The Boosting Effect of Exploratory Behaviors

Jivko Sinapov and Alexander Stoytchev
Developmental Robotics Laboratory
Iowa State University
Ames, IA 50011, U.S.A.
{jsinapov|alexs}@iastate.edu

Abstract

Active object exploration is one of the hallmarks of human and animal intelligence. Research in psychology has shown that the use of multiple exploratory behaviors is crucial for learning about objects. Inspired by such research, recent work in robotics has demonstrated that by performing multiple exploratory behaviors a robot can dramatically improve its object recognition rate. But what is the cause of this improvement? To answer this question, this paper examines the conditions under which combining information from multiple behaviors and sensory modalities leads to better object recognition results. Two different problems are considered: interactive object recognition using auditory and proprioceptive feedback, and surface texture recognition using tactile and proprioceptive feedback. Analysis of the results shows that metrics designed to estimate classifier model diversity can explain the improvement in recognition accuracy. This finding establishes, for the first time, an important link between empirical studies of exploratory behaviors in robotics and theoretical results on boosting in machine learning.

Introduction

Object exploration is one of the hallmarks of human and animal intelligence. Infants perform a large set of exploratory behaviors such as grasping, shaking, dropping, and scratching on most objects they encounter (Piaget 1952). Such behaviors are commonly used to learn about objects and their physical properties (Lederman & Klatzky 1987). Object exploration procedures have also been observed in a wide variety of animal species (Power 2000). Some birds, for example, perform almost their entire behavioral repertoire when exploring an object for the first time (Lorenz 1996).

Interactive object exploration is also an inherently multimodal process. For instance, surface texture can be perceived by sliding one's finger on the surface to obtain tactile sensations, but that behavior also produces auditory feedback, which can help to identify the texture (Lederman 1982). Many object properties can only be characterized using multiple modalities (Lynott & Connel 2009). In light of these findings, research in robotics has confirmed that the use of multiple exploratory behaviors and multiple sensory modalities improves interactive object recognition rates (Sinapov *et al.* 2009; Bergquist *et al.* 2009). But what causes this improvement?

This paper addresses this question by analyzing previously published datasets from two different interactive recognition tasks: 1) object recognition using auditory and proprioceptive feedback; and 2) surface texture recognition using tactile and proprioceptive feedback. More specifically, this paper examines whether metrics designed to measure classifier diversity can be used to estimate the expected improvement of accuracy when combining information from multiple modalities or multiple behaviors. The results explain, for the first time, why using multiple exploratory behaviors and multiple sensory modalities leads to a boost in object recognition rates.

Related Work

The use of behaviors in robotics has a long history (Brooks 1986; Arkin 1987; Matarić 1992). Initially, they were introduced as an attempt to simplify the control problem by splitting the robot's controller into tiny modules called behaviors (Brooks 1986). At that time, the behavior-based approach outperformed other existing control methods, which quickly increased its popularity. Recently, the research focus has shifted from using behaviors for controlling the robot to using behaviors for extracting information about objects (Fitzpatrick *et al.* 2003; Stoytchev 2005).

It was also realized that each behavior produces sensory signatures across one or more sensory modalities. This insight was used to improve the robot's knowledge about objects and their properties. For example, it was shown that integrating proprioception with vision can bootstrap a robot's ability to interact with objects (Fitzpatrick *et al.* 2003). Interaction with objects could also enable a robot to recognize them based on the sounds that they produce (Krotkov 1995; Torres-Jara *et al.* 2005) or based on the proprioceptive data generated by the robot's hand as it grasps the objects (Natale *et al.* 2004). Other experimental results show that using multiple modalities leads to a boost in recognition performance (Saenko & Darrell 2008; Morency *et al.* 2005).

