

# Active Learning for Vision-Based Robot Grasping

MARCOS SALGANICOFF

salganic@asel.udel.edu

*Applied Science and Engineering Labs, University of Delaware/Alfred I. duPont Institute,  
P.O.B. 269, Wilmington, DE 19899*

LYLE H. UNGAR

ungar@central.cis.upenn.edu

RUZENA BAJCSY

bajcsy@central.cis.upenn.edu

*Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, 19104*

**Editors:** Judy A. Franklin, Tom M. Mitchell, and Sebastian Thrun

**Abstract.** Reliable vision-based grasping has proved elusive outside of controlled environments. One approach towards building more flexible and domain-independent robot grasping systems is to employ learning to adapt the robot's perceptual and motor system to the task. However, one pitfall in robot perceptual and motor learning is that the cost of gathering the learning set may be unacceptably high. Active learning algorithms address this shortcoming by intelligently selecting actions so as to decrease the number of examples necessary to achieve good performance and also avoid separate training and execution phases, leading to higher autonomy. We describe the IE-ID3 algorithm, which extends the Interval Estimation (IE) active learning approach from discrete to real-valued learning domains by combining IE with a classification tree learning algorithm (ID-3). We present a robot system which rapidly learns to select the grasp approach directions using IE-ID3 given simplified superquadric shape approximations of objects. Initial results on a small set of objects show that a robot with a laser scanner system can rapidly learn to pick up new objects, and simulation studies show the superiority of the active learning approach for a simulated grasping task using larger sets of objects. Extensions of the approach and future areas of research incorporating more sophisticated perceptual and action representation are discussed

**Keywords:** robots, vision, manipulation, active learning, grasping, interval estimation

## 1. Introduction

Vision-based grasping and retrieval of objects is an important skill in many tasks. A robotic system having the ability to perceive pertinent target object features, and the ability to select viable grasp approach and preshape values for a robotic hand would be able to carry out many useful functions. Possible scenarios for such a system would range from the handling of toxic materials in dangerous environments to the assistance of people with physical disabilities in relatively benign household environments. While significant progress has been made in the development of machine perception, highly dextrous robotics hands, and robotic grasp analysis and synthesis, limitations in the state of the art in vision and robotics still preclude the availability of "off-the-shelf" robot systems with truly effective vision and manipulation capabilities. This is primarily due to limitations in the reliability of machine perceptual systems in varied environments, as well as the difficulties in the control of dextrous manipulators. The environments into which a robot system is placed may have a variety of unfamiliar objects, materials and lighting conditions, making the perceptual and manipulation task difficult without the use of a significant number of domain-dependent techniques.

### 1.1. Previous Work

Previous work in robotic grasping may be divided into two major approaches: analytical approaches based on detailed theoretical analysis of the mechanics of contacts and heuristic approaches inspired by the research of neuropsychologists (Jeannerod & Biguer, 1981, Jeannerod, 1988) and orthopedic medicine (Schlesinger, 1916, Napier, 1956).

Analytical approaches to grasping have analyzed grasp selection based on mechanics, most often in terms of maximizing some objective function such as grasp stability, slip-resistance, or minimizing some objective function, such as internal forces (see(Cutkosky, 1989) for a survey). These approaches have given a great deal of theoretical insight into grasp planning, although they incorporate simplifying assumptions in order to be tractable and assume essentially perfect perception of relevant object properties such as shape and mass. Unfortunately, these *a-priori* assumptions may not be valid in all situations and may become less accurate as a system changes over time due to mechanical wear or other factors.

An alternative approach to selecting grasp approach and preshapes is to use a heuristic approach which employs domain-dependent rules based on objects and selections that are empirically observed. Heuristically-based grasp generators almost always include some equivalent of major grasp preshape types as categorized by Schlesinger (Schlesinger, 1916) including the fingertip, hook, cylindrical, palmar lateral and spherical grasp types. These different categories were formed based on functional analysis of human grasping behavior where different characteristics classes of hand shapes were observed to be associated with the prehension of different object shapes and tasks combinations.

