Places, from the Robot's Point of View

Benjamin Kuipers
Computer Science & Engineering
University of Michigan
(Since January 1, 2009)

The World is Infinitely Complex

- The world *itself* is continuous, and continuously, dynamically changing.
- *Perceptual input* from the world is also extremely complex:
  - "blooming, buzzing, confusion"
- An agent's representation and inference resources are finite, and quite limited.
- An intelligent agent (human or robot) must cope with this challenge

Spatial Representations

- I work on representing spatial knowledge.
  - The importance of multiple representations for incomplete knowledge of large-scale space.
  - How to combine rich sensory input about local space, to build useful representations of global space.
- Much of my work has used laser range sensors.
  - But the lessons are still useful for vision.
  - My students and I are beginning to use vision.

Incomplete Knowledge

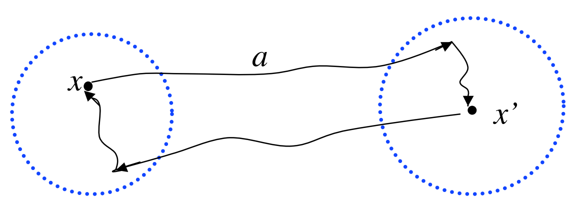
- The ability to represent incomplete knowledge is important due to:
  - Sensor errors and imprecision
  - Limited processing, slow storage and retrieval
  - Unexpected types of environments
- Humans are far more robust than any AI
  - In spite of fixed and sudden limitations
  - In surprising environments
- Incomplete knowledge is a relevant factor.

The Place Abstraction

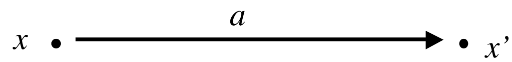
- Focus first on representing space.
  - Large-scale space is space whose structure is larger than the perceptual horizon.
  - Small-scale space has structure within the sensory horizon.
- What are places?
  - In LSS, places are decision points.
  - In SSS, places are regions with gateways.
- *Places* are made up of *distinctive states*.

Learn *Distinctive* States

- A *distinctive state* (location plus orientation) is the isolated fixed-point of a hill-climbing control law.



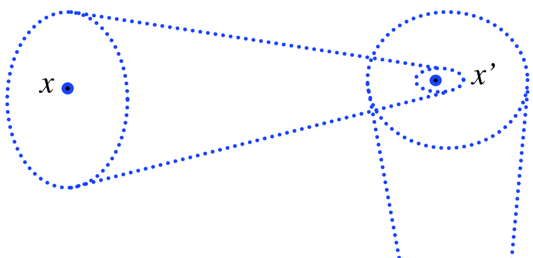
- Reliable motion abstracts to a schema $\langle x, a, x' \rangle$.



- Hill-climbing to a dstate reduces image variability due to pose variation.

Distinctive States

- Between distinctive states, actions are *functionally deterministic*
  - if all final-state uncertainty is contained within every initial-state basin of attraction
- Supports abstraction from continuous to discrete state space.



The basic Spatial Semantic Hierarchy

- The human cognitive map includes multiple ontologies for spatial knowledge:
  - **Control**: select *control laws* to move reliably among *distinctive states*.
  - **Causal**: *actions* link *states*, which have sensory *views*.
  - **Topological**: *places*, *paths*, and *regions* linked by connectivity, order, containment.
  - **Metrical**: *frames of reference*, distance, direction, shape.

[Kuipers, 2000]

What do we (need to) know?