Subsequent experiments have shown that robots can boost their object recognition rates by performing multiple exploratory behaviors as opposed to just one. This effect has been demonstrated with various sensory modalities, including audio (Sinapov *et al.* 2009), proprioception (Bergquist *et al.* 2009), and touch (Sukhoy *et al.* 2009; Hosoda *et al.* 2006). The source of this boosting effect, however, has not been adequately explained so far. The goal of this paper is to provide a theoretical link between the boosting effect and exploratory behaviors.

Theoretical Framework

This section describes the theoretical framework, which uses the concept of classifier diversity to study the recognition improvement attained when a robot uses multiple exploratory behaviors and multiple sensory modalities. We start with the observation that the boosting effect is similar to the classification improvement attained when using machine learning techniques such as bagging and boosting in conjunction with an ensemble of classifiers. Machine learning theory has attempted to explain the success of ensemble classifiers by introducing the concept of classifier diversity (Lam 2000; Kuncheva & Whitaker 2003). In this framework, combining predictions from diverse or complementary classifiers is thought to be directly related to the improvement in classification accuracy of the ensemble when compared to that of the individual base classifiers.

Problem Formulation

Let N be the number of behaviors in the robot's repertoire, and let M be the number of sensory modalities. Upon executing behavior i on a target object, the robot detects sensory stimuli $X_i^1, \ldots, X_i^M$, where each X_i^j is the sensory feedback from modality j. In the most general case, each stimulus can be represented either as a real-valued vector, or as a structured data point (e.g., a sequence or a graph).

The task of the robot is to recognize the target object by labeling it with the correct discrete label $c \in \mathcal{C}$. To solve this problem, for each behavior, i, and each modality, j, the robot learns a model $\mathcal{M}_i^j$ that can estimate the class label probability $Pr(c|X_i^j)$. In other words, for each combination of behavior and modality, the robot learns a classifier that estimates the class label probability for each $c \in \mathcal{C}$. The following two sub-sections describe how the robot integrates stimuli from multiple modalities and multiple behaviors in order to further improve the accuracy of its predictions.

Combining Multiple Modalities

For each behavior i, the robot learns a model $\mathcal{M}_i$, which combines the class-label probabilities of the modality-specific models $\mathcal{M}_i^j$ (for $j = 1$ to M). Given sensory stimuli $X_i^1, \ldots, X_i^M$ detected while performing behavior i on a given object, the robot estimates the class-label probabilities for this object as:

$$Pr(c|X_i^1, \ldots, X_i^M) = \alpha \sum_{j=1}^{M} w_i^j Pr(c|X_i^j)$$

In other words, given the stimuli from the M available sensory modalities, the robot combines the class-label estimates of the modality-specific models $\mathcal{M}_i^j$ using a weighted combination rule. The coefficient α is a normalizing constant, which ensures that the probabilities sum up to 1.0. Each weight w_i^j corresponds to an estimate for the reliability of the model $\mathcal{M}_i^j$ (e.g., its accuracy).

It is worth noting that humans integrate information from multiple modalities in a similar way when performing the same task (Ernst & Bulthof 2004). For example, when asked to infer an object property given proprioceptive and visual feedback, humans use a weighted combination of the predictions of the two modalities. Experimental results have shown that the weights are proportional to the estimated reliability of each modality (Ernst & Bulthof 2004). The weighted combination of predictions ensures that a sensory modality that is not useful in a given context will not dominate over other more reliable channels of information.

Combining Multiple Behaviors

To further improve the quality of its predictions, the robot uses not only multiple sensory modalities, but also applies multiple behaviors. After performing n distinct behaviors on the test object (where $n \leq N$), the robot detects sensory stimuli $[X_1^1, \ldots, X_1^M], \ldots, [X_n^1, \ldots, X_n^M]$. As in the case of combining multiple modalities, the robot uses a weighted combination rule and labels the test object with the class label $c \in \mathcal{C}$ that maximizes:

$$Pr(c|X_1^1, \ldots, X_1^M, \ldots, X_n^1, \ldots, X_n^M) = \alpha \sum_{i=1}^{n} \sum_{j=1}^{M} w_i^j Pr(c|X_i^j)$$

Intuitively, it is expected that by combining the predictions of the models $\mathcal{M}_i^j$ it is possible to achieve higher recognition accuracy than with any single model alone, especially if the weights w_i^j can be estimated accurately from the training dataset. This expected improvement is assumed to be directly related to the level of *diversity* between individual models (Lam 2000; Kuncheva & Whitaker 2003). The next subsection describes several metrics for estimating model diversity that are commonly used in the machine learning literature.

