

Learning Sensor, Space and Object Geometry

Jeremy Stober *

Department of Computer Science
The University of Texas at Austin
stober@cs.utexas.edu

Abstract

Robots with many sensors are capable of generating volumes of high-dimensional perceptual data. Making sense of this data and extracting useful knowledge from it is a difficult problem. For robots lacking proper models, trying to understand a stream of uninterpreted data is an especially acute problem. One critical step in linking raw uninterpreted perceptual data to cognition is dimensionality reduction.

Current methods for reducing the dimension of data do not meet the demands of a robot situated in the world, and methods that use only perceptual data do not take full advantage of the interactive experience of an embodied robot agent. This work proposes a new scalable, incremental and active approach to dimensionality reduction suitable for extracting geometric knowledge from uninterpreted sensors and effectors.

The proposed method uses *distinctive state abstractions* to organize early sensorimotor experience and *sensorimotor embedding* to incrementally learn accurate geometric representations based on experience. This approach is applied to the problem of learning the geometry of sensors, space, and objects. The result is evaluated using techniques from statistical shape analysis.

Introduction

In the early stages of development infants form ego-centric models of the world, which then form the basis for learning more advanced concepts. A robot waking up in an unfamiliar body faces a similar challenge, and acquiring an ego-centric model that includes details of the sensor, space, and object geometry would facilitate learning more advanced concepts. One immediate barrier to acquiring geometric representations is the high-dimensional nature of uninterpreted sensor signals. The real-time high-dimensional nature of the signals presents a real challenge for existing state-of-the-art manifold learning and dimensionality reduction methods.

*This work has taken place in the Intelligent Robotics Labs at the University of Texas at Austin and at the University of Michigan. Research in the Intelligent Robotics Labs is supported in part by grants from the National Science Foundation (IIS-0713150 to UT Austin and CPS-0931474 to UM) and from the TEMA-Toyota Technical Center to UM.
Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

When robots have access to sensor and action models as well as a target representation of space, then the models provide an effective method for reducing the dimension of sensory signals by allowing the agent to estimate the value of a small set of unobservable variables in the target representation. For example, solutions to the important problem of simultaneous localization and mapping (SLAM) involve an agent using sensor and action models to estimate both the world geometry and the robot pose.

The introduction of apriori models into any robotic system does have disadvantages. For example, the models have to be supplied by an engineer who is knowledgeable about the relevant aspects of the design of the robot. Once a robot is operating independently in the world, the assumptions that went into designing the original system may break down due to physical changes in the robot.

Unlike robots, a human's ability to interpret sensor signals is remarkably plastic. For example, humans can adapt over time to such radical changes as complete inversions of the visual field (Dolezal 1982) or even entirely new modes of visual input (Bach-y Rita 2004). State of the art robots would fail under the same difficult circumstances, yet humans show a remarkable ability to adapt. The goal of this work is to develop methods that give robots similar abilities.

Related Work

In manifold learning the variation in perceptual information is assumed to be caused by variation in a small number of continuous unobservable variables. As an example, in a static world the variation of a robot's sensor readings over time is the result of a robot's changing pose. Even though the variation in robot sensor readings is high dimensional, the cause of those variations is attributable to the variation in a small number of pose parameters.

For pure perceptual methods, attempting to infer changes in latent variables is a difficult task because of the complexity of the relationship between the latent space and the high-dimensional observations. In an interactive environment, where changes in agent perception are mediated by agent actions, better methods of discovering low dimensional representations are possible.

For instance, Bowling et al. (2007), in their work on action respecting embedding (ARE), demonstrated that including information about agent actions is useful for forming

accurate low-dimensional embeddings of sensor information. In other work, Philipona et al. (2010) have shown how ISOMAP can be extended using information derived from local agent actions.

The approach presented here builds on prior work and seeks to solve several outstanding issues. For example, bounds on computational resources makes applying manifold learning difficult in situations involving lifetime learning of embodied agents. In addition, methods that use only local information like ISOMAP can fail to recognize important aspects of the global state space geometry. Finally, many traditional methods for dimensionality reduction fail to identify the appropriate scale for data, opting instead for some normalized scale for the resulting representation. The method presented here offers a compelling alternative, where scale is determined based on agent actions and calibrated to an agent’s own body.

