Behavior-Grounded Object Identification, Grouping and Ordering by a Humanoid Robot

Connor Schenck, Jivko Sinapov, Kerrick Staley
Developmental Robotics Laboratory
Iowa State University
{cschenck, jsinapov, kerrick}@iastate.edu

***Abstract*—From an early stage in development, infants show a profound drive to explore the objects around them. Research in psychology has shown that in doing so, they solve a vast array of problems, including the formation and establishment of object representations, recognition of objects based on the stimuli they produce, object grouping and ordering, as well as learning words that describe objects and their properties. This proposal introduces a behavior-grounded framework for object perception that will enable a robot to solve these very same problems.**

I. Introduction

Our ability to explore physical objects is unparalleled in the natural world. From an early age, human beings spend much of their time manipulating objects while simultaneously observing the resulting stimuli (e.g., visual movement, auditory events, etc.). A long line of research in psychology has revealed that humans (as well as animals) acquire information about objects through the use of a number of manipulation behaviors, commonly referred to as *exploratory procedures* [23] or *exploratory behaviors* [9], [33]. For example, scratching an object can inform us of its roughness, while lifting it can inform us of its weight. In a sense, the exploratory behavior acts as a "question" to the object, which is subsequently "answered" by the sensory stimuli produced during the execution of the behavior.

Other research in psychology has established that the sensory feedback produced by objects can be crucial for solving several key tasks:

1) *object identification*, i.e., the ability to individuate objects, recognize the object identity of a given object stimulus, and recognize when a stimulus is produced by a novel object [19], [17]..
2) *object sorting*, i.e., the ability to spontaneously group items into sets, or orders, without given a specific criteria [48], [32].
3) *category and relational learning*, i.e., the ability to assign category membership to novel objects as well as infer how two objects should be ordered, based on a criteria specified by a series of example objects with known labels and/or orderings [2].

The goal of this proposal is the development of a multi-modal behavior-grounded framework for object perception that would enable a robot to solve these problems in an experimental setting. To achieve this aim, the robot in our framework actively performs exploratory behaviors (e.g., grasping, lifting, shaking, dropping, pushing and tapping) when learning about objects as opposed to just passively observing them. While most robots perceive objects using vision alone, the robot in our framework also uses the auditory, proprioceptive and tactile sensory modalities, which are necessary to capture many object properties [6], [26].

The rest of the proposal is organized as follows: Section II gives an overview of the related work in psychology and robotics. Section III provides a detailed description of the three main problems this proposal addresses, along with the approaches that will be used to solve them. Section IV describes our robot platform and the experimental design used to test our solutions. Finally, the last section provides extra information about the software libraries that will be used, the team members, and the timeline for the project.

II. Related Work

A. Psychology and Cognitive Science

The ability of humans to individuate objects and recognize their identities has been extensively studied in psychology. The problem of object identification is typically defined as that of inferring how many objects the environment contains (also referred to as individuation) as well as recognizing when the same object is encountered twice (sometimes referred to as identification as well as recognition) [17]. Studies in developmental psychology have shown that this process is fundamental to establishing an internal object representation that can handle the large number of objects that humans encounter in their day to day lives [52], [19].

For this reason, how infants establish an object representation and subsequently use it to recognize the identities of objects is a question of significant interest to developmental psychology. For example, a study in infants showed that even at the age of 12-months, humans are able to individuate objects using both shape and color information [52]. The study also found that while both object features were used for the task of figuring out how many objects exist, only the shape feature was used when recognizing the identity of an object that was previously individuated. Other studies have shown that when identifying objects, infants and adults often make different judgments based on the differences in the objects' features [54], indicating that at such an early age, the biological circuits that allow the problem to be solved are still developing.

In a typical scenario, the human participant observes (or interacts with) objects one at a time, where the next object may or may not be a previously encountered one. Subsequently, participants may be asked to enumerate the objects they observed, or match an object stimulus to one of the estimated object identities. For example, one such study with human adults showed that as the number of objects observed increases, the likelihood that a novel object will be classified as a previously observed object goes down [17]. The same study also found that humans rely on prior information when solving identification problems - based on this finding, the robot's object identification model will also be evaluated when prior information about objects with known identities is available.

A closely related area of developmental psychology studies how infants group objects. An important finding is that certain experimental settings can elicit spontaneous sorting and grouping behaviors by infants [30], [48]. Starkley [48] reports that both 9 and 12 month-old infants exhibit sorting behaviors when presented with a set of 8 objects, where the set contains 2 groups of four objects that are similar along some dimension (e.g., size, color, etc.).

Sorting and grouping behaviors have also been observed with non-human primates [32], [46]. For example, Spinozzi *et al.* [46] found that human-encultured Bonobos and Chimpanzees are capable of spontaneously partitioning a set of objects into two categories. The authors also report that chimpanzees' predominant means of partitioning a set of objects is by manipulating objects from one object class only. This procedure is consistent with the behavior of 3 year old infants [46]. Overall, these findings suggest that the ability to sort objects is fundamental to primate intelligence.

