Symmetry as a Basis for Perceptual Fusion

Thomas C. Henderson¹, Elaine Cohen¹, Anshul Joshi¹, Edward Grant², Mark Draelos², and Nikhil Deshpande²

Abstract—We propose that robot perception is enabled by means of a common sensorimotor semantics arising from a set of symmetry theories (expressed as symmetry detectors and parsers) embedded a priori in each robot. These theories inform the production of structural representations of sensorimotor processes, and these representations, in turn, permit perceptual fusion to broaden categories of activity. Although the specific knowledge required by a robot will depend on the particular application domain, there is a need for fundamental mechanisms which allow each individual robot to obtain the requisite knowledge. Current methods are too brittle and do not scale very well, and a new approach to perceptual knowledge representation is necessary. Our approach provides firm semantic grounding in the real world, provides for robust dynamic performance in real-time environments with a range of sensors and allows for communication of acquired knowledge in a broad community of other robots and agents, including humans. Our work focuses on symmetry based multisensor knowledge structuring in terms of: (1) symmetry detection in signals, and (2) symmetry parsing for knowledge structure, including structural bootstrapping and knowledge sharing. Operationally, the hypothesis is that group theoretic representations (G-Reps) inform cognitive activity. Our contributions here are to demonstrate symmetry detection and signal analysis and for 1D and 2D signals in a simple office environment; symmetry parsing based on these tokens is left for future work.

I. INTRODUCTION AND BACKGROUND

Physical robot systems have been steadily improving for many years now in terms of their capabilities, robustness, compliance, etc., and there is a strong push to introduce these systems into human environments as cooperative agents to assist people in their daily activities. A major roadblock to this goal is the lack of strong and robust cognitive abilities in robots, and more specifically inadequate knowledge acquisition, representation and manipulation. Robots need various kinds of knowledge to perform effectively in real applications, and the current approaches to providing that knowledge are to (1) have the robot learn from scratch, (2) spoon feed the knowledge by human programming, or (3) have robots share knowledge.

Our goal is to explore the use of symmetry analysis as a basis for the semantic grounding of multisensor sensorimotor affordance knowledge; this includes symmetry detection in signals, symmetry parsing in knowledge representation, and symmetry exploitation in perceptual fusion. We hypothesize that symmetry-based structuring of knowledge provides a more robust semantic basis than current methods, and in particular symmetry as applied to the acquisition of affordances from signals, representation for modeling actions and their effects, and exploitation in generative action discovery.

One form of knowledge of particular interest is self-knowledge about the robot’s own structure and capabilities: this includes sensors, actuators, kinematic and dynamic structures, energy consumption and replenishment, and computational capabilities (speed, space, parallel processing, signal processing, internet connectivity, etc.). This provides a basis for knowledge of affordances in the external world, i.e., the recognition of entities appropriate for the performance of a task. Finally, working knowledge is needed for the interactions between the robot and the environment for both physical actions and social interactions. Of course, a robot will also need to be able to understand and formulate goals and the plans necessary to achieve those goals, but we do not address this aspect of cognition here.

Hypothesis: We propose that robot affordance knowledge acquisition and perceptual fusion can be enabled by means of a common sensorimotor semantics which is provided by a set of symmetry theories embedded a priori in each robot. These theories inform perception, and thus the production of structural representations of sensorimotor processes, and these representations, in turn, permit perceptual fusion to broaden categories of activity.

Symmetry [28] plays a deep role in our understanding of the world in that it addresses key issues of invariance, and as noted by Viana [27]: “Symmetry provides a set of rules with which we may describe certain regularities among experimental objects.” Symmetry to us means an invariant, and by determining operators which leave certain aspects of state invariant, it is possible to either identify similar objects or to maintain specific constraints while performing other operations (e.g., move forward while maintaining a constant distance from a wall). Michael Leyton has described the exploitation of symmetry [11] and the use of group theory as a basis for cognition [12]. Our approach is motivated by Leyton’s work, but does not exploit the technical aspects of the wreath products used by Leyton.

Operationally, the hypothesis is that group theoretic representations (G-Reps) inform cognitive activity. We exploit symmetry-based signal analysis and concept formation in sensorimotor reconstruction and scene analysis. A schematic view of our symmetry-based affordance architecture (the Symmetry Engine) is given in Figure 1.