Modeling Geometry

In the completed and proposed work, an agent takes a multi-step approach to acquiring geometric knowledge. First, an agent forms a *distinctive state abstraction (DSA)* using a simple set of hill-climbing control laws. The fixed points of these control laws are the distinctive states. The *DSA* allows the agent to scale the learning process to large domains by separating large domains into local learning problems around each distinctive state. The *DSA* provides a concrete set of policy improvement goals for efficiently moving to and between distinctive states.

As the agent learns progressively better policies, the agent applies *sensorimotor embedding* to extract geometric knowledge implicit in the learned policies. Sensorimotor embedding first computes similarities between all states in the region of a single distinctive state. The similarity between any two states, s and s' , is computed by comparing the sequences of actions, $\langle a_i \rangle_1^{n-1}$ and $\langle a'_j \rangle_1^{m-1}$, generated by the policy π_i , that bring an agent to a shared distinctive state. Formally,

$$\delta_{\pi_i}(s, s') \equiv \text{DTW}(\langle a_i \rangle_1^{n-1}, \langle a'_j \rangle_1^{m-1}),$$

where DTW stands for dynamic time warping.

Within each region of a distinctive state, the geometric representation of the local state space is generated using multidimensional scaling applied to the matrix of dynamic time warping comparisons (Figure 1).

Completed and Proposed Work

Stober et al. (2009) have shown how sensorimotor embedding can be applied to learning the structure of a foveated (non-uniform) sensor. In that work the sensor geometry was derived from a learned saccade policy. This method was shown to have several important advantages, as the learned sensor geometry could adapt dynamically to considerable changes in actual sensor geometry, including total visual inversion and simulated macular degeneration.

Proposed work will focus on learning the structure of space and objects using the same process. For learning spatial geometry, this method will be compared directly with

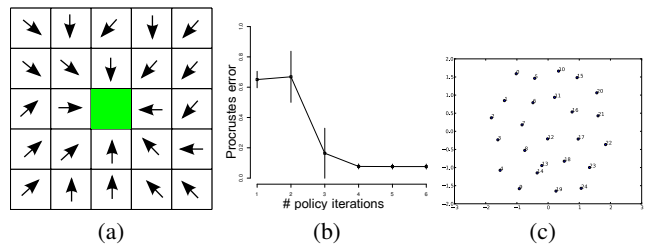


Figure 1: (a) shows an optimal policy in a grid-world with one distinctive goal state. (b) shows the decrease in geometry error as the policy improves. (c) shows the final representation of the relative positions of the states using sensorimotor embedding and an optimal policy.

ARE in the IMAGEBOT domain. In addition, recent work on learning object geometry in computer vision has relied on relating distinctive views of objects using homographies (Savarese and Fei Fei 2007). The proposed method will build object models using distinctive states and learned policies for navigating to and between these distinctive states.

The primary method of evaluation will involve comparing computed geometry with ground truth geometry using sample points drawn from both representations. These point sets can be compared using statistical shape analysis. Secondary analysis such as scree diagrams and subjective comparisons of the resulting representations will also be performed.

Conclusion

The proposed and completed work provides a incremental, scalable and active method for learning geometry from sensorimotor experience. Both *DSAs* and sensorimotor embedding provide developing agents with new tools to understand and adapt to uninterpreted sensors and effectors in complex environments.

References

Bach-y Rita, P. 2004. Tactile sensory substitution studies. *Annals of the New York Academy of Sciences* 1013:83.

Bowling, M.; Wilkinson, D.; Ghodsi, A.; and Milstein, A. 2007. Subjective localization with action respecting embedding. *Robotics Research* 28:190–202.

Dolezal, H. 1982. *Living in a World Transformed: Perceptual and Performatory Adaptation to Visual Distortion*. Academic Press.

Philipona, D., and O’Regan, J. 2010. The sensorimotor approach in CoSy: The example of dimensionality reduction. *Cognitive Systems* 95–130.

Savarese, S., and Fei Fei, L. 2007. 3D generic object categorization, localization and pose estimation. In *IEEE 11th International Conference on Computer Vision*, 1–8.

Stober, J.; Fishgold, L.; and Kuipers, B. 2009. Learning the sensorimotor structure of the foveated retina. In *Proceedings of the Ninth International Conference on Epigenetic Robotics*.