GraphNav

A Behavioral Approach to Visual Navigation with Graph Localization Networks

Paul Hölzen

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Technical Aspects of Multimodal Systems

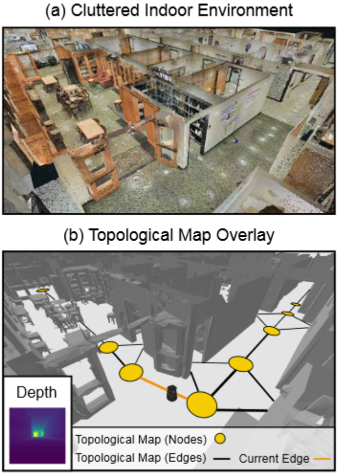
20. January 2020

Outline

1. Motivation
2. Method
   - Graph Neural Network
   - Graph Localization Network
   - Particle Filtering for GLN
   - Behavior Networks
3. Results
   - Baselines
   - Evaluation
4. Conclusion
5. References

Motivation

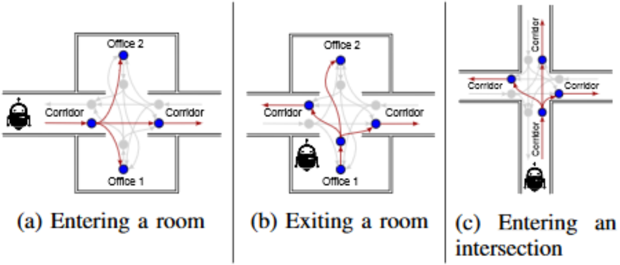
- Navigating cluttered spaces is difficult for robots
- Humans are really good at it
- Behavioral approach founded in psychology
  - proposed by Chen et al. [1]
  - *Cognitive Maps* $\rightarrow$ graph-like structure



Environment and corresponding topological map [1]

Motivation

- Benefits of a graph-like map
  - Coarse/Sparse topological information
  - Navigation planning on a graph
  - High-level abstraction



Navigation examples on topological map [1]

Graph Neural Network (GNN)

- Neural Network performing on graph-like structures
- Captures relational inductive biases
- Graph $G = (\mathbf{u}, V, E)$
  - $\mathbf{u}$ global feature
  - $V = \{\mathbf{v}_i\}_{i=1:n}$ node features
  - $E = \{(\mathbf{e}_k, r_k, s_k)\}_{k=1:m}$ edge features
- Edge features correspond to behaviors
  - *corridor follow*
  - *find door*
  - *turn left*
  - *turn right*
  - *straight (into room)*

Graph Neural Network (GNN)

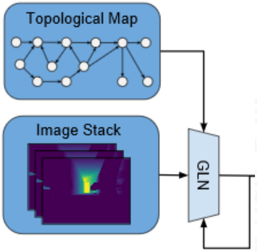
- Graph network blocks
  1. $\phi^e(\cdot)$ update edge features
  2. $\rho^{e \rightarrow v}(\cdot)$ aggregate edge features
  3. $\phi^v(\cdot)$ update node features
  4. $\rho^{e \rightarrow u}(\cdot), \rho^{v \rightarrow u}(\cdot)$ aggregate edge and node features
  5. $\phi^u(\cdot)$ update global feature

Info

The update functions $\phi^e(\cdot), \phi^v(\cdot), \phi^u(\cdot)$ were implemented using multilayer perceptrons (MLPs), the aggregation functions $\rho^{e \rightarrow v}(\cdot), \rho^{e \rightarrow u}(\cdot), \rho^{v \rightarrow u}(\cdot)$ use elementwise summation to ensure symmetry of the function (permutation agnostic)

Graph Localization Network (GLN)

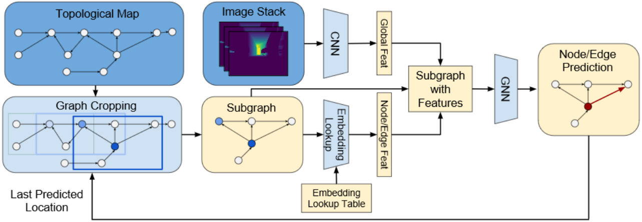
- Predicts location of the agent in the topological map
- Inputs
  - Current visual observation
  - Last predicted location
  - Graph with edge and node features



Graph localization network overview [1]

Graph Localization Network (GLN)

- Topological map is cropped to region around last location
- Edge/Node features from embedding lookup table
- Global feature from CNN processing visual observation
- GNN predicts the current node/edge



GLN architecture in detail [1]

Particle Filtering for GLN

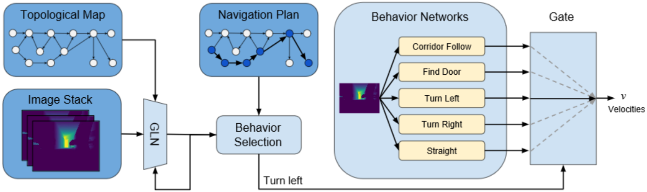
- used to improve GLN predictions
- based on statistical model
- $p(x_t | u_t, x_{t-1})$
  - x_t current state at time step t
  - u_t control input
- $p(z_t | x_t)$
  - z_t observation/measurement at time step t