A number of rule-based systems for grasp selection (Iberall, et al., 1988, Liu, Iberall & Bekey, 1989), were subsequently developed based on how skilled individuals chose grasps. These systems incorporate production rules that capture how humans choose particular preshapes and grasp sites. These approaches share a significant limitation with the analytic systems discussed so far in that they treat grasp selection as a completely separate process from perception, assuming perfect perception of object shape and other object properties. For these techniques to be practical, they require the machine perception system to provide descriptions at a level of detail and accuracy that is far beyond what can be provided by existing vision systems, especially if the variability of objects and the environment is taken into account.

One approach to solving the perceptual bottleneck is to retreat from general purpose systems and constrain the domain in which the autonomous system is to perform. Given sufficient constraints, a system may be engineered by taking advantage of *a priori* knowledge to simplify visual reconstruction of the scene and motor processing. Stansfield (1990) describes a heuristic rule-based system for grasp selection that uses a symbolic multi-component representation of object views derived from a range scanning system. If 2-D object contours can reliably be extracted from vision processing in a particular domain, then techniques employing analytical grasp point selection for optimal grasp on smooth closed 2-D contours can be used (Taylor, Blake & Cox, 1994). Alternative approaches such as Bard (1991, 1995, *in press*) compute ellipsoidal decompositions of voxel description that rely on fusing multiple stereo vision views in a space occupancy grid of voxels. The preshape

generation process attempts to identify feasible ellipsoid parts for grasping according to heuristics based on part size and accessibility, followed by a physics-based simulation to verify a grasp before execution.

Unfortunately, retreating from a truly general purpose system by hand selecting reliable domain-dependent perceptual attributes (e.g. smooth object contours), feature extractors and special purpose grasping routines requires custom programming and engineering which can add significant cost to robot systems. Additionally, it is a great burden on the engineers and programmers to guarantee system reliability over a combinatorially large number of possible scenarios. Each time the system is deployed in a different domain, significant re-engineering effort may be necessary to adapt the system to function correctly. This additional cost makes it difficult to justify the use of robots as a truly flexible task solution. In future situations that may require true long-term autonomy, such as space exploration, there may not be engineering personnel available to carry out the required adaption.

### ***1.2. Learning Approaches to Vision Based Grasping***

One approach to avoiding the pitfalls of domain-dependent systems is to employ perceptual and/or motor learning techniques. Many researchers (Kuniyoshi, Inaba & Inoue, 1989, Kang & Ikeuchi, 1993, Dunn & Segen, 1988, Bennett, 1991, Kamon, Flash & Edelman, 1994, Salganicoff, 1992, Salganicoff, 1993) have taken the approach that it is preferable for robotic systems to learn a grasping strategy. The learning approach has the benefit that the system adapts to the characteristics of objects in its domain as well as to its perceptual and motor system, providing a system with higher autonomy. Robot learning systems can incorporate the true idiosyncrasies of their sensing and motor apparatus rather than rely on possibly inaccurate or outdated expectations of the system's designers. By taking a learning approach to vision-based robotic grasping (and robotic tasks in general), the noisiness and variability of both the perceptual, action systems and the environment can be incorporated into the robotic system to better ensure success, since many learning algorithms can handle noisy or missing inputs. Secondly, gradual drift or even catastrophic failure in sensor and motor systems and in the properties of the environment may be compensated for using learning without additional programming effort, leading to systems that are more robust and autonomous. If a sufficiently rich perceptual and motor representation is chosen, then the learning system can use these representations as needed over the many different environments where the robot might be placed.

Techniques in learning to grasp have taken both supervised and unsupervised approaches. Supervised approaches are exemplified by a rote learning or "teaching by showing" approach developed by Kuniyoshi et al. (1989) and Kang and Ikeuchi (1993). The observed history of actions is used to form a plan composed of action primitives that the robot can carry out. In general, these systems require significant *a priori* knowledge about the task and controlled environmental conditions in order to enhance the speed and reliability of the vision system in recognizing actions and objects in real-time.

