta_{1:k-1}}[\mathcal{X}_{k-1}]\mathbb{P}(x_k\mid x_{k-1},a_{k-1})\mathbb{P}(\mathcal{Z}_k\mid\mathcal{X}_k,C,\beta_k)$

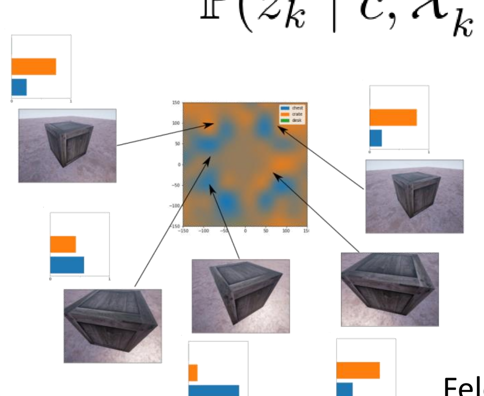


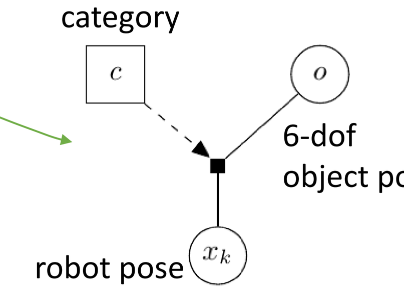
2.a Viewpoint-Dependent Semantic Models

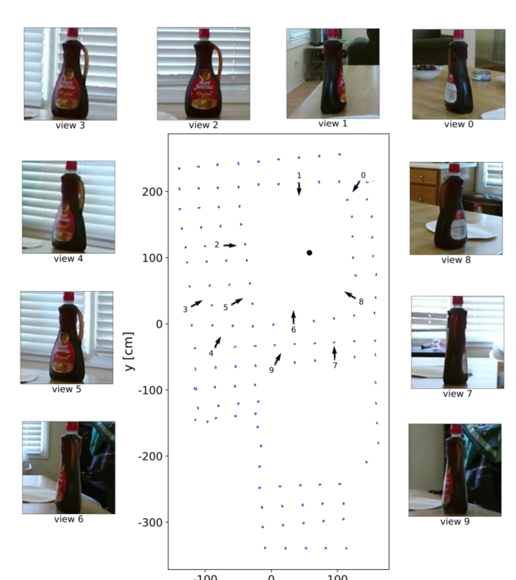
- Viewpoint-dependent models couple inference of geometry and semantics, making them mutually beneficial (see [9-11]).