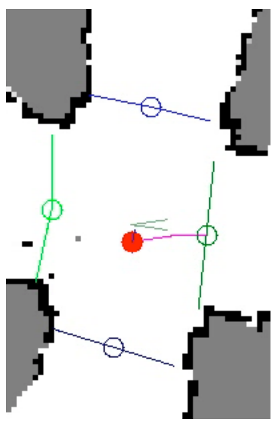
- In the **basic Spatial Semantic Hierarchy**:
  - We don’t need to know sensor semantics at all!
  - We only need reliable hill-climbing (HC) and trajectory-following (TF) control laws.
  - They define distinctive states, places, paths, and the topological map.
- In the **Hybrid Spatial Semantic Hierarchy**:
  - We know enough sensor semantics to build the Local Perceptual Map.
  - Localization in the LPM replaces hill-climbing
  - Stronger assumptions: more powerful mapping

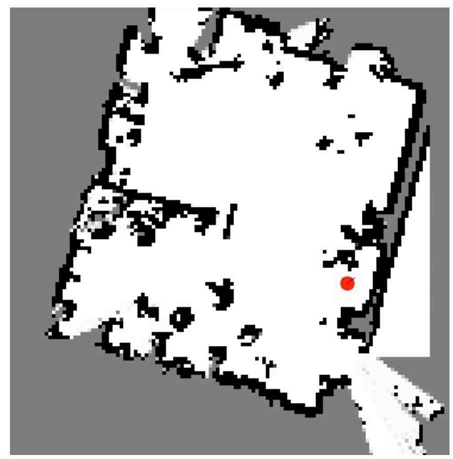
Hybrid Spatial Semantic Hierarchy

	Metrical Mapping	Topological Mapping
Small-scale space	Local SLAM	Local decision structure
Large-scale space	Global metrical map	Global topological map

Hybrid Spatial Semantic Hierarchy

- Accurate localization in small-scale space replaces hill-climbing.

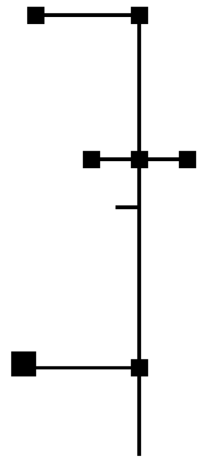




Exploration and Mapping

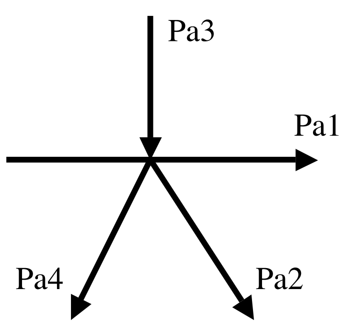
- Maintain a map of local perceptual space.
  - A bounded scrolling map of small-scale space around the agent.
  - The coordinate frame may drift.
- Continually parse the local topological structure.
- Identify places. Build the topological map as a side-effect.

Exploration and Mapping

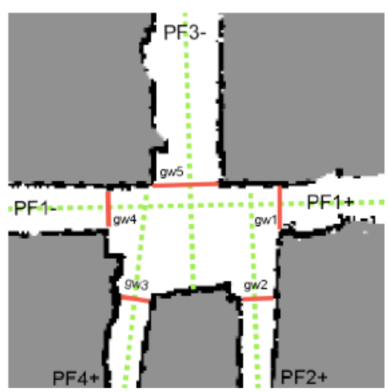


Places at Different Spatial Scales

- Decision point versus trajectory through local place neighborhood.



in large-scale space



in small-scale space

Does a place abstraction always exist?

- Not in truly pathological environments
  - open ocean (but Polynesian navigators use places!)

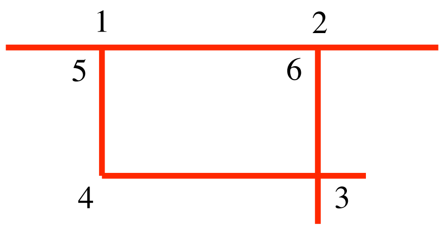
  or with pathological sensors
  - video snow

- **Conjecture: Yes**, with sufficiently rich sensors in a sufficiently rich environment.
  - home and office environments
  - campus/urban indoor/outdoor environments