For humans in particular, object grouping skills are thought to be fundamental for language acquisition – for example, Nelson argued that children form primitive conceptual categories which are later used when binding the meaning of a word [30]. Similarly, based on a large volume of experimental research, Bloom argues that a large part of early language learning is about establishing a relation that maps language symbols (e.g., individual nouns) to already existing concepts that are formed independently of the language in question [4]. An example of what this may look like is provided by Kemp *et al.* [16] who write:

> *"Before learning her first few words, a child may already have formed a category that includes creatures like the furry pet kept by her parents; and learning the word 'cat' may be a matter of attaching a new label to this preexisting category." [16, p. 216]*

Not surprisingly, a large volume of research has focused on revealing how humans learn the names of categories [2]. In this framework, the participants are typically presented with several examples from each object category and subsequently asked to categorize a novel item. Researches postulate that humans use two different strategies (sometimes in combination) to learn categories from examples - the first involves finding the common features of members of an individual category, while the second consists of identifying the distinctive features among the non-members of that category [14], [13]. Experiments have shown that adults can learn categories even when presented only with pairs of objects of different categories [14]. Children between the ages of 6-9 years old, however, could only learn the same categories when provided with object pairs in which the two objects are of the same category class, indicating that the two strategies for solving the task have different developmental trajectories [14].

In addition to learning discrete categories, researchers have also examined how adult and infant humans learn real-valued comparative relations such as "A is bigger than B" [44], [8]. As with category learning, humans can learn such relations when presented with paired examples for which the relation is provided by the instructor or inferred by some other means. Hence, the robot in this work will be tested in a similar fashion – after initially interacting with the objects, computational models will be evaluated using both discrete categorization as well as real-valued ordering tasks.

B. Robotics

Traditionally, most object recognition systems used by robots have relied heavily on computer vision techniques [34], [47], [35] and/or 3D laser scan data [38]. But studies in psychology indicate that not only is there a link between neural activations and different sensory inputs for the same object in the brain [1], but that often multiple senses are necessary to correctly recognize an object. In a study by Sapp *et al.*, toddlers were presented with sponges painted as rocks and only by grasping the sponges could they realize that they were being deceived [39]. Other studies involving proprioception or audition have also shown that not only is it possible to use sensory modalities other than vision to recognize objects and their properties, but in some cases it is necessary [15], [7], [10], [11].

Recently, there have been multiple studies in robotics that have focused on object recognition using sensory modalities other than vision or 3D laser scan data. A study by Natale *et al.* [29] showed that proprioceptive information obtained by grasping an object can be used to successfully recognize objects. Other studies have estimated physical parameters of objects from proprioceptive data [21], [22], which can be used to recognize objects. A study by Bergquist *et al.* [3] showed that a robot can use proprioceptive information alone to recognize an object from a large set of objects. A study by Sinapov *et al.* [43] showed a similar result using auditory information alone. Other studies have confirmed that audition can be used for object recognition [37], [36] as well as for determining properties of objects [20]. Another study by Metta *et al.* showed that integrating proprioception and vision can bootstrap a robot's ability to manipulate objects. All of these studies strongly imply that sensory modalities other than vision (e.g. audition, proprioception) are useful for object recognition in addition to vision. This research will take advantage of this by combing multiple sensory modalities in order for the robot to perform tasks.

One of the major drawbacks of virtually all of the methods cited above is that during the training stage, the robot has to be told which object it is exploring at any given trial. In other words, the training trials must be grouped by object identity. In order to relax that assumption, a robot must be able to autonomously figure out how many objects it has interacted with as well as organize its sensorimotor data according to object ID (i.e., solve the object individuation problem). There has been relatively little work in robotics in that area - a study by Modayil and Kuipers [27] showed how a robot could use data gathered from a laser range finder to build an ontology of objects. Another study by Southey and Little [45] used a stereo camera to detect depth features in the robot's environment, which were combined based off 3D movement patterns to create representations of each object in the environment.

In addition to object recognition, there has been much work in robotics studying how robots can form object categories in an unsupervised manner. Some of them have focused on how robots can estimate similarity between objects and use that similarity to develop meaningful object categories [31], [29], [28], [51], [43], [50]. In [29] a Self-Organizing Map was used to illustrate the haptic similarities between objects, while [43] showed that a robot can use auditory data generated from performing multiple behaviors on an object to estimate similarities.

