Robot Cognition using Bayesian Symmetry Networks

Anshul Joshi, Thomas C. Henderson and Wenyi Wang
University of Utah

UUCS-13-005

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

19 November 2013

Abstract

[10] proposed a generative theory of shape, and general cognition, based on group actions on sets as defined by the wreath product. Our position expressed here is that this approach can provide a strong basis for robot cognition when:

1. tightly coupled to sensorimotor data and analysis,
2. used to structure both general concepts and specific instances, and
3. combined with a probabilistic framework (Bayesian networks) to characterize uncertainty.

We describe a roadmap to achieve these and provide some evidence of feasibility.

1 INTRODUCTION

Our goal is to develop cognitive capabilities for autonomous robot agents. [11] state that cognition "can be viewed as a process by which the system achieves robust, adaptive, anticipatory, autonomous behavior, entailing embodied perception and action." For us, this includes the ability to:

1. analyze sensorimotor data in order to determine a rational course of action,
2. represent and recognize important concepts describing the world (including its own self),
3. recognize similarities between concepts,
4. extend concepts to new domains, and
5. determine likelihoods of assertions about the world.

We have previously argued that an effective basis for robot cognition requires some form of innate knowledge; see [6]. In fact, most current robot implementations rely mostly on innate knowledge programmed in by the system builder (e.g., sensor data analysis algorithms, object recognition, kinematics and dynamics, planning, navigation, etc.), although some aspects of the robot's knowledge may be learned from experience. This approach to establishing robot cognitive faculties does not scale well, and as a more effective and efficient paradigm, we have proposed that a collection of symmetry theories form the basic innate knowledge of an autonomous agent [5, 8, 7, 9]. These are defined as formal, logic-based systems, and then the robot proceeds to use this knowledge by analyzing sensorimotor data to discover models for the theories. For example, the axioms which define a group require a set of elements and an operator that acts on the set such that the closure, identity, inverse and associativity properties are satisfied. Figure 1 shows how this works. Namely, sets and operators are found in the sensorimotor data and checked to see is they satisfy the axioms. If so, then the robot may use the theory to test new assertions (i.e., theorems) that follow from the axioms and have direct meaning for the model, but which may never have been experienced by the robot.

We propose that in addition to a logical framework to allow assertions about the world, robot cognition requires a representational mechanism for sensorimotor experience. The wreath product provides this; for a more detailed account of wreath products, see [2, 4, 10]. The wreath product is a group formed as a splitting extension of its normal subgroup (the direct product of some group) and a permutation group. As an example, consider the wreath

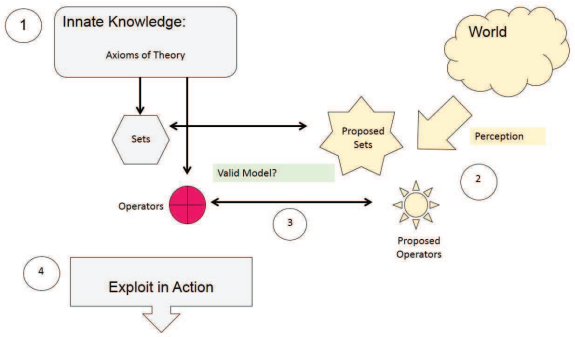

Figure 1: Models for Theories are Discovered: (1) A theory is defined; (2) Sets and Operators are Hypothesized from Sensorimotor Data; (3) Axioms are Validated for Hypotheses; (4) Theory is Exploited.

product representation of the outline of a square shape: $\{e\} \wr \mathcal{Z}_2 \wr \Re \wr \mathcal{Z}_4$, where $\{e\}$ represents a point, $\mathcal{Z}_2$ is the mod 2 group meaning select the point or not, $\Re$ is the continuous (one-parameter Lie) translation group, $\mathcal{Z}_4$ is the cyclic group of order 4, and $\wr$ is the wreath product. [Note that: $\Re \wr \mathcal{Z}_4 = \Re \times \Re \times \Re \times \Re \rtimes \mathcal{Z}_4$ where $\rtimes$ is the semi-direct product.] Leyton [10] describes this representation in great detail and explains the meaning of the expression as follows:

1. $\{e\}$: represents a point. (More precisely, the action of the group consisting of just the identity acting on a set consisting of a point.)
2. $\{e\} \wr \mathcal{Z}_2$: represents the selection of a point or not; i.e., to get a line segment from a line, just select the points on the line segment by assigning the 1 element to them.
3. $\{e\} \wr \mathcal{Z}_2 \wr \Re$: represents a line segment; i.e., one side of the square.
4. $\{e\} \wr \mathcal{Z}_2 \wr \Re \wr \mathcal{Z}_4$: represents the fact that each side of the square is a rotated version of the base segment; i.e., the four elements of $\mathcal{Z}_4$ represent rotations of 0, 90, 180 and 270 degrees.