The two major research thrusts (see Figure 2) are:

1) Symmetry Detection: this involves the detection of

---

¹This work was supported by an NSF grant 10-21038.
²T.C. Henderson, Anshul Joshi and E. Cohen are with the School of Computing, University of Utah, SLC, UT, USA; tch at cs.utah.edu
³E. Grant, M. Draelos and N. Deshpande are with the Dept of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA; egrant at ncsu.edu
symmetry in signals, and this means the extraction of symmetry features (lines, color, surfaces, etc.), and symmetry axes. We are developing a wide range of methods, including signal processing and spline-based symmetry axis determination.

2) **Symmetry Parsing**: the individual symmetry elements must be composed into hierarchically structured representations. The production of hierarchical G-rep descriptions requires new analysis methods of the available symmetries. It would be useful if some form of prime factorization were possible in order to make comparisons and manipulation more efficient. Specific sensorimotor data for behaviors must be included in these descriptions. Methods to allow robust communication and sharing of the symmetry representation must be developed. This means finding a way to share the sensorimotor grounding of the concepts produced by the G-reps.

### II. SENSORIMOTOR RECONSTRUCTION IN 1D

Given a set of unknown sensors and actuators, sensorimotor reconstruction is achieved by exploiting relations between the sensor data and the actuator control data to determine sets of similar sensors, sets of similar actuators, necessary relations between them, as well as sensorimotor relations to the environment. Several authors have addressed this problem, and we propose here a principled approach that exploits various symmetries and that achieves more efficient and robust results. A theoretical position is defined, the approach shown more efficient than previous work, and experimental results given.

Early on, Pierce [23] described an approach to learning a model of the sensor set of an autonomous agent. Features are defined in terms of raw sensor data, and feature operators are defined which map features to features. The goal is to construct a perceptual system for this structure. One of the fundamental feature operators is the *grouping operator* which assigns features to a group if they are similar. This work was extended to spatio-visual exploration in a series of papers [14], [15], [23]. For a detailed critique of Pierce’s work, see [3]. Olsson extended this work in a number of papers [5], [6], [7], [8], [9], [20], [21]. He used information theoretic measures for sensorimotor reconstruction, and no innate knowledge of physical phenomena nor the sensors is assumed. Like Pierce, Olsson uses random movements to build the representation and learns the effect of actions on sensors to perform visually guided movements. The major contributions are the analysis of information theoretic measures and motion flow. O’Regan and Noé [22] use the term sensorimotor contingencies and give an algorithm which can determine the dimension of the space of the environment by “analyzing the laws that link motor outputs to sensor inputs”; their mathematical formulation is elegant.

A symmetry defines an invariant. The simplest invariant is identity. This can apply to an individual item, i.e., a thing is itself, or to a set of similar objects. In general, an invariant is defined by a transformation under which one object is mapped to another. Sensorimotor reconstruction can be more effectively achieved by finding such symmetry operators on the sensor and actuator data (see also [1], [4]).

Invariants are very useful things to recognize, and we propose that various types of invariant operators provide a basis for cognitive functions, and that it is also useful to have processes that attempt to discover invariance relations among sensorimotor data and subsequently processed versions of that data.

#### A. Symmetry Detection in Signals

Assume a set of sensors, \( S = \{ S_i, i = 1 \ldots n_S \} \) each of which produces a finite sequence of indexed sense data
values, \( S_{ij} \) where \( i \) gives the sensor index and \( j \) gives an ordinal temporal index, and a set of actuators, \( A = \{ A_i, i = 1 \ldots n_A \} \) each of which has a finite length associated control signal, \( A_{ij} \), where \( i \) is the actuator index and \( j \) is an ordinal temporal index of the control values.

We are interested in determining the similarity of sensorimotor signals. Thus, the type of each sensor as well as the relation to motor control actions play a role. It is quite possible that knowledge of the physical phenomenon that stimulates a sensor may also be exploited to help determine the structure of the sensor system and its relation to motor action and the environment.