Griffith *et al.* [12] showed that a robot can form categories of "container" and "non-container" by observing the movement of an object dropped in the vicinity of another object. Sinapov and Stoytchev [42] showed that a robot can use these object similarities to detect which object in a set of objects is the odd one out. While all of these studies showed how a robot can group objects in an unsupervised manner, they all suffer from one main drawback: They all require the type of sorting to be specified in advance - for example, in [12], the robot's categorization model used the X-means algorithm, which can find clusters in data, but not orders or hierarchies. In [41], on the other hand, the categorization algorithm assumed that the objects can be organized in a hierarchy, as opposed to some other structure. The research project that we propose plans to implement methods such as the one described in [18] to allow the robot to determine which structure type should be used to organize a particular set of objects – in other words, the structure used to sort the object is induced by the model, rather then specified by the programmer.

Supervised learning for object category classification has also been studied in robotics, though not as extensively as identification. A study by Lopes and Chauhan [25] had a robot use vision to extract features from an object. They then used a set of classifiers to classify each object into different categories specified by a human. A study by Sinapov and Stoytchev [40] showed how a robot can use proprioceptive and auditory feedback to classify objects into six human-labeled categories.

Other studies have examined relations among objects. The study by Griffith *et al.* [12] examined the relationship between objects dropped in the vicinity of a container/non-container, and how the two objects moved when the robot interacted with them. This research will present methods for categorizing objects into pre-defined categories and learning relations between objects as they relate to ordering objects (e.g. bigger than). To the authors' knowledge, there has been no previous research in robotics on ordering objects.

III. Experimental Platform

A. Robot and Sensors

We will use an upper torso humanoid robot, which has as its actuators two 7-DOF Barrett WAMs, each with an attached 3-finger BarrettHand. The WAMs have built-in proprioception that measures joint angle and torque at 500 Hz; auditory feedback is captured by an Audio-Technica U853AW cardioid microphone mounted in the head, which samples 1 channel (mono) at the standard 16-bit/44.1 kHz resolution and rate. A digital accelerometer device [49], mounted on one of its fingertips, samples acceleration of the fingertip at 1600 Hz, allowing detection of minute vibrations due to rubbing between the robot's fingertip and the objects' surfaces. Vision is supplied by a ZCam, which is a color+depth camera from 3DV systems that records standard 640×480 RGB video in addition to 320×240 depth images accurate to within 1-2 cm.

B. Objects

In this project, the robot will explore 100 different household objects. To our knowledge, this is the largest number of objects explored by a robot for this type of study. The total object set will consist of 20 individual object sets (of 5 objects each) such that each set corresponds to an object category, within which the objects vary along 1 or more dimensions. Some of the object categories include:

- decorative eggs of varying material
- small inflated rubber balls of varying color and material
- pink water noodles of varying length
- small oval-shaped cups of varying color
- tupperware containers of varying contents
- styrofoam cones of varying size
- cans of food of varying size and weight
- medicine bottles of varying color, contents and weight
- soda cans of varying sizes

Fig. 1. Example object sets that will be used in our experiments.

- plastic bottles of varying color and material
- PVC pipe pieces of varying diameter
- cups of varying material and color
- boxes of pasta of varying color, contents, size, and weight
- metal objects of varying color, shape, and size

Figure 1 shows some of the object sets that will be used in our experiments.

C. Behaviors

The robot in our experiments will perform 10 different behaviors on the objects. Some of the behaviors will require that the object's location is detected on the table, which is done in the following way:

1) A background model will be created by taking a snapshot of the empty table before any objects are placed on it.
2) When an object is in place and the robot needs to determine its position, the robot will move its hand out of its field of view, calculate the deviation of each pixel observed from the value

predicted by the background model, and then use a threshold to classify them as either "background" or "foreground".

3) The "foreground" areas will then be grouped, and the largest group will be deemed the object; the shape representing the object will then be surrounded with a rectangle. An example is shown in Figure 2.
4) The pixel coordinates of the lower left corner of this rectangle will be provided to a 3-nearest neighbors algorithm, which will interpolate the appropriate hand position from a set of pretrained positions.

The "Look" behavior will be the only non-interactive behavior; for this behavior, visual modalities will be later analyzed (color, depth). For the other behaviors, proprioceptive, auditory, and vibrotactile data will be recorded.

1) Look: The robot will record a sequence of images from its cameras (both RGB and depth) which will be used to extract visual object features

2) Grasp: The aforementioned algorithm will be used to position the arm near the object, and the gripper will be closed.

3) Slow Lift: After grasping, the object will be lifted with little acceleration.

4) Fast Lift: After grasping, the object will be lifted with sharp acceleration.

5) Shake: After lifting, the object will be shaked by applying torque on the wrist and elbow joints.

6) Drop: After lifting, the object will be positioned 2.5 cm off the table and released; the motors will be moved only slightly to minimize noise from the fingers.

7) Tap: The aforementioned algorithm will be used to position the hand in a "scooping" position next to the object, and the hand will hit the object.

8) Crush: The object will be positioned in a fixed place, and the hand will move in a fixed trajectory down onto it. Once the object is struck, and once the torque values reach a threshold, the robot will back its hand up.