Each new group to the right side of the wreath symbol defines a control action on the group to the left, and thus provides a description of how to generate the shape. E.g., to draw a

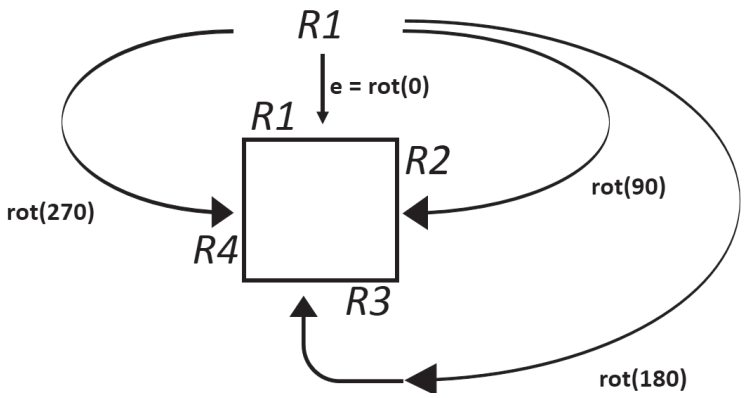

Figure 2: Control Flow in a Wreath Product gives an Explicit Definition in Terms of Actuation of how to Generate the Shape; e, the group identity (rotate by 0 degrees) acts on the first copy, $\Re_1$ to obtain the top of the square; r_{90} acts on $\Re_1$ by rotating it 90 degrees about the center of the square to obtain the right side, $\Re_2$; r_{180} acts on $\Re_1$ by rotating it 180 degrees to obtain $\Re_3$, and finally, r_{270} acts on $\Re_1$ rotating it 270 degrees to obtain $\Re_4$.

square, start at the specified point (details of the x,y values of the point, etc. are left to annotations in the representation), translate some length along a straight line, then rotate 90 degrees and draw the next side, and repeat this two more times. Figure 2 shows how control flows from the rotation group, $\mathcal{Z}_4$, down to copies of the translation group, $\Re$.

Thus, taken together, innate theories allow the discovery of the basic algebraic entities (i.e., groups) of the wreath products, and wreath products allow the description of concepts of interest (here restricted to shape and structure). In the following sections, we expand on the exploitation of sensorimotor data (actuation is crucial) in operationally defining wreath products.

Finally, we propose a probabilistic framework for characterizing the uncertainty in wreath products. The wreath product maps directly onto a Bayesian network (BN) as follows:

1. The rightmost group of the wreath product forms the root node of the BN. Its children are the direct product group copies to the left of the semi-direct product operator.
2. Recursively find the children of each interior node by treating it as the rightmost group of the remaining wreath product.

For example, consider the square expressed as $\{e\} \wr \mathcal{Z}_2 \wr \Re \wr \mathcal{Z}_4$. Figure 3 shows the corresponding BN.

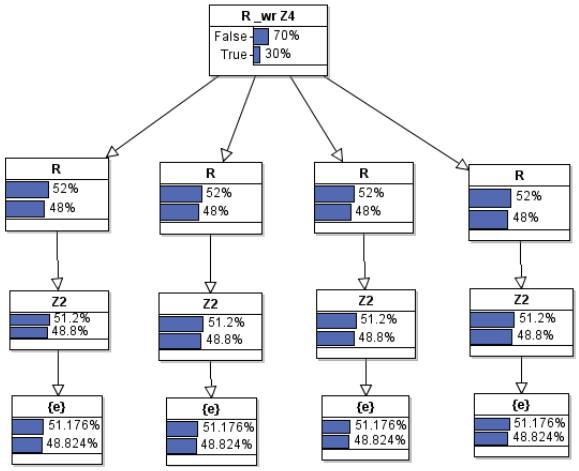

Figure 3: Bayesian Network for the Wreath Product of a Square Created by Rotation.