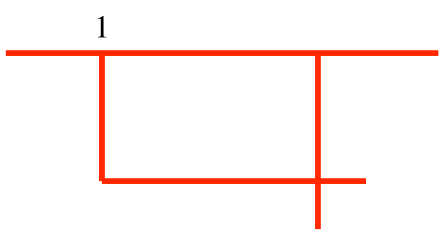
Build the Global Topological Map

- Define a tree of *all possible* topological maps consistent with exploration experience.
  - They are the leaves of this tree.
- For each new action+observation
  - If the map predicts the observation, *OK*.
  - If it contradicts the observation, *prune it*.
  - Otherwise, *branch* on maps with new edges:
    - All possible loop-closing hypotheses
    - One hypothesis of a brand-new place
  - Identify the current best map.

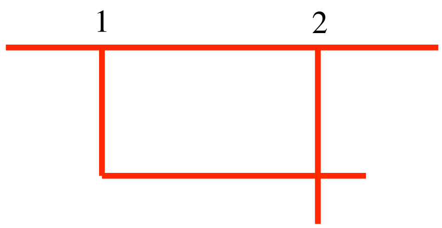
Building the Tree of Maps

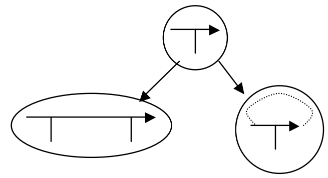


Tree of Maps (1)

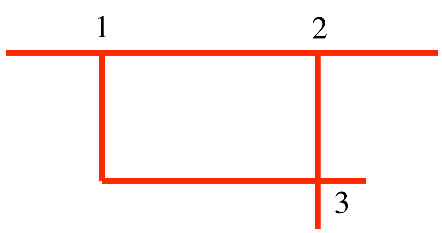


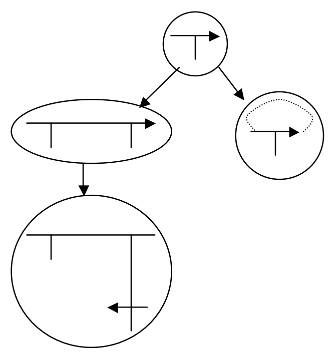
Tree of Maps (2)



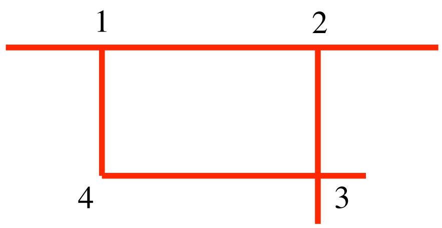


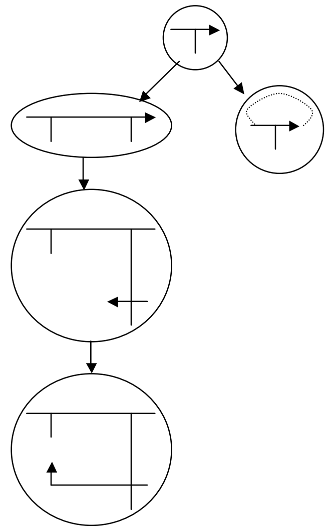
Tree of Maps (3)