Estimating Model Diversity

Combining predictive or recognition models (e.g., classifier ensembles, mixture of experts, etc.) is an established area of research within the machine learning community. A wide variety of metrics have been developed to measure the level of *diversity* among classifiers, with emphasis on establishing a relationship between diversity and accuracy (Kuncheva & Whitaker 2003). Traditionally, such metrics have been used to compare classifiers that are trained on biased or reweighted subsets of the original dataset. In contrast, each of the robot's recognition models $\mathcal{M}_i^j$ is trained and tested on data from a particular behavior-modality combination. Next, we show how several of the proposed metrics can be extended in order to measure the diversity of the robot's recognition models derived from the N exploratory behaviors.

Let $[X_1^1, \ldots, X_1^M, \ldots, X_N^1, \ldots, X_N^M]_k$ constitute the k^{th} interaction trial (where $k = 1$ to K) during which the robot sequentially performs all N behaviors on a test object and

Table 1: The relationship between a pair of recognition models $\mathcal{M}_a$ and $\mathcal{M}_b$ can be expressed using a 2 x 2 table, which shows how often their predictions coincide (N^{11} and N^{00}) and how often they disagree (N^{01} and N^{10}).

	$\mathcal{M}_a$ correct	$\mathcal{M}_a$ wrong
$\mathcal{M}_b$ correct	N^{11}	N^{10}
$\mathcal{M}_b$ wrong	N^{01}	N^{00}

detects the sensory stimuli from all M modalities. The output of some recognition model $\mathcal{M}_a$ can be represented as a $K-$dimensional binary vector $\mathbf{y_a} = [y_{1,a}, \ldots, y_{K,a}]^T$, such that $y_{k,a} = 1$ if the model $\mathcal{M}_a$ correctly labels the object present during trial k, and 0 otherwise. One strategy for measuring the pairwise diversity between two models $\mathcal{M}_a$ and $\mathcal{M}_b$ is to compare the corresponding vectors $\mathbf{y_a}$ and $\mathbf{y_b}$.

The first metric used in this study is the *disagreement measure*, which was previously used by Skalak (1996) to quantify the diversity between a base model and a complementary model. The disagreement measure is defined as:

$$DIS_{a,b} = \frac{N^{01} + N^{10}}{N^{11} + N^{10} + N^{01} + N^{00}}$$

where N^{pq} is the number of trials (out of K) for which $y_{k,a} = p$ and $y_{k,b} = q$ (see Table 1).

In other words, the disagreement measure is simply the ratio of the number of trials in which one model was correct and the other was wrong to the total number of trials. The measure is always in the range of 0.0 to 1.0. Low values indicate that the predictions of the two models mostly agree (whether right or wrong).

The second metric used in this study is Yule's Q-Statistic (Kuncheva & Whitaker 2003), which is defined for two models $\mathcal{M}_a$ and $\mathcal{M}_b$ as:

$$Q_{a,b} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}$$

The Q-statistic ranges from -1.0 to 1.0. For statistically independent models, the expectation of $Q_{a,b}$ is 0 (Kuncheva & Whitaker 2003). A high value of Q indicates that both models label objects either correctly or incorrectly during the same interaction trials, while a low value of Q indicates that the two models commit errors on different trials.

Experimental Setup

This section briefly describes the two previously published datasets from our lab, which were obtained from their authors (along with the corresponding source code) for the purposes of this study. For more details, please refer to the original papers.