We suppose that certain 1D signal classes are important and are known a priori to the agent (i.e., that there are processes for identifying signals of these types). The basic signals are:

- **zero**: \( y = 0 \) (at all samples)
- **constant**: \( y = a \) (for some fixed constant \( a \))
- **binary**: \( y \) takes on either the value 1 or 0
- **linear**: \( y = at + b \) (function of time index)
- **periodic**: has period \( P \) and the most significant Fourier coefficients \( C \)
- **Gaussian**: sample from Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \)

Thus, a first level symmetry is one that characterizes a single signal as belonging to one of these categories. Of course, composite signals can be constructed from these as well, e.g., the impulse signal is a non-zero constant for one step, followed by the zero signal.

Next, pairwise signal symmetries can exist between signals in the same class:

- **linear**
  - same line: \( a_1 = a_2, b_1 = b_2 \)
  - parallel: \( a_1 = a_2, b_1 \neq b_2 \)
  - intersect in point: rotation symmetry about intersection point
- **periodic**
  - same period
  - same Fourier coefficients
- **Gaussian**
  - same mean
  - same variance

### B. Sensorimotor Reconstruction

The sensorimotor reconstruction process consists of the following steps: (1) perform actuation command sequences, (2) record sensor data, (3) determine sensor equivalence classes, and (4) determine sensor-actuator relations. An additional criterion is to make this process as efficient as possible.

Olsson, Pierce and others produce sensor data by applying random values to the actuators for some preset amount of time, and record the sensor sequences, and then look for similarities in those sequences. This has several problems: (1) there is no guarantee that random movements will result in sensor data that characterizes similar sensors, (2) there is no known (predictable) relation between the actuation sequence and the sensor values, and (3) the simultaneous actuation of multiple actuators confuses the relationship between them and the sensors.

To better understand sensorimotor effects, a systems approach is helpful. That is, rather than giving random control sequences and trying to decipher what happens, it is more effective to hypothesize what the actuator is (given limited choices) and then provide control inputs for which the effects are known. Such hypotheses can be tested as part of the developmental process. The basic types of control that can be applied include: none, impulse, constant, step, linear, periodic, or other (e.g., random).

Next, consider sensors. Some may be time-dependent (e.g., energy level), while others may depend on the environment (e.g., range sensors). Thus, it may be possible to classify ideal (noiseless) sensors into time-dependent and time-independent by applying no actuation and looking to see which sensor signals are not constant (this assumes the spatial environment does not change). Therefore, it may be more useful to not actuate the system, and then classify sensors based on their variance properties. That is, in realistic (with noise) scenarios, it may be possible to group sensors without applying actuation at all.

Consider Pierce’s sensorimotor reconstruction process. If realistic noise models are included, the four types of sensors in his experiments (range, broken range, bearing and energy) can all be correctly grouped with no motion at all. (This assumes some energy loss occurs to run the sensors.) All this can be determined just using the equals symmetry operator (identity) and the means and variances of the sensor data sequences.

### C. Exploiting Actuation

Of course, actuation can help understand the structure of the sensorimotor system. For example, consider what can be determined by simply rotating a two-wheeled robot that has a set of 22 range sensors arranged equi-spaced on a circle. Assume that the control signal results in a slow rotation parallel to the plane of robot motion (i.e., each range sensor moves through a small angle to produce its next sample) and rotates more than \( 2\pi \) radians. Then each range sensor produces a data sequence that is a shifted version of each of the others – i.e., there is a translation symmetry (of periodic signals) between each pair. The general problem is then:

**General Symmetry Transform Discovery**

**Problem**: Given two sensors, \( S_1 \) and \( S_2 \), with data sequences \( T_1 \) and \( T_2 \), find a symmetry operator \( \sigma \) such that \( T_2 = \sigma(T_1) \).

Full details of the algorithms and methodology are given in [3]. Here we simply give the results for sensor grouping based on symmetries in the sensorimotor data (see Figure 3 for performance results in simulation). Sensor data sampling time was varied from 1 to 20 seconds for binary noise of 5%, 10% and 25%, and Gaussian variance values of 0.1, 1, and 10. Ten trials were run for each case and the means are shown in the figure. As can be seen, perfect sensor grouping is achieved after 20 seconds without any actuation
cost. Previous methods required driving both wheels for a longer time and they cost about $30k_a/s$ more in energy than our method ($k_a/s$ is the actuation to sensing cost ratio).