current observation | object class | viewpoint relative to object

$$\mathbb{P}(z_k\mid c,\mathcal{X}^{(rel)}_k)$$



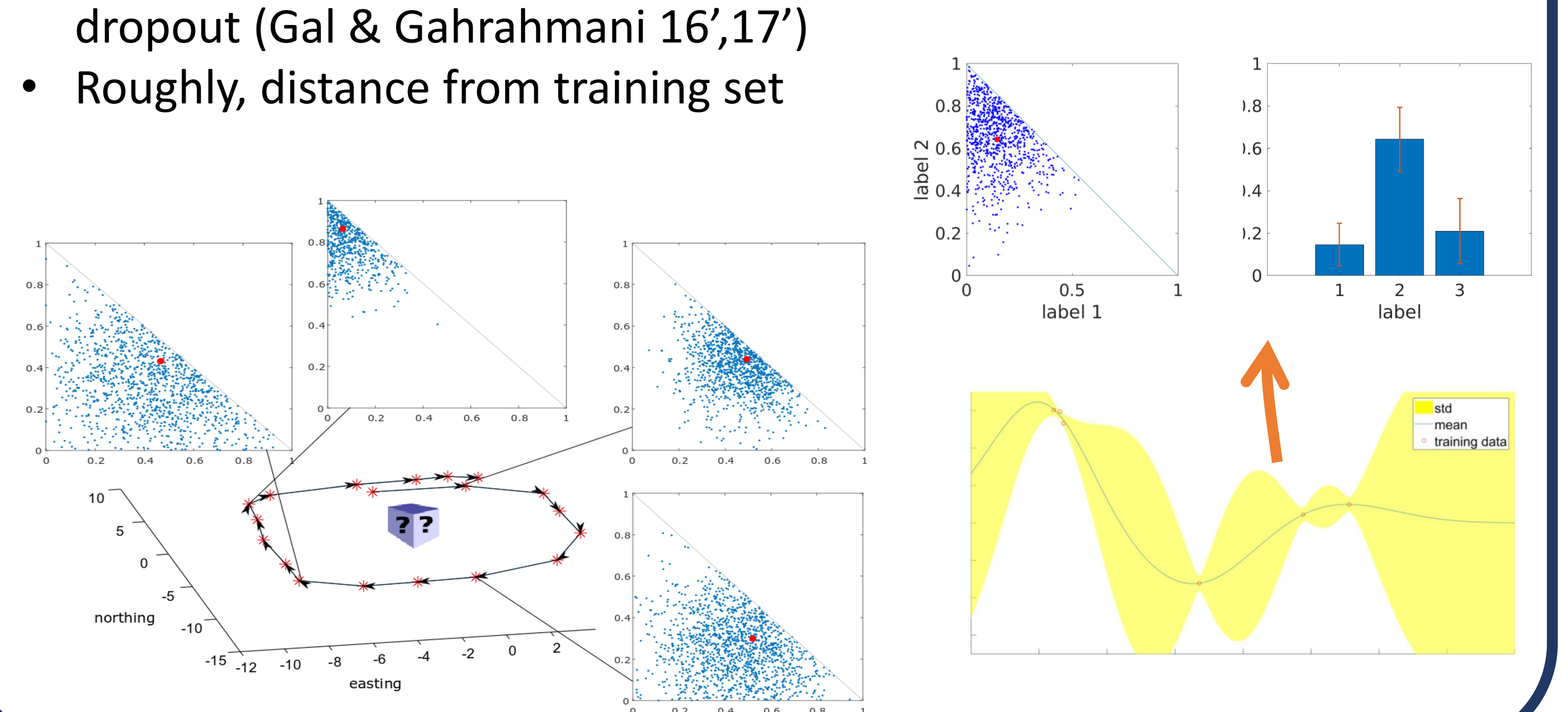




Feldman and Indelman. "Spatially-dependent Bayesian semantic perception under model and localization uncertainty." Autonomous Robots (2020): 1-29.

2.b Model Uncertainty

- Approximate posterior $\mathbb{P}(s\mid z)$ using multiple forward passes with dropout (Gal & Gahrahmani 16',17')
- Roughly, distance from training set



2.c Spatial Class Model

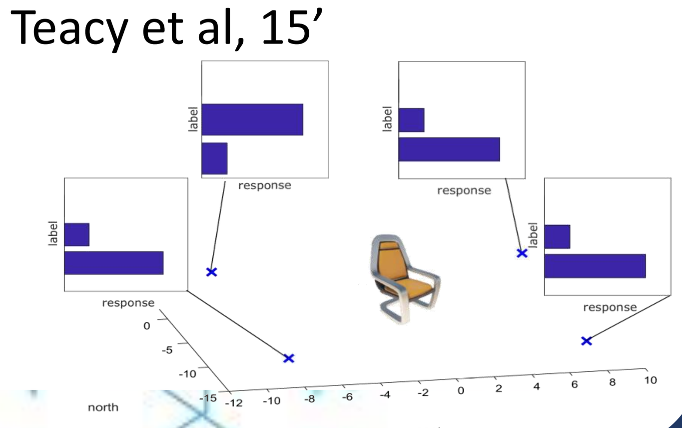
For every known class^c, model similar to Teacy et al, 15'

$$s=f_c(x^{(rel)})+\epsilon$$

$$f_c(x^{(rel)})\sim\mathcal{GP}\left(\mu_c(x^{(rel)}),k_c(\cdot,\cdot)\right)$$

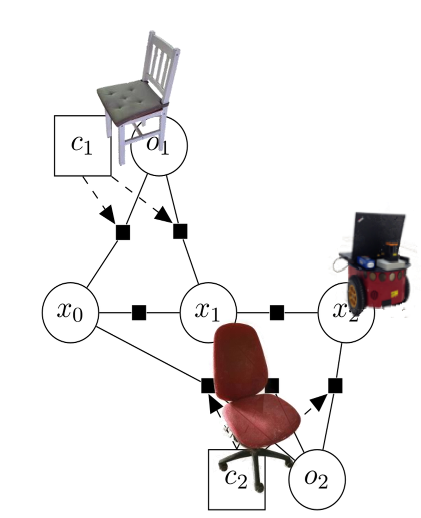
Inducing joint distribution

$$\mathbb{P}(\mathcal{S}_{0:k}\mid c,\mathcal{X}^{(rel)}_{0:k})=N\left(\mu_{0:k},\Sigma_k\right)$$