9) Poke: The aforementioned algorithm will be used to position the hand in a position similar to the "Tap" position, except that all but one finger will be curled up. The finger will hit the object, moving faster than during the "Tap" behavior.

10) Push: The aforementioned algorithm will be used to position the hand in a "clapping" position next to the object, and the hand will push the object with the palm rather than the fingers.

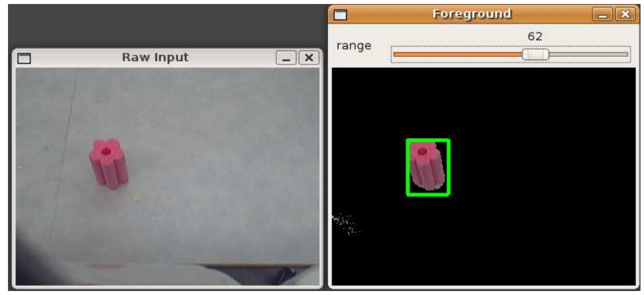

Fig. 2. Illustration of object detection in the robot's visual field of view.

Over the course of each behavior execution, the robot will record the sensory feedback from the following sensors: joint-torque sensors in the motors, microphones in the head, RGB images from the ZCam, Depth images from the ZCam and vibrotactile feedback from the robot's fingertip accelerometer. Each behavior will be executed on each object a minimun of 5 times.

IV. Theoretical Model

A. Notation

Let $\mathcal{B}$ be the set of exploratory behaviors and let $\mathcal{S}$ be the set of sensory modalities available to the robot. Let $\mathcal{C}$ be a set of behavior-modality contexts such that each context $c_j \in \mathcal{C}$ refers to a unique combination of a behavior and a sensory modality (e.g., *drop-audio*). Note that it is not necessary for every combination to be present in the set $\mathcal{C}$, since in our case certain behaviors do not produce sensations in certain modalities.

During each object exploration trial, the robot is presented with an object $o \in \mathcal{O}$, the set of all objects, and subsequently applies its set of exploratory behaviors on the object. Hence, when executing behavior $b \in \mathcal{B}$, the robot observes a set of sensory signals $\mathcal{X}_b = \{x_1 \dots x_{m_b}\}$ where each x_j represents the sensory feedback observed from some known sensory modality in $\mathcal{S}$. Note that the number of sensory feedback signals detected when performing some specific behavior, $|\mathcal{X}_b| = m_b$, may be less than the number of sensory modalities, $|\mathcal{S}|$, since certain behaviors do not produce sensations in certain modalities (e.g., looking at an object does not produce tactile sensations).

After all behaviors are applied on the test object, the i^{th} exploration trial may be summarized by the collection of observed sensory feedback signals, $T_i = \{X_b\}_{b \in \mathcal{B}}$. In practice, the signals x_j may be encoded as numerical vectors, real-valued time series, or discrete sequences. For this project, several different representations will be used, including sequences

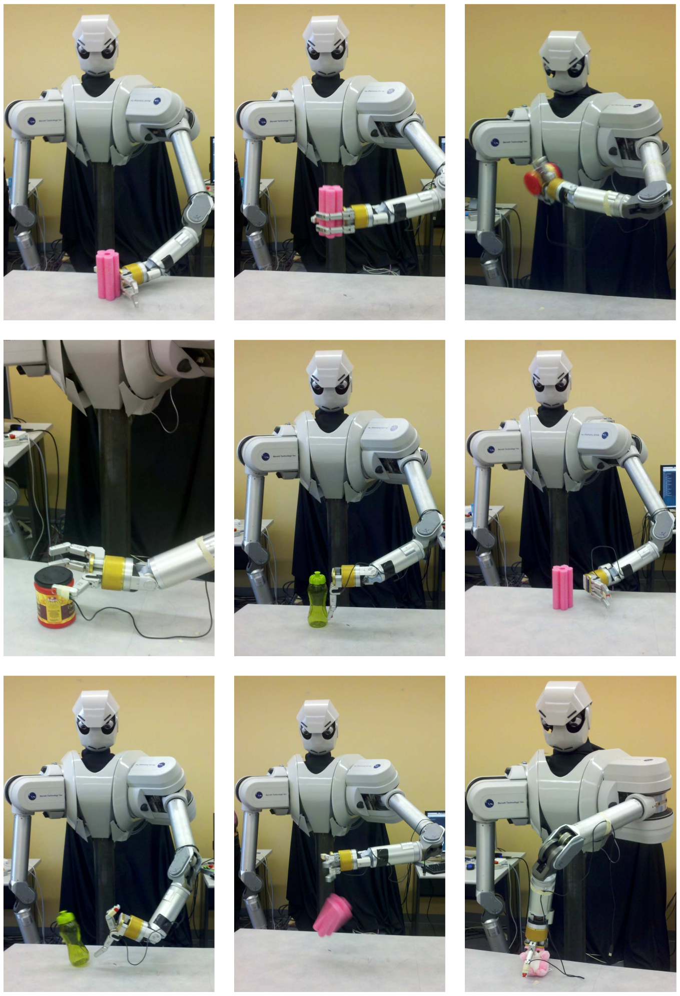

Fig. 3. The exploratory behaviors (excluding the “look” behavior) that the robot performs on objects.