Unsupervised learning approaches do not require human intervention and so are better suited for high autonomy. A number of systems have used the output of computer vision, primarily in the form of two dimensional edge contours, to drive either recognition and

indexing of previously grasped objects, or to use the local object contour information as the basis of features for classifiers. Dunn & Segen (1988) employed an unsupervised memory-based learning approach with a vision derived two dimensional polygonal representation; grasp was coded by relative location and orientation to the object. During execution, a 2-D model matcher was used to index the presented object to a previously manipulated object and to invoke the previously successful grasp.

Kamon et al. (1994) used a contour-based two dimensional representation of the perimeter of objects and measured grasp location and human selected heuristic quality parameters for the object. During on-line learning the system tried to apply previously attempted grasps to the current object and then compared their predicted fitness using a nearest-neighbor learning rule. If no grasp had high predicted fitness, a randomized domain-specific heuristic was used to select a new grasp.

A knowledge intensive approach to learning to grasp was taken by Bennett (1991) who worked in robotic grasping of polygonal 2-D puzzle piece tasks using explanation-based learning and simple domain theories about uncertainty and grasping.

Tan (1990, 1993) employed a feature-based sonar depth representation and cost-sensitive learning (CSL) extension of ID-3 (Quinlan, 1986) to learn to recognize object labeled by high-level grasp-approach combinations (e.g. approach object from top gripping rim in pinch grasp).

Stansfield has also investigated replacing expert system production rules for grasp selection (Stansfield, 1990) with the batch-learning of associations using a connectionist representation with back-propagation learning (Rumelhart, Hinton & Williams, 1986). The system was trained to associate feature/view bindings to grasp categories and grasp sites.

Salganicoff (1992, 1993) developed a system for grasping objects based on full three-dimensional information from a laser range scanner and a parameterized superquadric representation. It learned to select from a predetermined set of canonical approach directions using a density-adaptive learning algorithm along with a forgetting component to allow learning set updates to incorporate changes in the task.

### *1.3. Motivation for Active Learning*

The advantages of a learning approach to vision-based grasping, such as increased autonomy, and self adaptation to limits in sensing, action and the particulars of a given domain, are, of course, not without tradeoffs. In robot learning, as in many other learning tasks, the costs of acquiring the learning set may be the single greatest barrier towards the application of learning in a task. The acquisition costs for each exemplar can be particularly expensive, since each sensing and/or motor action often takes a significant amount of time to execute, and has other concomitant risks as well. In fact, the computational complexity of invoking the learning algorithm may be small in comparison to the cost in time and material of gathering the learning set (Tan, 1993). Therefore it is imperative to decrease learning set cost, and active learning provides an unsupervised way of doing so. Active learning systems allow the learner to control where in the input space their exemplars are drawn. They thus permit the learners to use strategies which balance the costs of gathering exemplars for

learning (exploration) against the cost of misclassification during the execution of the task (exploitation).

Another way of looking at active learning is to view it as an optimization problem in which one is simultaneously building a model of a process and optimizing its performance. These problem has been studied by several different sets of researchers. In applied statistics, e.g. for the chemical process industries, evolutionary optimization (EVOP) has long been influential (Box & Draper, 1969). In EVOP, one fits a local linear or quadratic model to the available data, moves the process toward an optimum, fits the model again, moves again and iterates. These active learning methods are being applied today in large chemical plants (see (Moreno & Yunker, 1992)).

A second line of research comes from statistical decision theory and multi-armed bandits, and uses methods such as the Sequential Probability Ratio Test (SPRT), which rigorously computes whether one should collect further data points (see (Berger, 1985), and Gittins indices (Gittins, 1988)), which allow one to compare multiple arms (courses of actions) against a single reference arm, and thus avoid pairwise comparisons. As with all Bayesian methods, one must assume models of the distributions of the observations, including their correlation structures (if any) and one needs priors on the parameters in the models. The computations become exceedingly difficult for complex models of the probability distributions or costs.

A third line of research comes from the machine learning community, in particular workers in reinforcement learning and robot learning. Because of the unique demands of robot learning a number of active exploration approaches have been developed (Thrun & Moller, 1992, Moore, 1990, Schneider, 1993, Sutton, 1990, Atlas, et al, 1990, Cohn, 1994) to accelerate learning in those domains. Reinforcement learning researchers have also developed a variety of active exploration heuristics (Sutton, 1990, Kaelbling, 1990, Thrun & Moller, 1992) that tradeoff exploration and exploitation during adaptation.