Tactile Surface Recognition Dataset

In the first dataset, the task of the robot was to recognize surface textures by applying exploratory scratching behaviors on them (Sukhoy *et al.* 2009). The robot was programmed with five different exploratory behaviors, which constitute scratching trajectories performed at different speeds and in different directions. During each scratching interaction, the robot recorded the tactile feedback from an artificial fingernail with an embedded 3-axis accelerometer and the proprioceptive joint-torque feedback from all 7 joints. Twenty different surfaces were included in the experiments. The robot performed all five scratching behaviors on each surface ten different times for a total of 1000 behavioral interactions.

Interactive Object Recognition Dataset

In the second dataset, the task of the robot was to (interactively) recognize objects using only proprioceptive and auditory feedback (Bergquist *et al.* 2009). The robot was programmed with five exploratory behaviors: *lift*, *shake*, *drop*, *crush*, and *push*. Each of these behaviors was applied ten times on fifty different objects, for a total of 2500 behavioral interactions. During each interaction, the robot recorded auditory feedback through a microphone and proprioceptive feedback in the form of joint-torque values.

Feature Extraction and Learning Algorithm

For all three modalities (auditory, tactile, and proprioceptive), the sensory stimuli X_i^j were encoded as a sequence of states in a Self-Organizing Map (SOM). A separate SOM was trained on input from each modality. Given a recorded audio signal, the Discrete Fourier Transform (DFT) was computed, which resulted in a matrix containing the intensity levels of each frequency bin over time. This high-dimensional feedback, was transformed into a sequence over a discrete alphabet by mapping each column vector of the DFT matrix to a state in a trained SOM (see Sinapov *et al.* (2009) for details). Similarly, the DFT was computed for the tactile sensory feedback as described by Sukhoy *et al.* (2009), and subsequently mapped to a discrete sequence of activated states in a SOM. The proprioceptive feedback was also represented as a sequence by mapping each recorded joint-torque configuration to a state in a SOM, which was trained on proprioceptive data as described in (Bergquist *et al.* 2009).

Each recognition model $\mathcal{M}_i^j$ was implemented as a k-Nearest Neighbor classifier with $k = 3$. The global pairwise sequence alignment score was used as the k-NN similarity function, which was computed for sequences of the same sensory modality. See (Sinapov *et al.* 2009) and (Bergquist *et al.* 2009) for more details.

Experiments and Results

Boosting Accuracy with Multiple Modalities

The first experiment explores whether the improvement attained when using multiple sensory modalities is related to the pairwise diversity metrics defined earlier. In this scenario, the robot is first evaluated on how well it can recognize the target object (or surface texture) from a single behavioral interaction with it. Table 2 shows the recognition rates for the surface texture recognition dataset when using either modality alone, as well as when the two modalities are combined. For comparison, the expected chance accuracy is $1/20 = 5.0\%$. Table 3 shows the results from the same experiment performed on the object recognition dataset. In this

Table 2: Surface Recognition from a Single Behavior

Behavior	Tactile	Proprioceptive	Combined
Lateral, fast	50.0 %	30.5 %	55.5 %
Lateral, medium	53.5 %	35.5 %	62.5 %
Lateral, slow	48.5 %	35.0 %	57.0 %
Medial, fast	42.0 %	48.5 %	57.0 %
Medial, slow	33.5 %	52.5 %	56.0 %
Average	45.5 %	40.4 %	57.6 %

Table 3: Object Recognition from a Single Behavior

Behavior	Auditory	Proprioceptive	Combined
Lift	17.4 %	64.8 %	66.4 %
Shake	27.0 %	15.2 %	29.4 %
Drop	76.4 %	45.6 %	80.8 %
Crush	73.4 %	84.6 %	88.6 %
Push	63.8 %	15.4 %	65.0 %
Average	51.6 %	45.1 %	66.0 %

case, a chance predictor is expected to achieve $1/50 = 2.0\%$ accuracy. It is clear that the reliability of each modality is contingent on the type of behavior being performed on the object. For example, when the object is lifted, the proprioceptive model fares far better than the audtory model (since little sound is generated when an object is lifted). When the object is pushed by the robot, however, the auditory modality dominates in performance.