[Note that the continuous translation group $\Re$ calls for an uncountable number of cross product groups (one for each real), but we do not implement this explicitly.] The following sections describe how the conditional probability tables can be determined from frequencies encountered in the world (e.g., the likelihood of squares versus other shapes), and experience with the sensors and actuators (e.g., the likelihood of the detection of the edge of a square given that the square is present in the scene). [1] proposed a Bayesian network approach for object recognition based on generalized cylinders, but our method differs in that we use the more general wreath product representation (which can represent generalized cylinders) and include actuation in the description, and we derive the BN's from the wreath product; also see [12].

2 ROBOT CONCEPT FORMATION

We now describe an architecture to allow object representations to be constructed from groups discovered in sensorimotor data, and combined to form wreath products. Figure 4 gives a high-level view of the interaction of innate knowledge and sensorimotor data to produce derived knowledge. A lower level architecture is given in Figure 5 involving long-term memory, short-term memory, behavior selection and sensorimotor data. Cognition can be driven by the data (bottom-up) or by a template presented in the behavior selection unit (top-down). In the former case, symmetries are detected in the data which gives rise

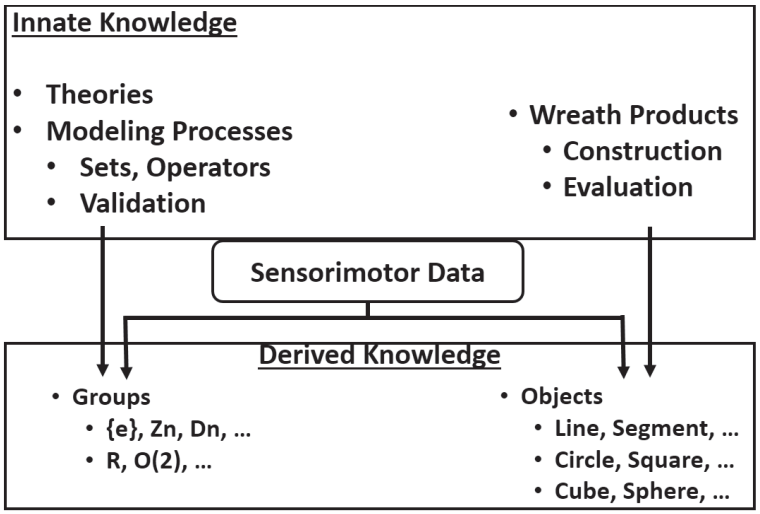

Figure 4: High-Level Cognitive Process.

to the assertion of groups. Then the structure of these groups is determined to form wreath products.

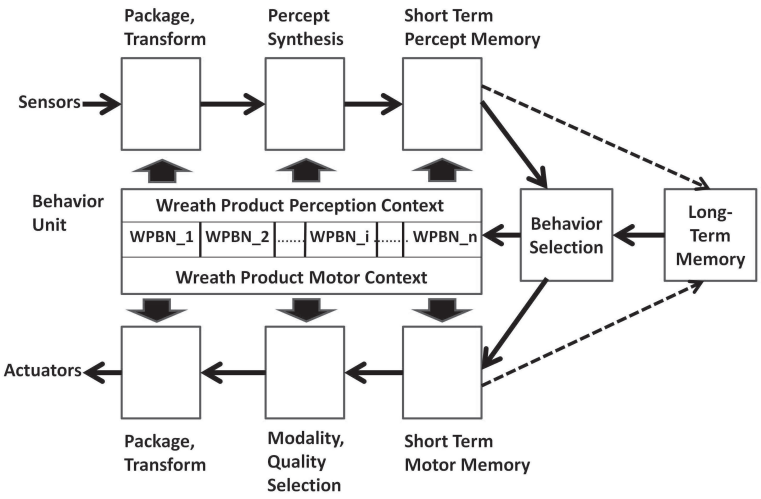

Figure 5: Lower Level Layout of Cognitive Process.