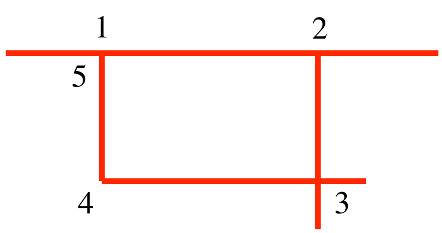


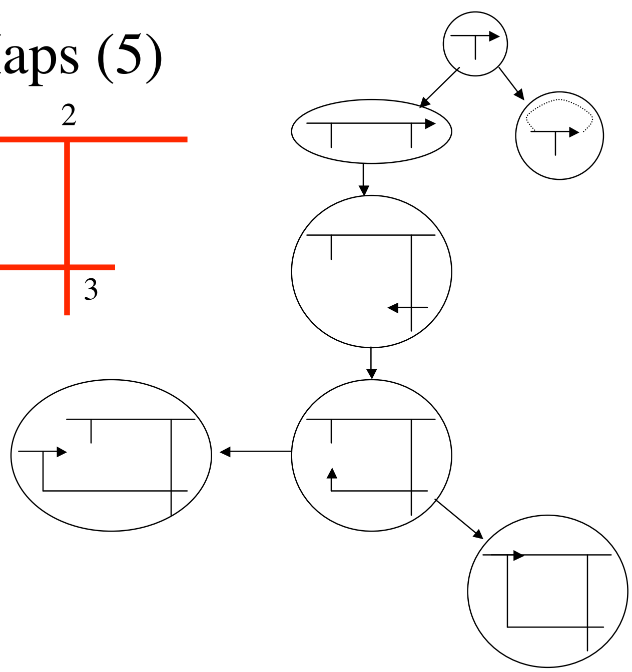
Tree of Maps (4)



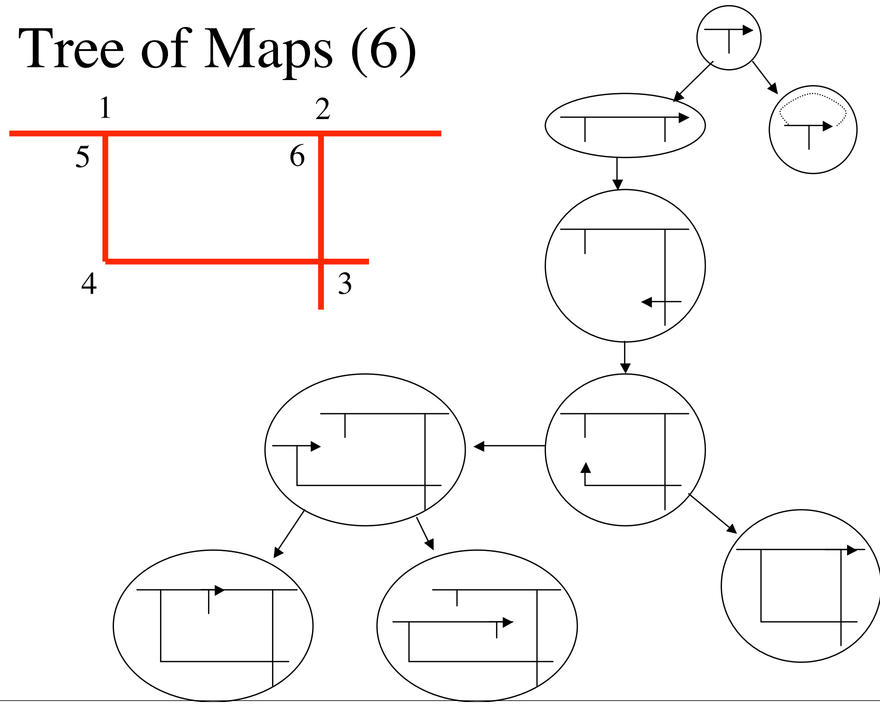


Tree of Maps (5)





Tree of Maps (6)