Particle Filtering for GLN

- Assumption 1: Two time steps $t-1$ and t don't differ a lot in topological location
  - $p(x_t|u_t, x_{t-1}) = p(x_t|x_{t-1})$
  - Chen et al. use $p(x_t = x_{t-1}|x_{t-1}) = 0.8$
- Assumption 2: $p(z_t)$ and $p(x_t)$ are uniform distributions for all time steps
  - $\gamma = \frac{p(z_t)}{p(x_t)} = \text{const.}$
  - Bayes rule: $p(z_t|x_t) = \gamma \cdot p(x_t|z_t) \propto p(x_t|z_t)$
  - Approximate $p(x_t|z_t)$ by aggregating edge probabilities from the GLN

Behavior Networks

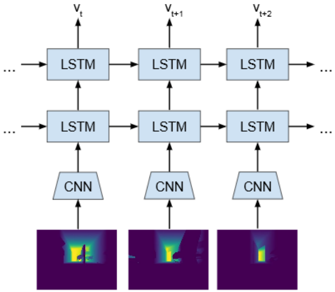
- Separate networks for each behavior
- Correspond to edge features
- CNNs and LSTMs used to implement



Overall architecture of *GraphNav* including behavior networks [1]

Behavior Networks

- CNN-based behavior networks
  - *corridor follow*
  - *find door*
- LSTM-based behavior networks
  - *turn left*
  - *turn right*
  - *straight (into room)*



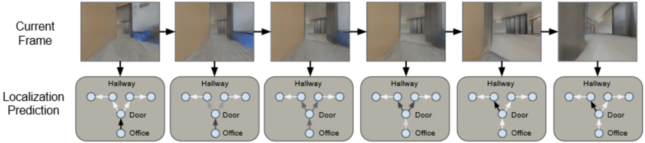
Architecture of LSTM-based behavior networks [1]

Baselines

- Evaluation of the results by comparing to baselines
  - *PhaseNet*[2]: LSTM-based, predicts temporal progress of behavior and when to switch to a new one
  - *BehavRNN*[3]: Sequence-to-sequence deep learning model, behavior classification from visual input
  - *GTL*: Ground Truth Localization, used to evaluate behavior networks independently

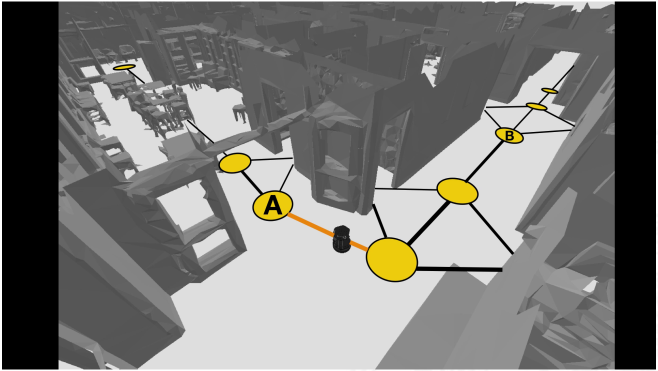
Evaluation

- *GraphNavPF* (with Particle Filtering) has highest performance compared to baselines
- Per-behavior success (90%) and path completion rate (70%) are resonable
- *PhaseNet* and *BehavRNN* perform significantly worse on seen and unseen environments
- *GTL* baseline shows that behavior networks work well, struggles in open spaces



Output of the localization network [1]

Video



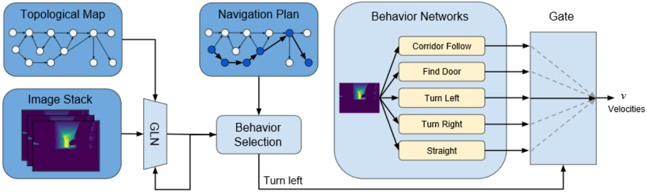
Video example of the GraphNav approach working [4]

Future Research

- Topological map has to be created and annotated by hand
- Set of behaviors has to be pre-defined
- Chen et al. propose data-driven approach to automate this
- Simulation-to-reality has to be tested

Conclusion

- Navigation approach that uses topological map and visual information as input
- Graph neural networks for localization
- Separate behavior networks with behavior selection
- Outperforms several baselines



Overall architecture of *GraphNav* including behavior networks [1]

References

[1] Chen, Kevin, et al. "A behavioral approach to visual navigation with graph localization networks." arXiv preprint arXiv:1903.00445 (2019).

[2] Yu, Tianhe, et al. "One-shot hierarchical imitation learning of compound visuomotor tasks." arXiv preprint arXiv:1810.11043 (2018).

[3] Sutskever, Ilya, et al. "Sequence to sequence learning with neural networks." In Advances in neural information processing systems, pages 3104–3112 (2014).

[4] Chen, Kevin, et al. "GraphNav: A behavioral approach to visual navigation with graph localization networks." March 2019, URL: www.youtube.com/watch?v=nN3B1F90CFM, Acessed 17.01.2020.