B. Object Identification

1) Problem Formulation: Object identification is the problem of deciding how many objects have been perceived, as well as establishing what psychologists refer to an object representation (OR) [17]. While there is no clear agreement about what exactly an OR encodes, it has been theorized that at the very least, an OR must contain a mapping between *object tokens* (i.e., individual experiences with an object) to *object identities* (i.e., a unique identifier of the object) [17]. In addition, the object identification problem cannot be considered fully solved, unless a novel object can be recognized as being novel.

In the scenario proposed in this paper, the robot is presented with a set of objects $\mathcal{O}$, such that each object is presented one at time to the robot. The objects are explored in an order such that after performing one exploration trial with an object, the object is switched for the next one and so forth. This process is repeated n times, such that in the end, the robot has explored each object in n different trials. Hence, each object identification task can simply be represented by the set of trials $\mathcal{T} = T_1, \ldots T_{n\times|\mathcal{O}|}$.

The robot in this setting has no information regarding which object was present at trial i or how many objects were present in the object set $|O|$. Hence, to solve the problem, the robot's object identification model must be capable of findings how many objects there were, enumerate them using object IDs, and associate each of the trials in the set $\mathcal{T}$ with a specific object identity. Thus, this model would allow a robot to group its sensory motor experiences into sets such that each set corresponds to experience with a unique object. This is highly desirable since variants of such representations have been used for a wide variety of tasks, but always require a human programmer to specify the object identity of a given trial or sensorimotor experience with the object.

2) Approach and Evaluation: Given a sensorimotor data for a set of trials $\mathcal{T}$, a possible solution to solve the problem is to use a clustering algorithm to partition the trials into individual sets, such that each set corresponds to a particular object. Popular clustering methods include k-Means, as well as graph-based algorithms such as spectral clustering. While such an approach might succeed on the task, these algorithms suffer from the drawback that they require careful parameter selection. For example, k-Means requires that the programmer specify the desired number of clusters (in our case, object identities). The choice of these parameters often determines how coarse the resulting partitioning is.

This implies that if the objects are always explored one at a time, there is no guarantee that a particular clustering algorithm will solve the task. One way to overcome this problem is to give the robot's model some prior information regarding experience with objects with known identities. In fact, there is direct evidence from psychological experiments that humans also need such prior experience [17]. Hence, the second approach we consider will be one that is trained on a certain object set with known identities $\mathcal{O}_{train}$ while the subsequent individuation model is evaluated on set of objects $\mathcal{O}_{test}$ whose identities are not known. An initial implementation for such a model will consist of an algorithm which estimates the best parameters of a given clustering algorithm for solving the task from the training data. Other approaches will be considered as well.

The object individuation model will be evaluated based on how close the resulting partitioning of the trials matches the way they would be partitioned when grouped into sets based on the true object identity. Since both, the model's output, and the true individuation, are represented by two clusterings, we will use metrics designed to compute the mutual information between two clusterings (e.g., the ones used in [53]). The object recognition model, on the other hand, will be evaluated on how well it can match the sensory feedback from individual trials to the identity of a known object that was previously experience (typically measured in percent accuracy). Finally, the recognition model from previous work will be improved so that it can estimate when the object in the interaction is a novel object, not previously explored by the robot – this will be done by measuring how uncertain the recognition model is given the sensory feedback produced by the object. As with all subsequent tasks, success will be determined by whether the performance of the models (measured by whichever metrics apply) is substantially better than what a chance model would produce.

C. Spontaneous Object Sorting

1) Problem Formulation: In a typical psychological experiment, the participant is presented with a set of objects and then either asked to group them or allowed to freely explore them to see if spontaneous sorting behavior occurs. Hence, the task in this setting is to learn a model which, given a set of object identities and some amount of sensorimotor experience with the objects, outputs one (or possibly more) ways that the objects may be grouped, or ordered. More specifically, the algorithm has to output a structure, S which can either represent a discrete categorization (e.g., finding two clusters in the set of objects), a continuous ordering (e.g., if the objects vary along weight), or even a grid-like structure (e.g., if the objects vary in visual size as well as weight).

2) Approach and Evaluation: While there are many algorithms and frameworks for unsupervised object categorization, virtually all of them require that the programmer specify the structure (e.g., a discrete clustering vs. a hierarchy vs. a grid) in advance. Given a set of objects, however, the robot cannot automatically know which structure is appropriate - instead the structure has to be induced from the sensorimotor feedback experienced with the objects. To solve this problem, we will use the probabilistic model of Kemp *et al.* [18] which is capable of estimating what type of structure best explains the data, as well as finding the instance of that structure. For example, given a set of 5 black and 5 white balls, the model would (ideally) discover that the *partition* structure is most optimal and that the specific instance involves two clusters, one for each color. On the other hand, if the 10 objects were identical in all respects except for weight, the model should find that the *order* structure is the best fit.