Suppose that data from a single object (a square) is to be analyzed. Figure 6 shows a number of group elements that can be discovered by symmetry detectors applied to image data of a square; this includes translation, reflection and rotation symmetries; we have described symmetry detectors for 1D, 2D and 3D data elsewhere [8, 7, 9]. The symmetry groups would allow the synthesis of $\Re \wr \mathcal{D}_4$ (this is the same as $\Re \times \Re \times \Re \times \Re \rtimes \mathcal{D}_4$) as

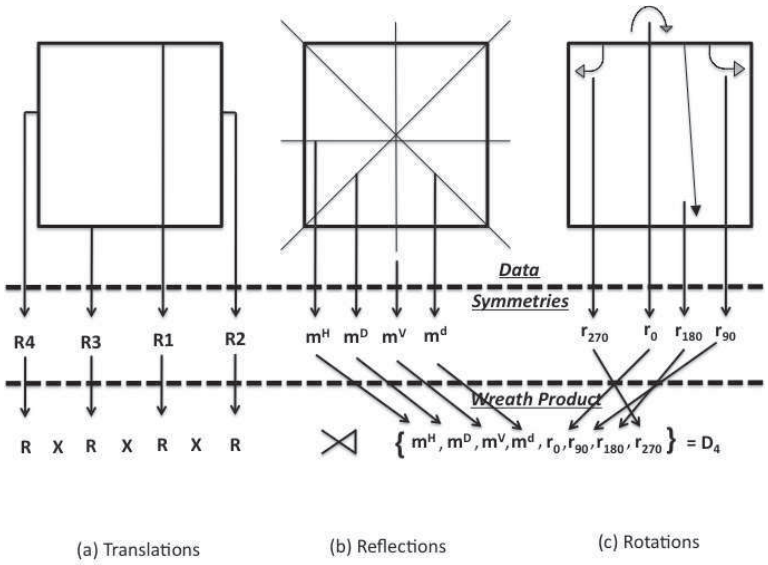

Figure 6: Symmetry Detection in Image of a Square.

a representation of the data. On the other hand if the drawing of the square were observed, and it was done by starting with the top most edge and drawing clockwise around the square, then $\Re \wr \mathcal{Z}_4$ would be generated. If first the top edge were drawn, then the bottom edge, then the right, then the left, this would give rise to: $\mathcal{Z}_2 \wr \mathcal{Z}_2$ Actuation used to observe the data can also be used to determine the appropriate wreath product; e.g., if motors move the camera so that the center of the image follows the contour of the square, then this gives rise to a symmetry in the angles followed (one angle ranges from 0 to 2π, while the other follows a periodic trajectory. [Also, note that all these wreath products could have a pre-fixed $\{e\} \wr \mathcal{Z}_2$ which represents selecting the points on the line segment.]

Next consider the 3D cube. Just like the square, there are several wreath product representations, each corresponding to a distinct generative process. Actual 3D data (e.g., from a Kinect), will most likely come from viewing one face of the cube or three faces.; Figure 7 shows an example Kinect image for these two situations in the case of a cube-shaped footstool. For the 3-face view, each visible face has a corresponding hidden parallel face which can be viewed as either generated from the visible face by reflection or rotation. The $\mathcal{Z}_2$ group represents either of these (in the specific encoding, we choose reflection). Figure 8(a)(Top) shows the cube viewed along a diagonal axis through two corners of the cube, and the $\mathcal{Z}_3$ rotational symmetry can be seen; i.e., a rotation of 120 degrees about the axis (K in Figure 8(a)(Bottom)) sends each pair of parallel faces into another pair. The wreath product for this is then $\Re^2 \wr \mathcal{Z}_2 \wr \mathcal{Z}_3$ while the tree structure shown in Figure 8(b) gives the control action tree.

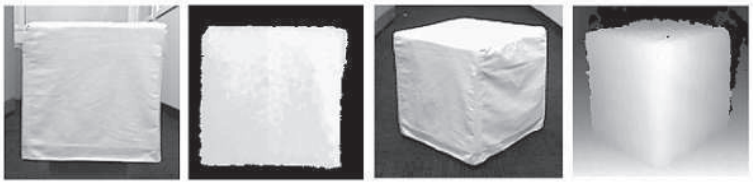

Figure 7: (a) One Face View RGB (b) One Face View Depth Map (c) Three Face View RGB (d) Three Face View Depth Map

