Semantic Visual Understanding of Indoor Environments: from Structures to Opportunities for Action

Grace Tsai, Collin Johnson, and Benjamin Kuipers
EECS, University of Michigan, Ann Arbor, Michigan

Introduction

Problem to Solve: Visual perception for agents

Input: **a stream of images**

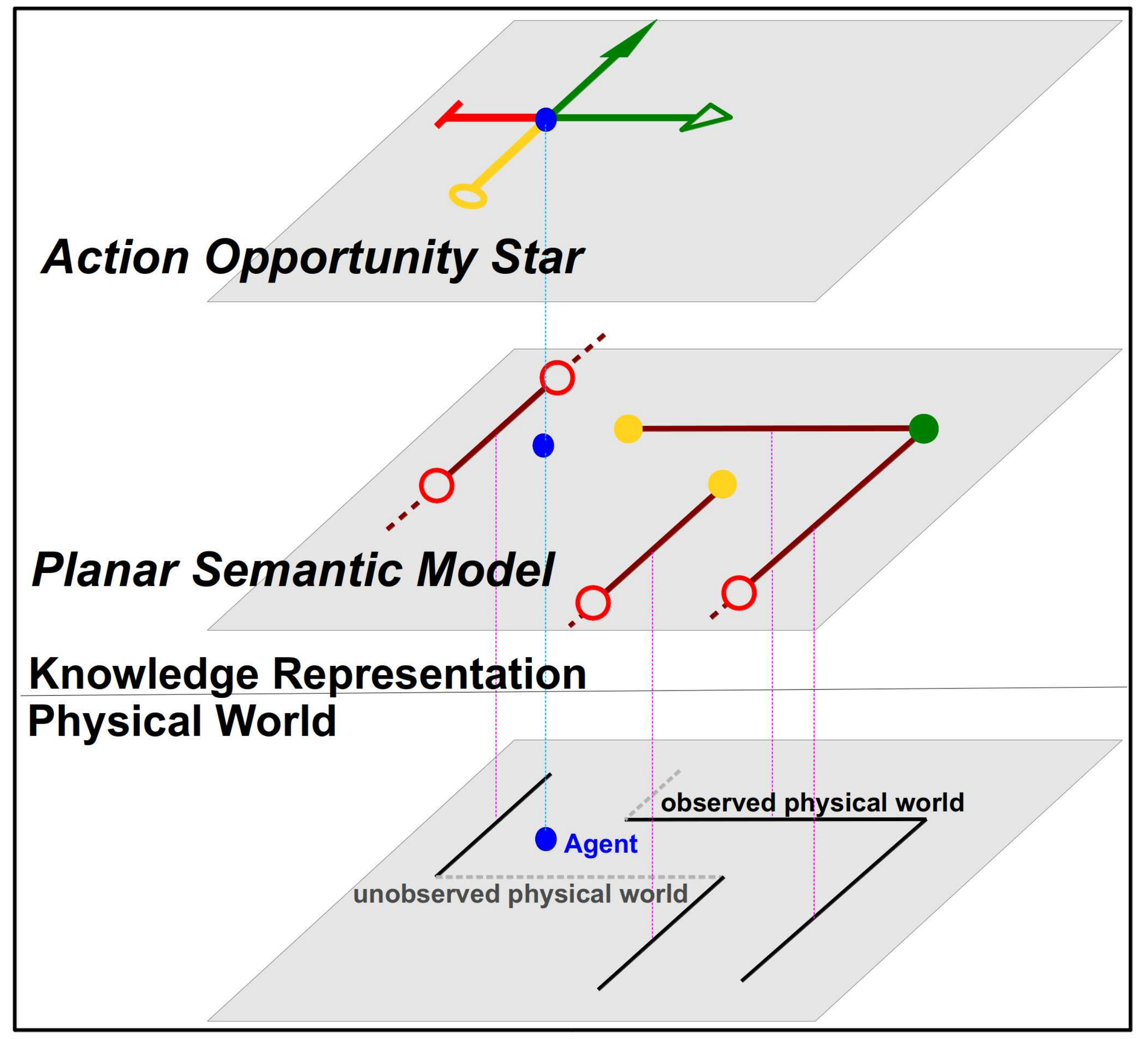
Process: **on-line** and **efficient**

Output: **concise** and **semantically meaningful** representation

- Captures information critical for plans and actions
  - For an indoor navigating robot: free-space and obstacles
  - Manipulation and grasping would have other needs

Contribution

- On-line framework for incremental scene understanding
  - Planar Semantic Mode (PSM): a concise planar representation that captures ground plane and rich relationships among wall segments.
- Move forward from geometric modeling to **reasoning about opportunities for action**.
  - Action Opportunity Star (AOS): an abstraction for navigation opportunities at a given location.
- **Express incomplete knowledge** so that unknown areas can be incrementally built as observations become available.