Fig. 3. Grouping Correctness vs. Number of Samples; left to right are for binary salt and pepper noise of 5%, 10%, and 25%; curves for 0.1 (leftmost of 3), 1.0 (middle of 3), and 10.0 (rightmost of 3) variance are given in each plot and time in 0.1 second units.

1) Sensor Grouping (Actuated): Given a set of sensors that characterize the group operation nature of an actuator (in this case rotation), the sensors can be grouped based on the fact that similar sensors produce data that has a translation symmetry along the temporal axis. Figure 4 shows representative data for the range and compass sensors. The simple determination of a translation symmetry between signals allows both grouping (i.e., the signals match well at some time offset), and the angular difference between the sensors (given by the $t_{offset}$ at which the symmetry occurs); $t_{offset}$ is proportional to the angle between the the sensors in terms of actuation units. Figure 5 shows the perfect grouping result with noise of 1% in the compass sensor data and 0.1 variance in the range sensor data (the figure shows a 29x29 similarity matrix where white indicates sensors are in same group, and black indicates that are not).

We have performed physical experiments with physical sensors to validate the proposed approach. Data was taken for both the static case (no actuation) and the actuated case (camera rotation).

Two sensors were used in this unactuated experiment: a camera and a microphone. The camera was set up in an office and a sequence of 200 images was taken at a 10Hz rate. Figure 6 shows one of these images. The 25x25 center set of pixels from the image comprise a set of 625 pixel signals each of length 200. An example trace and its histogram are given in Figure 7. As can be seen, this is qualitatively a Gaussian sample. Figure 8 shows a 200 sequence signal of microphone data, and its histogram which also looks Gaussian.

The application of our symmetry detectors classified all pixel and microphone signals as Gaussian signals, and grouped the pixel signals separately from the microphone due to the difference in their variance properties.

We also took a set of images in an actuated experiment by rotating the camera by one degree for 360 degrees. Domain translation symmetry allows the identification of all the pixel

Fig. 5. Grouping Matrix: 29 × 29 binary matrix; sensors 1-24 are range sensors (sensor 21 returns constant value); 25 is energy; 26-29 are compass sensors.

Fig. 6. One of the 200 Static Images.
Next we turn to the detection of symmetries in 2D images; much work has been done in this area (e.g., see [13] for a survey, and [2], [24] for more related work). We give a method here to detect the following symmetries on pixel sets:

- $C_1$: only the identity map is invariant.
- $C_n$: rotation symmetry about a fixed point with rotations of $\frac{360}{n}$.
- $D_1$: single axis of bilateral symmetry.
- $D_n$: rotation and reflective symmetries (rotational of $\frac{360}{n}$ and $n$ axis of reflective symmetry).

- $O(2)$: orthogonal group of continuous rotations about a fixed point and any line through that point is a reflective axis of symmetry.

The Planar Reflective Symmetry Transform is defined by Podolak et al. [24], and we use that to measure the reflective symmetry about a line at a specific point (see Figure 9 for the PRST values for a vertical standing ellipse). Generally, this is an expensive operation, so we have developed a way to limit the amount of computation. In order to more efficiently locate axes of reflective symmetry, we exploit the properties of the frieze expansion pattern (FEP); for more on this, see [10]. For example, Figure 10 shows the ellipse image (a) and its FEP (b). We extract the upper curve of the FEP (c) and locate maxima and minima (an axis of symmetry must pass through a min or max on this curve). This restricts the number of reflective lines for which the PRST must be computed. Figure 10(d) shows the set of reflective axes found for the ellipse. We also use the same PRST measure for rotational symmetry, where instead of flipping the image as required for reflective symmetry, we rotate the image based on aligning peaks and valleys in the FEP curve (e). All these symmetry axes are combined (f) to determine that this shape has $C_2$ symmetry. Figure 11 shows the same analysis for a leaf image. The resulting group classification is $D_1$ as there is one reflective axis of symmetry and no rotational symmetry.

**III. Symmetry Detection in 2D**

signals along a row as similar to each other (i.e., they are all in the plane of the rotation). Due to the translation amount, the offset between the signals is also discovered.