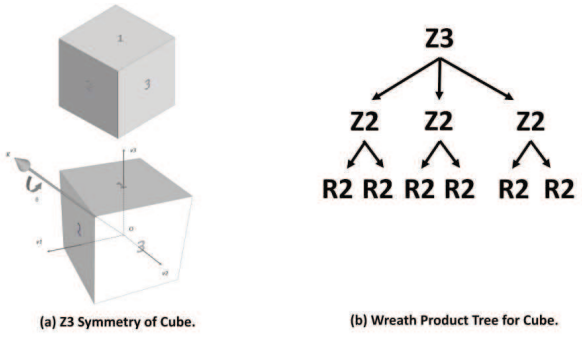

Figure 8: (a) View of Cube along Diagonal Axis with $\mathcal{Z}_3$ Symmetries (b) the Control Action Tree for $\Re^2 \wr \mathcal{Z}_2 \wr \mathcal{Z}_3$

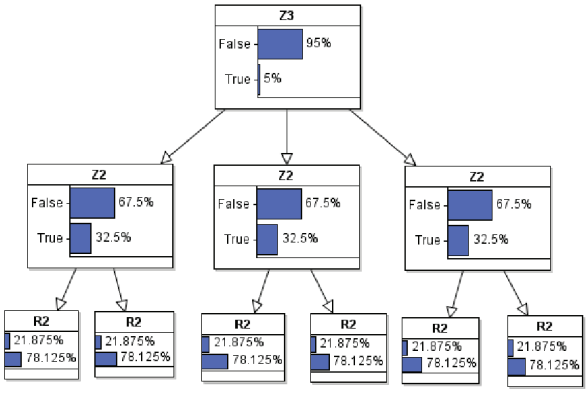

Figure 9: Bayesian Network for Cube.

3 BAYESIAN SYMMETRY NETWORKS

We have already described how the wreath product provides a nested description of an object which maps in a natural way onto a causal graph structure for a Bayesian network. As a working example, we continue with the cube as described in Figure 8(b) for which the corresponding Bayesian network[1] is shown in Figure 9. The graph structure is defined by the wreath product, and the conditional probability tables are determined by considering the context (indoors), the sensor data noise, and the algorithm error. The top-most node $\mathcal{Z}_3$ has a prior of 5% true, since this condition holds for normals of any right-angle 3-faced corner. The $\mathcal{Z}_2$ nodes have a probability of 30% if there is no known $\mathcal{Z}_3$ symmetry, otherwise 80%, and the $\Re^2$ flat face nodes have conditional probability of 70% if no $\mathcal{Z}_2$ symmetry is known, otherwise 95%. The figure shows the likelihoods of each symmetry assertion with no evidence. Figure 10 shows the changes in probabilities for the network when it is known that 3 faces exist, and that there is a $\mathcal{Z}_3$ symmetry for them. Note that the likelihoods for the unseen parallel faces rise to 91% in this case. This type of information may not be readily available to a robot without this cognitive structure.

This works reasonably well in practice; Figure 11 shows Kinect data for a cube shape viewed along a diagonal and surface normals. To determine if the $\mathcal{Z}_3$ symmetry exists, we determine the average normal for each face, then check the symmetry of the three normals

[1]This network was constructed using the AgenaRisk software which accompanies [3].

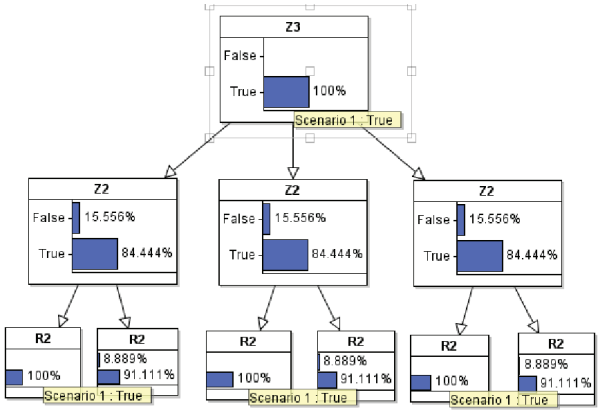

Figure 10: Bayesian Network for Cube with Observations Asserted.