The input to the structure discovery model consists of a similarity relation that specifies how similar two objects are. In our case, the robot will estimate such a relation for each behavior-modality combination, and then attempt to fit the best structure for each one. The different ways in which the objects may be sorted may subsequently be visualized. The model will be tested on sets of objects which vary along one or more given dimensions (e.g., shape) while remaining constant on others (e.g, color, weight, etc.). If the sorting produced by the model captures the variation among the objects, than the sorting will be considered meaningful. In addition, once the model fits a structure for each sensory-modality context, these structures can then be used to extract features for each object – for example, given a partition structure with 2 clusters, the cluster membership for an object may be used as an input feature for an object classification task, which is discussed in the next subsection.

D. Category and Relational Learning

1) Problem Formulation: The final task consists of training the robot to recognize the category labels of objects given a certain amount of objects with known labels. For example, if the robot interacts with a large set of objects, and if the programmer specifies that two of those objects are called "cups", then the robot's model should be able to infer what other objects are cups as well. Similarly, given a comparative relation (e.g., A bigger than B), and a few example objects for which the relation is known, the robot's model should be able to learn a ranking rule which can compare novel objects according to the same rule, given some sensorimotor experience with them.

2) Approach and Evaluation: For the category recognition task, we will consider up to 20 objects categories, where each one consists of five objects. Some of the categories group objects based on material (e.g., the metal objects category), while others group them based on shape (e.g., balls). Some are based on common names that humans assign to them (e.g., pop cans category). To recognize the category of an object, a classifier approach will be used, such that the classifier established a mapping between object features (extracted over the course of interaction with the objects) and object category label. In this setting, each object is represented as a feature vector which will be generated by looking at how the objects are sorted in various sensorimotor contexts (as described in the previous section). Alternatively, graph-based learning methods (such as the one used in [40]) will also be explored, which directly exploit the estimated object similarity in a given behavior-modality combination. As done in prior work, the category recognition accuracy will be compared for different machine learning algorithms (e.g., k-Nearest Neighbors, Suppor Vector Machine, etc.).

For the comparative relational learning task, the task of the model is to determine the order for a pair of objects, given features used to represent the objects. Similarly, this model will be trained with example objects for which the relation is known, and will be tested on objects for which the relation is unknown. Several relations will be considered for evaluation, including size relation (bigger then), and weight relation (heavier then). In machine learning, several algorithms have been proposed to solve this task, most of which are designed to rank the order of documents such as web pages. For this project, we plan to test any available standard machine learning methods, as well as develop ordering models that use regression (i.e., mapping from features to a real valued number) to order objects.

V. Appendix

A. Team

1) **Kerrick Staley** is a first-year student in Computer Engineering. He is interested, in general, in computer science, mathematics, and the physical sciences; he has specific interests in robotics, cryptography and data security, user interface design, and the practicalization of open source software. He programs primarily in C/C++ and Python. He enjoys reading Slashdot.org, and his Kirby skills in SSB64 will stomp most competitors. He has a website with further biographical details at kerrickstaley.com.
2) **Connor Schenck** is a senior in Computer Science. He has experience with C/C++, Java, and Matlab. He has used OpenCV, Weka, Java Swing, and MATLAB's Image Processing Toolkit. He has taken courses on Machine Learning, Artificial Intelligence, Algorithms, and Statistics. He is a coauthor for the paper *Interactive Object Recognition Using Proprioceptive Feedback* and *Interactive Object Recognition Using Proprioceptive and Auditory Feedback*. He has also worked on multiple projects in the Developmental Robotics Laboratory at Iowa State University.

3) **Jivko Sinapov** received the B.S. degree in Computer Science from the University of Rochester, NY in 2005. He is currently a PhD student in Computer Science and works at the Developmental Robotics Laboratory at Iowa State University, Ames. His research interests include developmental robotics, robotic perception, manipulation, and machine learning.

B. Software Packages

We anticipate that the following list of software libraries will be used in for this project:

1) **The WEKA Java Machine Learning Library :** contains a number of implementations for popular machine learning algorithms for the tasks of classification, and unsupervised clustering [55].
2) **Structural Form Discovery MATLAB package:** implementation of the model proposed by Kemp *et al.* [18] for the purposes of fitting structures to data.
3) **OpenCV:** C++ computer vision library, used when detecting the object on the table, as well as extracting visual object features.
4) **GHSOM package:** a Java library implementing the Growing-Hierarchical Self-Organizing Map algorithm [5] for dimensionality reduction. The package will be used to turn high-dimensional sensory feedback data into low dimensional discrete sequence.
5) **Sparse Coding MATLAB package:** a MATLAB library developed by Lee *et al.* [24], which will be used to extract features given depth images taken by the robot's ZCam.
6) **robocop:** C++ software, written by Vlad Sukhoy, which wraps the Barett WAM API and is used for recording the robot's sensorimotor data during object exploration trials.