For both datasets, combining modalities significantly improves recognition performance as compared to using either modality alone. But what is the source of this improvement? To answer this question, we can quantify the improvement in recognition accuracy and relate it to the diversity of the models. For each behavior i, let $acc(\mathcal{M}_i^j)$ be the % accuracy of the modality-specific recognition model $\mathcal{M}_i^j$ and let $acc(\mathcal{M}_i)$ be the % accuracy of the modality-combining model $\mathcal{M}_i$. We define the *Recognition Improvement* (RI) for the i^{th} behavior as:

$$RI_i = acc(\mathcal{M}_i) - \frac{\sum_{j=1}^{M} acc(\mathcal{M}_i^j)}{M}$$

To see if there is a relationship between model diversity and recognition improvement, the disagreement metric was computed for each possible combination of modality-specific models. Figure 1 shows that for both datasets this relationship is approximately linear. As predicted by machine learning theory, high pairwise disagreement generally results in higher recognition improvement. This result shows that the concept of classifier diversity can indeed be applied to the robot's behavior-derived recognition models.

Boosting Accuracy with Multiple Behaviors

The next set of experiments examines the improvement in recognition rate achieved by performing multiple ex-

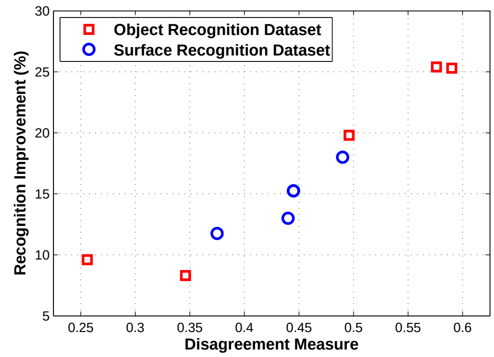

Figure 1: Pairwise disagreement measure vs. recognition improvement. Each point corresponds to one of the five behaviors in the two datasets. The horizontal axis shows the disagreement measure between the two modality-specific models, $\mathcal{M}_i^1$ and $\mathcal{M}_i^2$, for each behavior. The vertical axis shows the recognition improvement attained when both modalities are combined. In the surface recognition dataset, the points for two of the behaviors coincide.

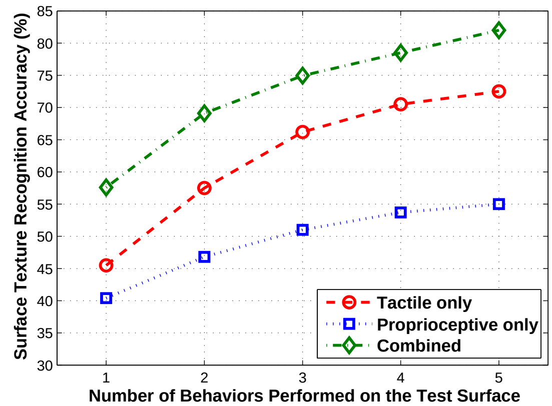

Figure 2: Surface texture recognition accuracy as the number of scratching behaviors is varied from 1 (the default, used to generate Table 2) to 5 (i.e., performing all five scratching behaviors).

ploratory behaviors on the test object/surface. Figure 2 shows the recognition accuracy for the surface texture recognition problem as the number of behaviors applied on the test surface is varied from 1 (the default, used to generate Table 2) to 5 (i.e., performing all five scratching behaviors). The results clearly show that the robot can significantly improve its recognition accuracy by applying multiple exploratory behaviors. Furthermore, the recognition rate increases at a faster pace when the predictions of the tactile models are combined, than when the predictions of the proprioceptive models are combined as shown in Figure 2. To understand the reasons why, we look at how this improvement is related to different measures of model diversity.