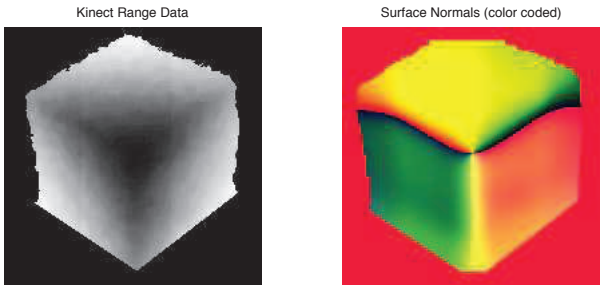

Figure 11: Kinect Cube Data, and Surface Normals.

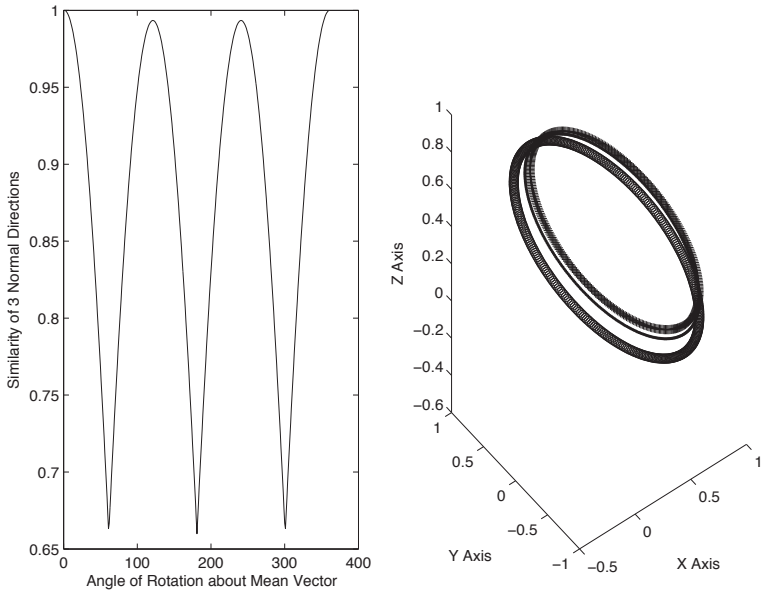

Figure 12: (a) Similarity of 3 Surface Normals under Rotation about their Mean Vector; (b) Trajectories of Normal Endpoints under Rotation.

rotated through 2π radians about their mean vector. Figure 12(a) shows the similarity measure (best match of three original normals with rotated versions of themselves, while (b) shows the trajectories of the three normal endpoints under the rotation.

4 CONCLUSIONS AND FUTURE WORK

Our position is that wreath product representations of objects provide a very powerful object concept mechanism, especially when combined with deep connections to sensorimotor data, tied to specific object descriptions, and embedded in a probabilistic inference framework. Current issues include:

- **General Concept Representation:** Arbitrary objects can be represented by use of tensor spline groups, as well as shape modification processes as described by Leyton. Implementation will require careful attention. In addition, Leyton argued that wreath products could represent any concept; therefore, extensions need to be found for structural, material properties, social, and other types of relations.
- **Combining Sensor and Actuation Data:** A detailed characterization of how sensorimotor data is embedded in the wreath product representation is required. There

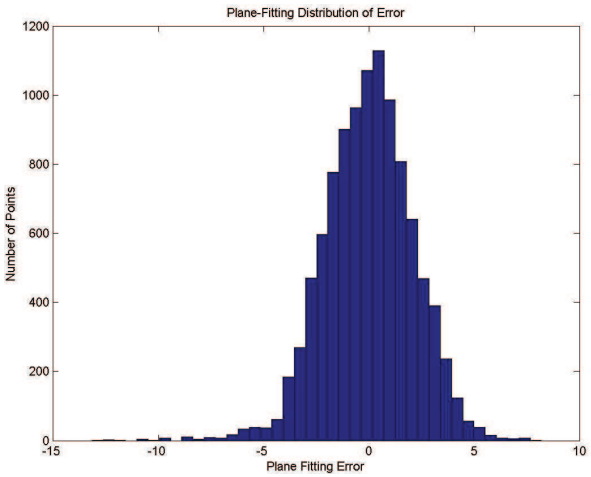

Figure 13: Error Distribution in Plane-Fitting for Kinect Data.

are limitations in what the actuators can achieve compared to what the sensors can perceive, depending on the robot used and its *Degrees of Freedom* (DoF). In the cube example, we mentioned that a robot with a head having 360° 3-DoF can *actuate the concepts* as motions. Similarly, a robot with a dextrous hand will be able to manipulate the object and be able to map the actuations to the concepts: e.g., in 2D the robot can draw the square, given the generative concept of a square, while in 3D it can trace its fingers along the the smooth faces of a cube and infer the object type based on the Bayesian network for a cube.