Rank the Consistent Maps

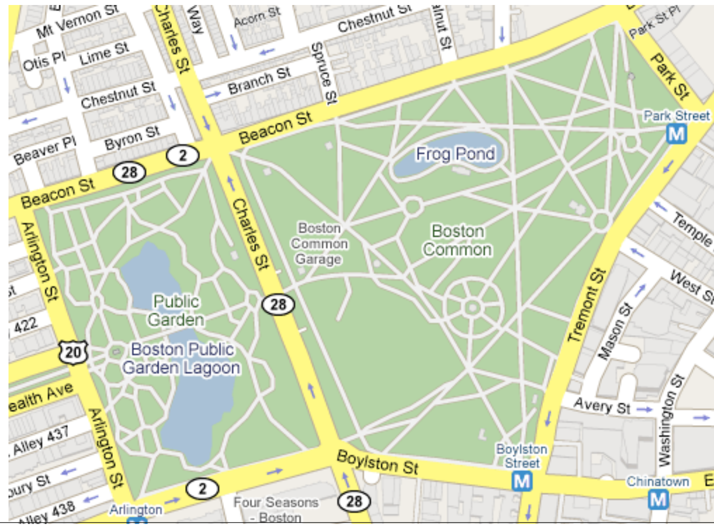
- The tree is **guaranteed** to contain the true map
  - All consistent maps are created.
  - Only inconsistent ones are deleted.
- Each map is a distinct loop-closing hypothesis.
  - Rank the consistent maps by simplicity (# places)
  - and/or likelihood, $p(odometry \mid layout)$.

Use the current best map for planning.

- Remember the tree.
- The current best map could be refuted.

Plausible maps may be wrong

- Especially in Boston!



Markov Localization

$$p(x'|a,z,m) = \alpha \, p(z|x',m) \int p(x'|x,a,m) \, p(x|m) \, dx$$

- Simplified in the SSH topological map.
  - Many fewer states x in m.
  - Reliable actions $\langle x, a, x' \rangle$.
  - Sensory images z clustered to views v.
  - A distinctive state has a single view: $view(x', v)$

$$p(x'|a,v,m) = \alpha \sum \{ p(x|m) \ : \ \langle x, a, x' \rangle \wedge view(x', v) \}$$

Recognizing Place Instances

- Closing a loop:
  - “*I have been here before*.”
- Problem 1: ***Perceptual aliasing***
  - Different places look the same (false pos).
- Problem 2: ***Image variability***
  - The same place looks different (false neg).
- Image variability is the most important problem with rich sensors.
  - So we start by abstracting it away!

Radical Abstraction of Sensory Images

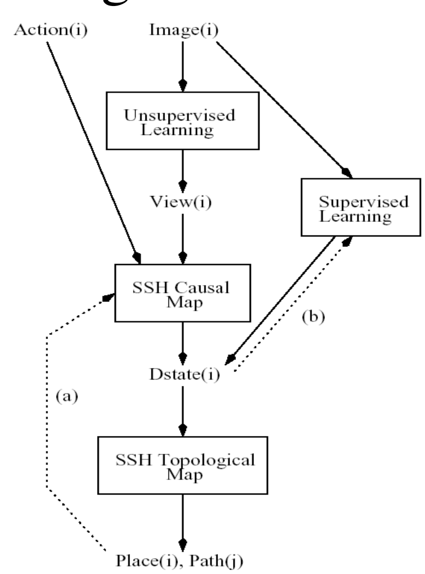
- Search for a clustering of sensory images that eliminates image variability.
  - Even by increasing perceptual aliasing.
  - Local place topology (+, T, L, etc.) will do.
- Each distinctive state has a single view: $view(x,v)$
  - Different dstates may have the same view.

Abstraction and Bootstrapping

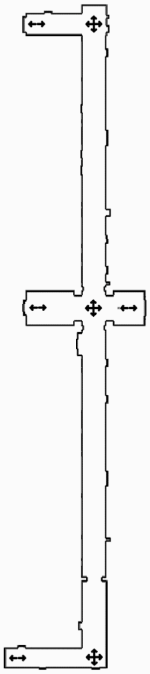
- Our approach [Kuipers & Beeson, AAAI-02]:
  - *Abstract the continuous world* to distinctive states in a topological map.
  - *Abstract rich sensory images* to views, simple enough to eliminate image variability.
  - *Build a topological map*, resolving ambiguity, to eliminate perceptual aliasing.
  - *Do supervised learning* for place recognition, using rich sensory images, and exploiting the topological map for supervision.
  - This can identify subtle discriminating features in the rich images.