Given two distinct behaviors i and j, let $acc(\mathcal{M}_i, \mathcal{M}_j)$ be the estimated recognition accuracy attained by combining the predictions of the models $\mathcal{M}_i$ and $\mathcal{M}_j$ (which can be either modality-specific models or modality-combining models). The recognition improvement for two behaviors i and j is defined as:

$$RI_{ij} = acc(\mathcal{M}_i, \mathcal{M}_j) - \frac{acc(\mathcal{M}_i) + acc(\mathcal{M}_j)}{2}$$

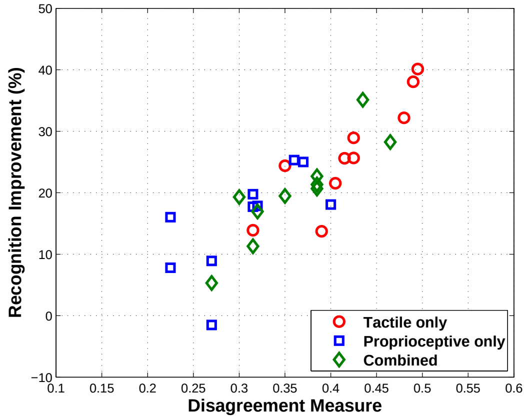

Figure 3: Pairwise disagreement measure vs. recognition improvement for the surface recognition dataset. For every unique combination of 2 behaviors (10 total for 5 behaviors), there are 3 points in the plot, one for each of the three conditions: touch, proprioception, or both. The horizontal axis shows the estimated disagreement measure between the two behavior-derived models, while the vertical axis shows the recognition improvement attained when applying both behaviors.

Figure 3 plots the disagreement measure vs. the recognition improvement for the surface recognition dataset. Because there are 5 behaviors in that dataset, we can form 10 different pairs of behaviors for which the improvement in recognition accuracy can be calculated under three different conditions: touch only, proprioception only, or both. We can also calculate the diversity between any two behavioral models. The results show that the amount of disagreement is directly related to the expected improvement. On average, the pairwise disagreement for the tactile recognition models is higher than that for the proprioceptive models. This explains why the improvement attained by applying multiple behaviors is greater with the tactile sensory modality.

The same plot can also be calculated for the object recognition dataset. A comparison plot in Figure 4 shows the relationship between the disagreement measure and the classification improvement for both datasets. There is a linear relationship between the diversity metric and the observed boost in the recognition rate. As predicted by machine learning theory, higher diversity results in higher accuracy improvement. This result shows that the disagreement measure is a good indicator for the expected recognition improvement, a finding that generalizes to both datasets.

Figure 5 shows the relationship between the Q-statistic and the recognition improvement for both datasets. The Q-statistic is approximately linearly related to the accuracy improvement in the surface recognition dataset, but there is no clear relationship in the object recognition dataset. This is indeed a surprising result, since the Q-statistic is typically the most common metric used for estimating the diversity between two classifier models and has been recommended as a good metric for measuring classifier model diversity (Kuncheva & Whitaker 2003). Several factors might explain this apparent discrepancy. First, the individual classifier models in Kuncheva & Whitaker (2003)'s study had

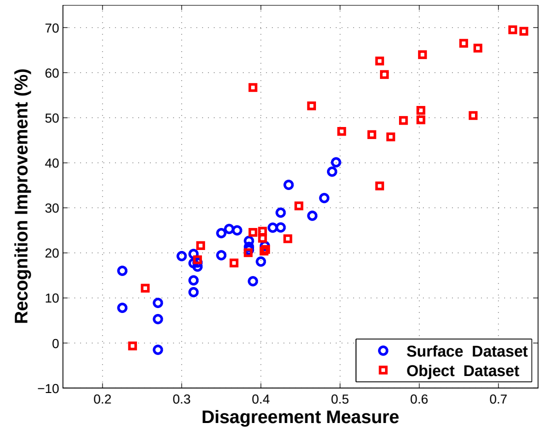

Figure 4: Pairwise disagreement measure vs. recognition improvement for each of the 10 possible pairs of behaviors, under three different modality conditions (modality 1 only, modality 2 only, or combined) for both datasets.