On the contrary, a two-wheeled robot with only a range sensor and no movable head or hands can only move on the ground, and will thus have limitations in relating actuations directly to concepts (especially 3D shape concepts).

- **Prior and Conditional Probabilities:** must be determined for the networks. This requires a rigorous learning process for the statistics of the environment and the sensors. For example, we are studying the error in fitting planes to Kinect data, and first results indicate that the noise appears Gaussian (see Figure 13).

- **Prime Factorization:** In the shapes we address, a Bayesian network can be created in multiple ways for the same shape; e.g., two square representations are $\Re \wr \mathcal{Z}_4$ and $\Re \wr \mathcal{Z}_2 \times \mathcal{Z}_2 \wr \mathcal{Z}_2$. The equivalence of resulting shape must be made known to the agent (even thought the generative mechanisms are different); we are exploring whether subgroups of the largest group generating the shape allow this to be identified.

- **Object Coherence and Segmentation:** Object segmentation is a major challenge,

and object classification processing will be more efficient if related points are segmented early on. We are looking at the use of local symmetries (e.g., color, texture, material properties, etc.) to achieve this. Moreover, object coherence can be found from motion of the object; namely, there will be a symmetry in the motion parameters for all parts of a rigid object which can be learned from experience.

References

[1] Thomas Binford, Tod Levitt, and Wallace Mann. Bayesian Inference in Model-Based Machine Vision. In *Uncertainty in Artificial Intelligence 3 Annual Conference on Uncertainty in Artificial Intelligence (UAI-87)*, pages 73–95, Amsterdam, NL, 1987. Elsevier Science.

[2] W.Y. Chang. Image Processing with Wreath Products. Master's thesis, Harvey Mudd College, Claremont, CA, May 2004.

[3] N. Fenton and M. Neil. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, Boca Raton, FL, 2013.

[4] R. Foote, G. Mirchandani, D. Rockmore, D. Healy, and T. Olson. A Wreath Product Group Approach to Signal and Image Processing: Part I – Multiresolution Analysis. *IEEE-T Signal Processing*, 48(1):102–132, 2000.

[5] R. Grupen and T.C. Henderson. Apparent Symmetries in Range Data. *Pattern Recognition Letters*, 7:107–111, 1988.

[6] T.C. Henderson, Y. Fan, A. Alford, E. Grant, and E. Cohen. Innate Theories as a Basis for Autonomous Mental Development. Technical Report UUCS-09-004, The University of Utah, October 2009.

[7] T.C. Henderson, Anshul Joshi, and Eddie Grant. From Sensorimotor Data to Concepts: The Role of Symmetry. Technical Report UUCS-12-012, University of Utah, School of Computing, Salt Lake City, UT, October 2012.

[8] T.C. Henderson, Hongchang Peng, Christopher Sikorski, Nikhil Deshpande, and Eddie Grant. Symmetry: A Basis for Sensorimotor Reconstruction. Technical Report UUCS-11-011, University of Utah, School of Computing, Salt Lake City, UT, May 2011.

[9] Thomas C. Henderson, Elaine Cohen, Edward Grant, M.T. Draelos, and Nikhil Deshpande. Symmetry as a Basis for Perceptual Fusion. In *Proceedings of the IEEE*

Conference on Multisensor Fusion and Integration for Intelligent Systems, Hamburg, Germany, 2012.

[10] M. Leyton. *A Generative Theory of Shape*. Springer, Berlin, 2001.

[11] D. Vernon, G. Metta, and G. Sandini. A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents. *IEEE Transactions on Evolutionary Computation*, 11(2):151–180, 2007.

[12] L. Zhang and Q. Ji. A Bayesian Network Model for Automatic and Interactive Image Segmentation. *IEEE Transactions on Image Processing*, 30(9):2582–2593, 2011.