Our Solution: Bootstrap Learning for Place Recognition

- *Use an* ***unsupervised learning*** *method*
  - cluster sensory images into views
- *to prepare for a* ***deductive*** *method*
  - build a topological map
- *that supports a* ***supervised learning*** *method*
  - nearest neighbor
- *to* ***recognize places*** *from rich images.*

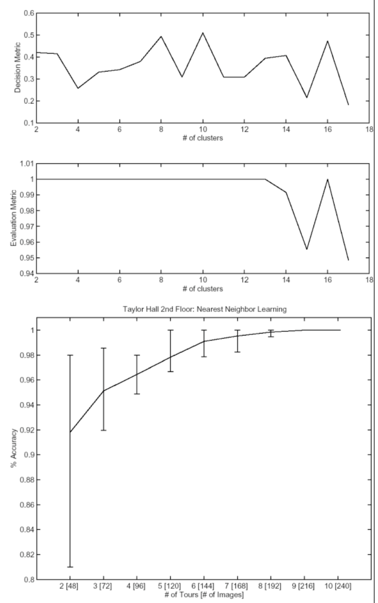


[Kuipers & Beeson, AAAI-02]



Experimental results

- 10 circuits, 240 images
  - 80 meters long
  - 10 independent trials
- Decision metric picks *k*=10 clusters (views)
  - perceptual aliasing
- Evaluation metric says *k*=13 would work.
- Mapper identifies
  - 20 dstates
  - 7 places, 4 paths
- Recognition accuracy rises rapidly to 100%
  - based on rich sensory images.



We are moving from Laser to Vision

- We are beginning to use visual SLAM to build the local perceptual map.
  - [Aniket Murarka, et al, 2006, 2008, 2009]
- The local 3D model identifies hazards that would be invisible to lasers:
  - Drop-offs, ramps, overhangs
- Project to a 2D safety map
  - Appropriate for mobile robot motion planning

Invariant Features Improve on Views

- At pose x, the agent observes a rich sensory image $z = g(x)$ with a set of features $F = F(z)$.
  - e.g., SIFT, SURF, corner, etc.
- Features are sparse in a very large space.
  - Reduces perceptual aliasing (false pos)
- Features are invariant to small pose variation.
  - Reduces image variability (false neg)
- Combine evidence to get strong confidence from many weak feature matches.
  - More flexible than deterministic views.

Instances and Categories

- An agent perceives *instances*.
  - *Categories* are inferred.
- A fully autonomous agent would learn its *own* categories from perceived instances.
  - But human designers can build in culturally-conventional categories, to meet their goals.
- Knowledge of image category (visual "gist") provides useful priors for recognizing objects and features.
  - Is the taxonomy of place categories interesting?

Summing Up

- How can an agent manage rich perceptions of an infinitely complex world?
  - By using multiple different abstractions.
  - E.g., “place” has several different meanings.
- Bootstrap learning to recognize places:
  - Use abstract views and places to build a map;
  - Use the map to identify places uniquely;
  - Use supervised methods to learn to recognize places from the rich sensory image.
- Visual sensing makes abstraction even more important.

References

- Kuipers, **The Spatial Semantic Hierarchy**. AIJ, 2000.
- Kuipers & Beeson, **Bootstrap learning for place recognition**. AAAI, 2002.
- Kuipers, **An intellectual history of the Spatial Semantic Hierarchy**. In Jefferies & Yeap (edited volume), Springer, 2008
- Murarka & Kuipers, **A stereo vision based 3D mapping algorithm for detecting ramps, drop-offs, and obstacles for safe local navigation**. IROS, 2009.
- Beeson, Modayil & Kuipers, **Factoring the mapping problem: Mobile robot map-building in the Hybrid Spatial Semantic Hierarchy**. IJRR, 2009.

- http://eecs.umich.edu/~kuipers/research/ssh/papers.html