Planar Semantic Model

Models the indoor environment by meaningful planes

- Ground plane
- Walls
  - Perpendicular to ground but not necessarily to each other
  - A collection of wall segments, delimiting where the wall is present and where there is an opening.

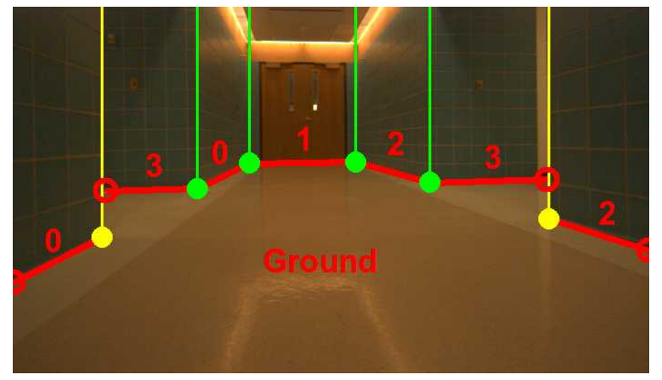
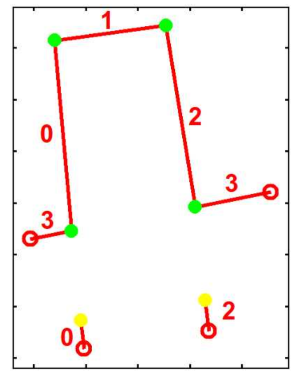


Image Projection of PSM

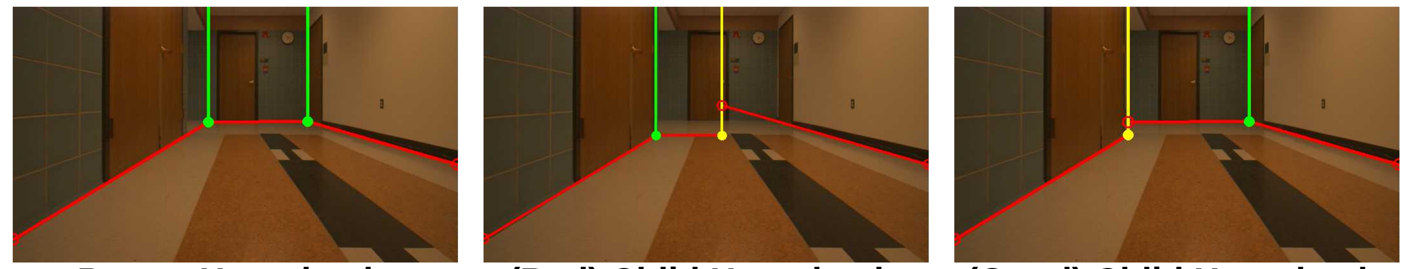


PSM visualized on the ground-plane map

The indices are the wall ID. There are 3 types of endpoints, 3D vertical bound of a wall segment. A **dihedral** endpoint corresponds to two observed wall segments. An **occluding** and an **indefinite** endpoints represent incomplete knowledge of the wall segments.

Construct PSM on-line from a stream of images

- Generate qualitatively distinct PSM hypotheses by image features through an incremental process. [2]



Parent Hypothesis | (Bad) Child Hypothesis | (Good) Child Hypothesis

- Test hypotheses based on their ability to explain 2D motion of a set of tracked features through a Bayesian filter. [1]

$$p(M_i|\mathbf{O}^1,\mathbf{O}^2,...,\mathbf{O}^t) \propto p(M_i)\prod_{j=1...t} p(\mathbf{O}^j|M_i)$$

Likelihood:

$$p(\mathbf{O}^j|M_i)=\prod_{k=0}^{n^t}\exp\frac{-||\hat{\mathbf{L}}^i(o_k^t)-\mathbf{L}(o_k^t)||^2}{2\sigma^2}$$