**IV. CONCLUSIONS AND FUTURE WORK**

We demonstrate symmetry theory as a basis for sensorimotor reconstruction in embodied cognitive agents and have shown that this allows the identification of structure with simple and elegant algorithms which are very efficient. The exploitation of noise structure in the sensors allows unactuated grouping of the sensors, and a simple one actuator rotation permits the recovery of the spatial arrangement of the sensors. This method was shown to hold for physical sensors as well. In addition, we describe a more efficient
approach to using the PRST for symmetry group detection in image shapes based on using properties of the Frieze Expansion Pattern. In future work, we intend to show how such tokens (i.e., groups) can be parsed into higher level descriptions of sensorimotor streams. This includes 3-D surface points, homogeneous 2-D surfaces (e.g., planes), and 3-D surface normals may all serve as a basic symmetry element for G-rep structuring. Data from Kinect or other range sensors along with an oct-tree organization can form the basis of this. For example, the Create robot shown in the left side (a) of Figure 12 has a Kinect sensor mounted on it. A sample image of an office scene is taken and range data taken with the Kinect is used to discover group symmetries for the scene; Figure 12 (b) shows a patch of a cone-shaped region (part of a waste basket), and the group expression for that patch.

![Ellipse Image](image1)

(a) Ellipse Image

(b) Frieze Expansion Pattern

(c) FEP Curve with Min/Max Indicated

(d) Reflective Axis.

(e) Rotation Directions.

(f) Reflective Axes and Rotation Directions.

Fig. 10. Symmetry Analysis for an Ellipse. (a) Original image. (b) FEP for Ellipse. (c) FEP curve with Max/Min Marked. (d) Reflective Axes. (e) Rotation Symmetry Directions. (f) All Symmetry Axes and Rotation Directions.

![Leaf Image](image2)

(a) Leaf Image

(b) Frieze Expansion Pattern

(c) FEP Curve with Min/Max Indicated

(d) Reflective Axis.

(e) Rotation Directions (None).

(f) Reflective Axes and Rotation Directions (None).

Fig. 11. Symmetry Analysis for a Leaf. (a) Original image. (b) FEP for Leaf. (c) FEP curve with Max/Min Marked. (d) Reflective Axes. (e) Rotation Symmetry Directions (None). (f) One Symmetry Axis and Rotation Directions (None).

Although generally not explicit in sensor data, symmetry axes are also important cognitive features. The medial axis gives the morphology of a 3D object and can be used to determine the intrinsic geometry (thickness) of both 2D and 3D shapes. Since it is lower dimensional than the object, it can be used to determine both symmetry and asymmetry of objects. In previous work we have obtained results on tracking the distance between a moving point and a planar spline shape and computed planar Voronoi diagrams between and within planar NURBS curves[26], [25] (see Figure 13), and found methods that allow us to characterize the correct topology as well as shape of the planar and 3D medial axis. We developed an approach that used mathematical singularity theory to compute all ridges on B-spline bounded surfaces of sufficient smoothness[17], and then extended the results to spline surfaces of deficient smoothness[16] and to compute ridges of isosurfaces of volume data[19].

Most recently we have extended that approach to compute the interior medial axis of regions in $R^3$ bounded by tensor product parametric B-spline surfaces[18]. The generic structure of the 3D medial axis is a set of smooth surfaces along with a singular set consisting of edge curves, branch curves, fin points and six junction points. In this work, the medial axis singular set (the set of transition points) is first computed directly from the B-spline representation using a collection of robust higher order techniques. Medial axis surfaces are computed as a time trace of the evolving self-intersection set of the boundary under the eikonal (grassfire) flow. The eikonal flow results in special transition points that create, modify or annihilate evolving curve fronts of the (self-) intersection set. The transition points are explicitly identified using the B-spline representation without ever actually computing the eikonal flow. Evolution of the self-intersection sets (the sheets and junction curves) are computed creating theoretically derived evolution vector fields that support accurate tracing of these entities. The algorithm is the first to accurately compute connected surfaces of the medial axis as well its singular set and its topological structure (see Figure 13). The algorithm can be simplified to compute accurate medial axes of planar NURBS curves, as well as the distance to boundary from each medial axis point. Thus, the topological and metrical symmetries can be

![a. Kinect on Create Robot](image3)

b. Wreath Product for Cone

Fig. 12. a. Create robot roaming and taking Kinect data; b. Data taken from wastebasket and represented as wreath product.
determined, and we intend to study how this can be used to segment objects of interest in a scene.

REFERENCES