C. Timeline

1) Final Object Selection: Week 1
2) Finish exploratory behavior scripting: Week 1
3) Conduct object exploratory trials: Week 1-2
4) Implement algorithms for feature extraction from sensory data: Week 3

5) Evaluate models for identification, sorting, and category learning on recorded data: Week 4-5
6) Generate result figures, write up manuscript(s) for publication: Week 6-8

References

[1] A. Amedi, K. Kriegstein, NM Atteveldt, MS Beauchamp, and MJ Naumer. Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, 166(3):559–571, 2005.

[2] F.G. Ashby and W.T. Maddox. Human category learning. *Psychology*, 56(1):149, 2005.

[3] T. Bergquist, C. Schenck, U. Ohiri, J. Sinapov, S. Griffith, and A. Stoytchev. Interactive object recognition using proprioceptive feedback. In *Proceedings of the 2009 IROS Workshop: Semantic Perception for Robot Manipulation, St. Louis, MO*, 2009.

[4] P. Bloom. How children learn the meanings of words: Learning, development and conceptual change, 2000.

[5] A. Chan and E. Pampalk. Growing hierarchical self organizing map (ghsom) toolbox: visualizations and enhancements. In *Proc. of the 9th Intl. Conf. on Neural Information Processing (NIPS)*, pages 2537–2541, 2002.

[6] M. Ernst and H. Bulthof. Merging the Senses into a Robust Percept. *Trends in Cognitive Science*, 8(4):162–169, 2004.

[7] W. Gaver. What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psych.*, 5:1–29, 1993.

[8] D. Gentner and L.L. Namy. Analogical processes in language learning. *Current Directions in Psychological Science*, 15(6):297, 2006.

[9] E. J. Gibson. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, 39:1–41, 1988.

[10] B. Giordano and S. McAdams. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *J. of the Acoustical Soc. of America*, 119(2):1171–81, 2006.

[11] M. Grassi. Do we hear size or sound? Balls dropped on plates. *Perception and Psychophysics*, 67(2):274–284, 2005.

[12] S. Griffith, J. Sinapov, M. Miller, and A. Stoytchev. Toward interactive learning of object categories by a robot: A case study with container and non-container objects. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on*, pages 1–6. IEEE, 2009.

[13] R. Hammer, A. Brechmann, F. Ohl, D. Weinshall, and S. Hochstein. Differential category learning processes: The neural basis of comparison-based learning and induction. *NeuroImage*, 52(2):699–709, 2010.

[14] R. Hammer, G. Diesendruck, D. Weinshall, and S. Hochstein. The development of category learning strategies: What makes the difference? *Cognition*, 112(1):105–119, 2009.

[15] M. Heller. Haptic dominance in form perception: vision versus proprioception. *Perception*, 21(5):655–660, 1992.

[16] C. Kemp, K.K. Chang, and L. Lombardi. Category and feature identification. *Acta psychologica*, 133(3):216–233, 2010.

[17] C. Kemp, A. Jern, and F. Xu. Object discovery and identification. 2009.

[18] C. Kemp and J.B. Tenenbaum. The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105(31):10687, 2008.

[19] P. Krojgaard. A review of object individuation in infancy. *British Journal of Developmental Psychology*, 22(2):159–183, 2004.

[20] E. Krotkov, R. Klatzky, and N. Zumel. Robotic perception of material: Experiments with shape-invariant acoustic measures of material type. In *Experimental Robotics IV*, volume 223 of *Lecture Notes in Control and Information Sciences*, pages 204–211. Springer Berlin, 1996.

[21] D. Kubus, T. Kroger, and F.M. Wahl. On-line rigid object recognition and pose estimation based on inertial parameters. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 1402–1408, 2007.

[22] D. Kubus and F.M. Wahl. Estimating Inertial Load Parameters Using Force/Torque and Acceleration Sensor Fussion. In *Robotic 2008, VDI-Beritche 2012 Munchen, Germany*, pages 29–32.

[23] S. Lederman and R. Klatzky. Haptic classification of common objects: knowledge-driven exploration. *Cognitive Psychology*, 22:421–459, 1990.

[24] H. Lee, A. Battle, R. Raina, and A.Y. Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19:801, 2007.

[25] L.S. Lopes and A. Chauhan. Scaling up category learning for language acquisition in human-robot interaction. In *Proceedings of the Symposium on Language and Robots*, pages 83–92. Citeseer, 2007.