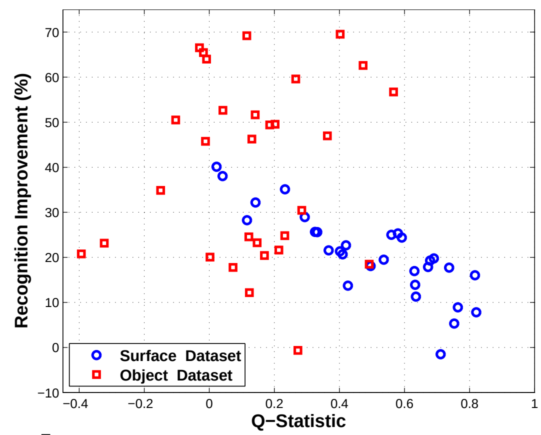

Figure 5: Pairwise Q-statistic vs. recognition improvement for each of the 10 possible pairs of behaviors, under three different modality conditions (modality 1 only, modality 2 only, or combined) for both datasets.

approximately the same individual accuracies. The individual recognition models used in the interactive object recognition task, however, have very different accuracies (see Table 3). For example, performing the *shake* behavior results in 29.4% recognition rate, while the *drop* behavior achieves 80.8%. Second, it has been shown by Dietterich (2000) that different methods for building collections of classifiers can result in different relationship patterns between diversity and improvement. Typically, it is assumed that each classifier model in the ensemble is trained on some biased subset (or otherwise modified version) of the original training set. In contrast, the recognition models learned by the robot are constructed in a profoundly different manner - each of the robot's recognition models is trained and tested only on data from a particular behavior-modality combination. Despite these differences, the concept of classifier diversity was still found to be useful for explaining the improvement in recognition accuracy in the robot experiments.

Summary and Conclusion

Exploratory behaviors play an important role in the object exploration patterns of humans and animals (Power 2000; Lorenz 1996). When these behaviors are applied on objects they act like "questions," which the object "answers" by producing effects across multiple sensory modalities. When multiple behaviors are performed the identity of the object can be uniquely identified. Recent studies have shown that robots can also use exploratory behaviors to improve their object recognition rates. The reasons for this improvement, however, have not been adequately explained so far.

This paper formulated a new metaphor to explain these results, namely, *behaviors are classifiers*. Thus, the behavioral repertoire of the robot can be viewed as an ensemble of classifiers, which can be boosted. The boosting effect generalizes not only to multiple exploratory behaviors, but also to multiple sensory modalities. Each new modality and each new behavior provides additional information that can be used to construct new classifiers.

Two large datasets with 50 objects and 20 surfaces were used to generate the results, which clearly show that the metrics designed to measure the diversity of classifiers can be applied to measure the diversity of the behaviors in the robot's behavioral repertoire. In particular, the *disagreement measure* for two behavior-derived recognition models was found to be linearly related to the observed boost in recognition rate when both behaviors are applied. This is an important contribution as it establishes for the first time a link between empirical studies of exploratory behaviors in robotics and theoretical results on boosting in machine learning.

References

Arkin, R. 1987. Motor schema based navigation for a mobile robot: An approach to programming by behavior. *Proc. of IEEE ICRA*, pp. 264–271.

Bergquist, T.; Schenck, C.; Ohiri, U.; Sinapov, J.; Griffith, S.; and Stoytchev, A. 2009. Interactive object recognition using proprioceptive feedback. *Proc. of the IROS Workshop: Semantic Perception for Mobile Manipulation.*

Brooks, R. 1986. A robust layered control system for a mobile robot. *J. of Robotics and Automation* 2(1):14–23.