C. Continuous Semantic Representation

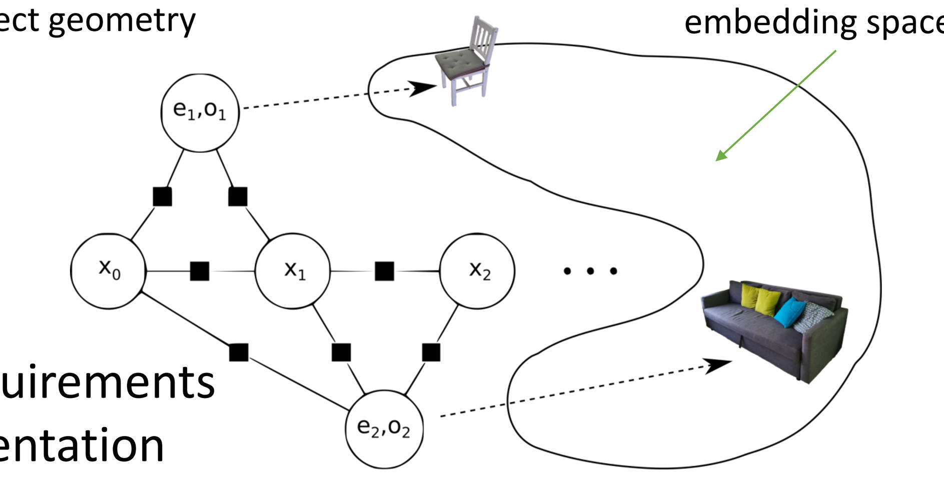
- Idea: learn a continuous semantic representation



robot track | object categories (discrete) | measurement and control history | continuous representation

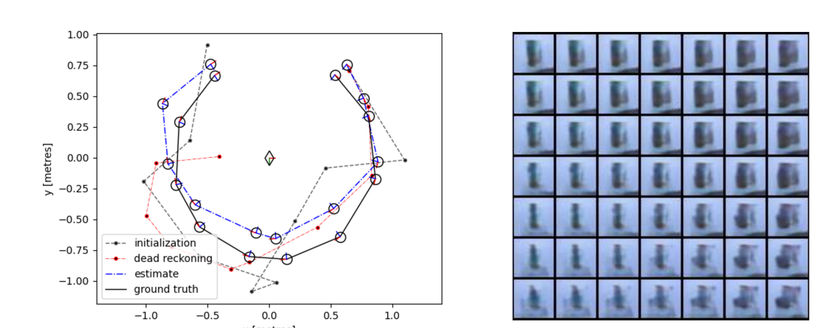
$$\mathbb{P}(\mathcal{X}_{0:k},\mathcal{C},\mathcal{O}\mid\mathcal{H}_k) \Longrightarrow \mathbb{P}(\mathcal{X}_{0:k},\mathcal{E},\mathcal{O}\mid\mathcal{H}_k)$$

object geometry | embedding space



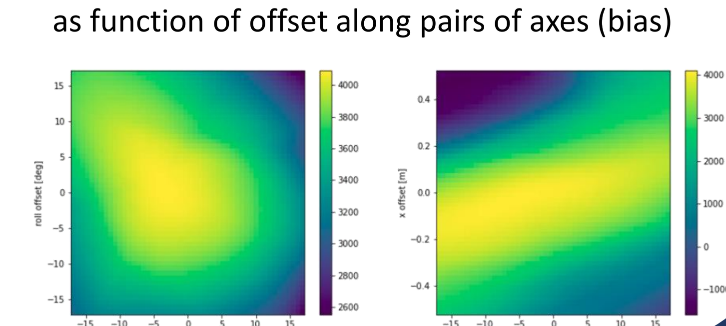
Advantages:

- Continuous inference
- Relaxed ground-truth requirements
- Learned semantic representation



Ground truth and estimated tracks | predictions for varying X

Learned likelihood as function of offset along pairs of axes (bias)



Bibliography

[1] Kaess et al. 08' TRO (iSAM)
[2] Kaess et al. 12' IJRR (iSAM2)
[3] Salas-Moreno et al. 13' CVPR
[4] Choudhary et al. 14' IROS
[5] Bowman et al. 17' ICRA
[6] McCormack et al. 18 3dv
[7] Nicholson et al. 19' RAL
[8] Yang and Scherer 19' TRO
[9] Feldman and Indelman 20' ARJ
[10] Tchuiev et al. 19' IROS
[11] Kopitkov and Indelman 18' IROS
[12] Teacy et al. 15' AAMAS

DEPARTMENT OF AEROSPACE ENGINEERING | TECHNION Israel Institute of Technology

TECHNION | The Henry and Marilyn Taub Faculty of Computer Science