[26] D. Lynott and L. Connell. Modality Exclusivity Norms for 423 Object Properties. *Behavior Research Methods*, 41(2):558–564, 2009.

[27] J. Modayil and B. Kuipers. Bootstrap learning for object discovery. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 1, pages 742–747. IEEE, 2005.

[28] T. Nakamura, T. Nagai, and N. Iwahashi. Multimodal object categorization by a robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2415–2420, 2007.

[29] L. Natale, G. Metta, and G. Sandini. Learning haptic representation of objects. In *Proceedings of the International Conference on Intelligent Manipulation and Grasping*, 2004.

[30] K. Nelson. Some Evidence for the Cognitive Primacy of Categorization and Its Functional Basis. *Merrill-Palmer Quarterly*, 19(1):21–39, 1973.

[31] S. Nolfi and D. Marocco. Active perception: A sensorimotor account of object categorization. In *From Animals to Animats 7: Proc. of the Sixth International Conf. on Simulation of Adaptive Behavior*, 2002.

[32] P. Potì. Logical structures of young chimpanzees' spontaneous object grouping. *International Journal of Primatology*, 18(1):33–59, 1997.

[33] Thomas G. Power. *Play and Exploration in Children and Animals*. Lawrence Erlbaum Associates, Publishers, Mahwah, NJ, 2000.

[34] M. Quigley, E. Berger, and A.Y. Ng. STAIR: Hardware and software architecture. *Presented at AAAI 2007 Robotics Workshop*, 2007.

[35] B. Rasolzadeh, M. Björkman, K. Huebner, and D. Kragic. An active vision system for detecting, fixating and manipulating objects in the real world. *International Journal of Robotics Research*, 29(2-3):133–154, 2010.

[36] J. Richmond. Automatic measurement and modelling of contact sounds. Master's thesis, University of British Columbia, 2000.

[37] J. Richmond and D. Pai. Active measurement of contact sounds. In *Proc. of ICRA*, pages 2146–2152, 2000.

[38] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz. Towards 3d point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11):927 – 941, 2008.

[39] F. Sapp, K. Lee, and D. Muir. Three-year-olds' difficulty with the appearance-reality distinction. *Developmental Psychology*, 36(5):547–60, 2000.

[40] J. Sinapov and A. Stoytchev. Object Category Recognition by a Humanoid Robot Using Behavior-Grounded Relational Learning.

[41] J. Sinapov and A. Stoytchev. From acoustic object recognition to object categorization by a humanoid robot. In *Proc. of the RSS 2009 Workshop on Mobile Manipulation, Seattle, WA.*, 2009.

[42] J. Sinapov and A. Stoytchev. The Odd One Out Task: Toward an Intelligence Test for Robots. 2010.

[43] J. Sinapov, M. Weimer, and A. Stoytchev. Interactive learning of the acoustic properties of household objects. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2518–2524, 2009.

[44] L.B. Smith, N.J. Cooney, and C. McCord. What Is" High"? The Development of Reference Points for" High" and" Low". *Child Development*, pages 583–602, 1986.

[45] T. Southey and J.J. Little. Object discovery through motion, appearance and shape. In *AAAI Workshop on Cognitive Robotics, Technical Report WS-06-03. AAAI Press*, 2006.

[46] G. Spinozzi, F. Natale, J. Langer, and K.E. Brakke. Spontaneous class grouping behavior by bonobos (Pan paniscus) and common chimpanzees (P. troglodytes). *Animal Cognition*, 2(3):157–170, 1999.

[47] S. Srinivasa, C. Ferguson, D. Helfrich, D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, and M. VandeWeghe. Herb: A Home Exploring Robotic Butler. *Autonomous Robots*, 28(1):5–20, 2009.

[48] D. Starkey. The origins of concept formation: Object sorting and object preference in early infancy. *Child Development*, 52(2):489–497, 1981.

[49] V. Sukhoy, R. Sahai, J. Sinapov, and A. Stoytchev. Vibrotactile recognition of surface textures by a humanoid robot. In *Proceedings of the Humanoids 2009 Workshop "Tactile Sensing in Humanoids - Tactile Sensors and Beyond", Paris, France*, pages 57–60, Dec 7, 2009.

[50] J. Sun, J.L. Moore, A. Bobick, and J.M. Rehg. Learning Visual Object Categories for Robot Affordance Prediction. *The International Journal of Robotics Research*, 29(2-3):174, 2010.

[51] S. Takamuku, K. Hosoda, and M. Asada. Object category acquisition by dynamic touch. *Advanced Robotics*, 22(10):1143–1154, 2008.

[52] P.D. Tremoulet, A.M. Leslie, and D.G. Hall. Infant individuation and identification of objects. *Cognitive Development*, 15(4):499–522, 2000.

[53] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[54] T. Wilcox and R. Baillargeon. Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology*, 37:97–155, 1998.

[55] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufman, San Francisco, 2nd edition, 2005.