Dietterich, T. 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning* 40(2):139–157.

Ernst, M.; and Bulthof, H. 2004. Merging the Senses into a Robust Percept. *Trends in Cognitive Science* 8(4):161–169.

Fitzpatrick, P.; Metta, G.; Natale, L.; Rao, S.; and Sandini, G. 2003. Learning about objects through action - initial steps towards artificial cognition. *In Proc. of IEEE ICRA*, pp. 3140–3145.

Hosoda, K.; Tada, Y.; and Asada, M. 2006. Anthropomorphic robotic soft fingertip with randomly distributed receptors. *Robotics and Autonomous Systems* 54(2):104–109.

Krotkov, E. 1995. Perception of material properties by robotic probing: Preliminary investigations. *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, pp. 88–94.

Kuncheva, L.; and Whitaker, C. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning* 51:181–207.

Lam, L. 2000. Classifier combinations: Implementations and theoretical issues. In W. Schiff, & E. Foulke (Eds.) *Tactual Perception: a sourcebook* (pp. 130-168), Springer.

Lederman, S. 1982. The perception of texture by touch. In J. Kittler, & F. Roli (Eds.) *Multiple Classifier Systems* Vol. 1857 of LNCS (pp. 78-86). Cambridge University Press.

Lederman, S.; and Klatzky, R. 1987. Hand movements: a window into haptic object recognition. *Cognitive Psychology* 19:342–368.

Lorenz, K. 1996. Innate bases of learning. In K. H. Pribram, & J. King (Eds.) *Learning as Self-Organization*. Lawrence Erlbaum Assoc., Publishers, Mahwah, NJ.

Lynott, D.; and Connel, L. 2009. Modality exclusivity norms for 423 object properties. *Behavior Research Methods* 41(2):558–564.

Matarić, M. 1992. Designing emergent behaviors: From local interactions to collective intelligence. *Proc. of the Intl. Conf. on Simulation of Adaptive Beh.*, pp. 432-441.

Metta, G.; and Fitzpatrick, P. 2003. Early integration of vision and manipulation. *Adaptive Beh.* 11(2):109–128.

Morency, L-P.; Sidner, C.; Lee, C.; and Darrell, T. 2005. Contextual recognition of head gestures. *Proc. of the Intl. Conf. on Multimodal interfaces*, pp. 18–24.

Natale, L.; Metta, G.; and Sandini, G. 2004. Learning haptic representation of objects. *Proc. of the Intl. Conf. on Intelligent Manipulation and Grasping.*

Piaget, J. 1952. The origins of intelligence in children. International Universities Press, New York.

Power, T. 2000. Play and exploration in children and animals. Mahwah, NJ: Layrence Erlbaum Assoc., Publishers.

Saenko, K.; and Darrell, T. 2008. Object category recognition using probabilistic fusion of speech and image classifiers. *Machine Learning for Multimodal Interaction* 4892:36–47. Springer, Berlin / Heidelburg.

Sinapov, J.; Weimer, M., and Stoytchev, A. 2009. Interactive learning of the acoustic properties of household objects. *Proc. of IEEE ICRA*, pp. 3937–3943.

Skalak, D. 1996. The sources of increased accuracy for two proposed boosting algorithms. *Proc. AAAI, Integrating Multiple Learned Models Workshop.*

Stoytchev, A. 2005. Behavior-grounded representation of tool affordances. *Proc. of IEEE ICRA*, pp. 3071–3076.

Sukhoy, V.; Sahai, R.; Sinapov, J.; and Stoytchev, A. 2009. Vibrotactile recognition of surface textures by a humanoid robot. *Proc. Humanoids Workshop: Tactile Sensing in Humanoids-Tactile Sensors and Beyond*, pp. 57-60.

Torres-Jara, E.; Natale, L., and Fitzpatrick, P. 2005. Tapping into touch. *Proc. of Intl. Workshop on Epigenetic Robotics*. pp. 79–86.