Tracked location

Predicted location based on hypothesis i

References

[1] Tsai & Kuipers. Dynamic visual understanding of the local environment for an indoor navigating robot. *IROS*, 2012
[2] Tsai et al. Real-time indoor scene understanding using Bayesian filtering with motion cues. *ICCV*, 2011
[3] Johnson & Kuipers. Efficient search for correct and useful topological maps. *IROS* 2012
Dataset: *http://web.eecs.umich.edu/~gstsai/release/Umich_indoor_corridor_2012_dataset.html*

Action Opportunity Star

Describe a circularly ordered list of opportunities for navigation around the agent

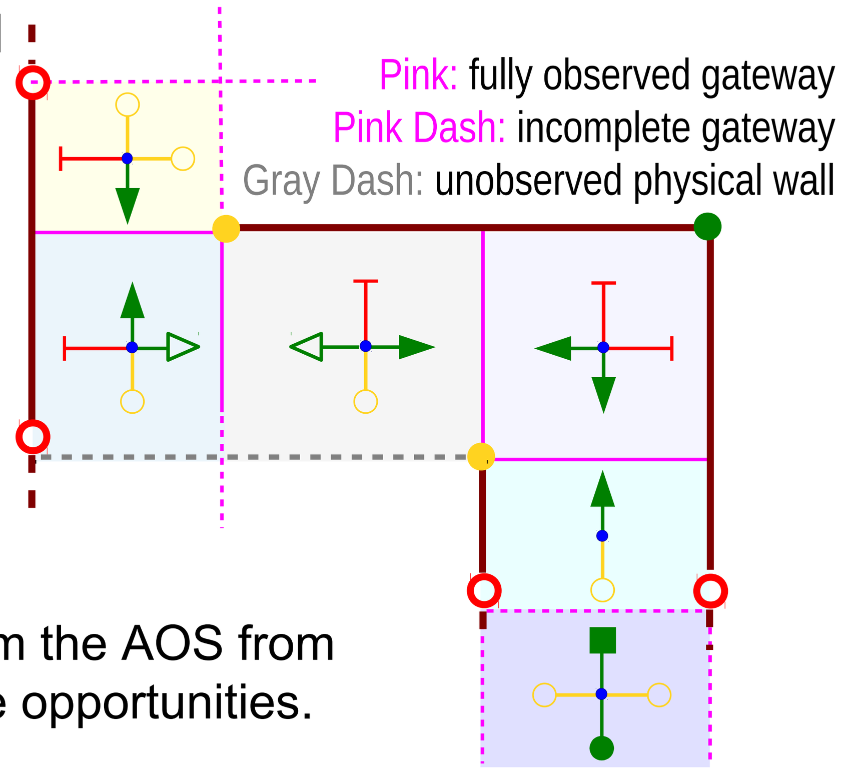
Opportunity is an abstraction, representing groups of trajectories that have the same qualitative effect on the agent's state.

- Path that the opportunity is on
- Direction along the path
- Gateway specifying which trajectory belongs to this opportunity
- Type capturing navigability & level of understanding

observed: navigable; full knowledge; leads agent in/out an intersection
partially observed: navigable but incomplete knowledge on its gateway
unnavigable: prohibits the robot to travel along the path
potential: undetermined navigability; leads agent to unobserved regions
exiting: navigable; leads agent outside its field of interest
beginning: navigable; leads agent into observed region at the beginning

Extract AOS from PSM

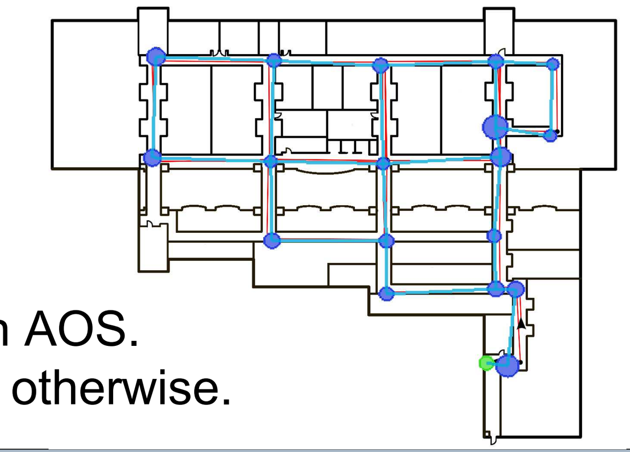
- Extract gateways by linking PSM endpoints and wall segments.
- Form an opportunity from each gateway and determine its type by the PSM endpoint.
- In each local region, form the AOS from simultaneously available opportunities.



- The AOS is invariant over local regions, capturing qualitative properties for the surrounding geometric structure.

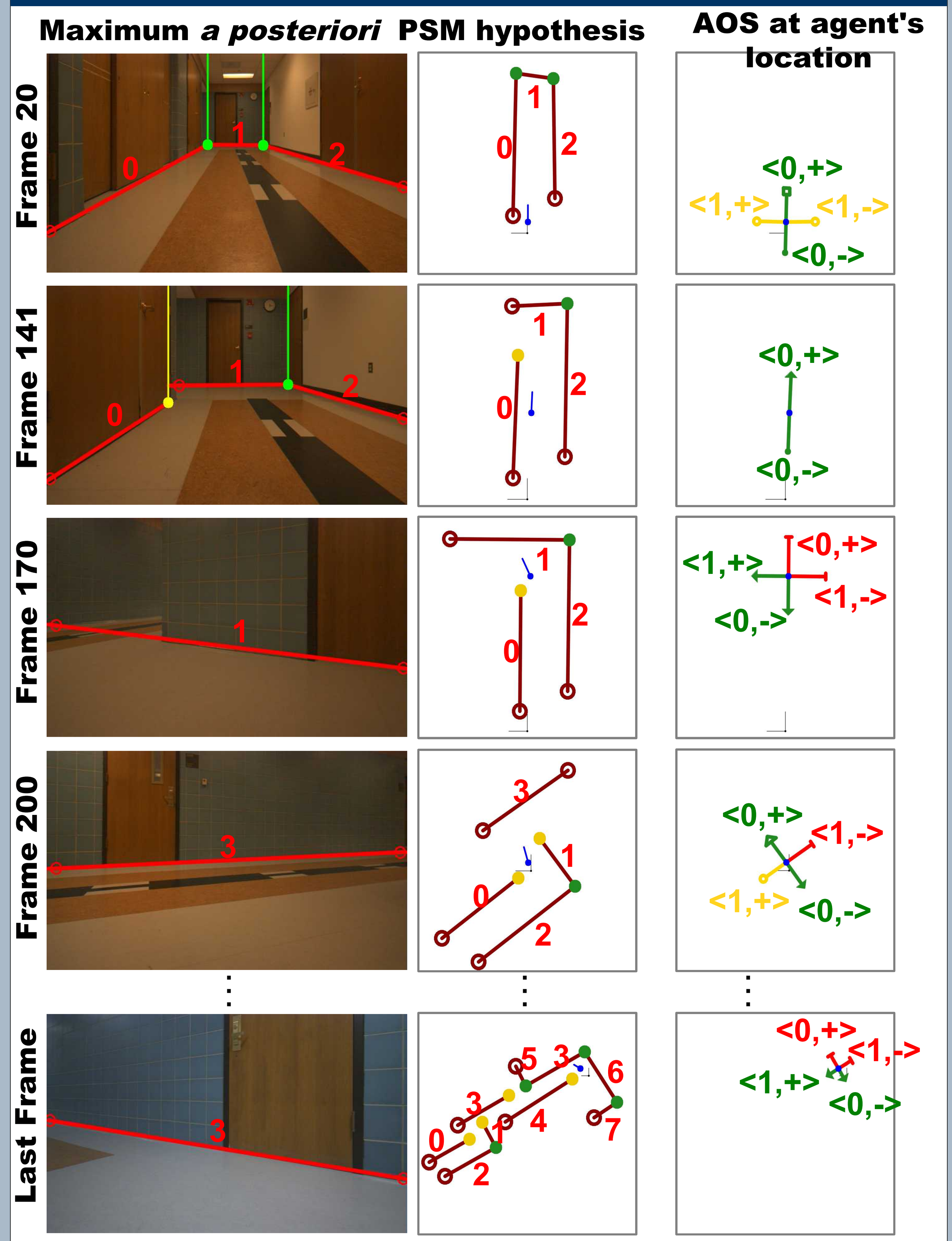
Application

Supports topological mapping in large scale environments. [3]



The agent is ...
on a path, when only one path in AOS.
at a place (and needs decision), otherwise.

Results



- The maximum *a posteriori* (MAP) PSM hypothesis is correct, 92.18% of the time
- AOS is always correct when MAP PSM hypothesis is correct.
- When PSM is incorrect, AOS is still correct, 73.73% of the time.

Acknowledgments

This work has taken place in the Intelligent Robotics Lab in the Computer Science and Engineering Division of the University of Michigan. Research of the Intelligent Robotics lab is supported in part by grants from the National Science Foundation (CPS-0931474, IIS-1111494, and IIS-1252987).