This article presents a general technique and evaluation for active learning. The approach combines the Interval Estimation (IE) (Kaelbling, 1990) exploration heuristic with the ID-3 inductive learning algorithm (Quinlan, 1986) and does not require separate training and performance phases. Actions are then chosen using an indirect indexing scheme (Salganicoff & Bajcsy, 1992) which indexes the leaves on the ID-3 tree that have a high expected probability of success (reward) conditioned on the current perceived attributes.

We apply this framework to a simplified version of the vision-based grasping problem, where the approach axis for a two fingered grasp must be selected based on the extents of the three semi-major superquadric axes of the target object. The result is a learning system that smoothly combines its training and exploitation of learned knowledge, decreasing the number of failures encountered until good grasping rules are induced. This allows the learner to benefit from learning as it occurs instead of forcing the system to wait until the end of the learning phase to take advantage of the acquired knowledge. The general framework for the active learning algorithm IE-ID3 is described in section 2. The specific grasping problem instance, representation, experimental setup, and empirical and simulation results for the method are elaborated upon in section 3. In experimental trials with a robotic system with a set of 6 objects, the system learns to select reliable approach direction in a small number

of attempts; convergence occurs in approximately 30 attempts per object. Simulations give a qualitative picture of the system's asymptotic performance as it interacts with a large number of objects. In simulations the IE-ID3 algorithm performs at almost twice the level of an open loop non-active learner, and also results in a higher classification accuracy. Section 4 compares IE-ID3 with other active learning and grasp learning approaches, and points out areas of future research, limitations and possible improvements. Finally, section 5 concludes with a synopsis of the approach and results.

## 2. An Active Learning Algorithm for Real-Valued Domains

Many approaches for action selection in active learning keep statistics relating to a particular discrete state or state/action pairs (Thrun, 1992). A number of different rationales have been used. For example, inversely weighing the raw number of times a state has been visited (Thrun, 1992) as a factor that favors execution of an action results in a system that is generally exploratory in nature, attempting to search the entire input space. Taking recency of a state visitation (Sutton, 1990) as a positive weighting for action selection gives a system which attempts to evenly visit the entire state space and track changes in the environment. These types of algorithms have been developed in the context of delayed reinforcement problems (Sutton, 1988, Watkins, 1989) with Markovian frameworks having discrete action and state framework, making them poorly suited for operation in continuous state and action spaces. However, it is possible to adapt techniques such as reinforcement learning to continuous domains through adaptive input-space partitions of those spaces. For example, Moore (Moore, 1991b) describes a modification to discrete state dynamic programming by the use of an adaptive resolution tree that partitions a continuous state-space into intervals. We employ a similar tree-based partitioning approach by using a classification tree algorithm (ID-3) (Quinlan, 1986, Quinlan, 1992) to form a partition over the continuous space, yielding an adaptively discretized space over which the Interval Estimation (IE) exploration algorithm ((Kaelbling, 1990), pp. 56), a discrete-space approach, can be applied.

The IE algorithm (Kaelbling, 1990) works by keeping statistics on the number of times a given action has been executed and the proportion of times that it has succeeded. A confidence interval is computed for the underlying probability of success for each of the feasible actions in a given context. The action whose upper-bound confidence interval is highest is chosen. The rationale is that an action may have a higher upper-bound for two reasons: either because the action is actually a good one to take, or because very few trials of the action have been executed, in which case there is insufficient empirical evidence to decide the goodness of the action in that context. Most importantly, this approach rapidly stops executing actions which have a very low likelihood of success and focuses on sampling those actions whose conditional probability of success cannot statistically be ruled out as best. If two actions share the same upper-bound, then they are chosen with equal probability.

We generalize the IE algorithm from its original formulation for a finite number of discrete actions to continuous valued perceptual and action spaces by combining it with the ID-3 algorithm (Quinlan, 1986) incorporating real-valued attribute splits. ID-3 automatically constructs decision trees by determining what questions (real-valued feature threshold values) give maximum information gain at each level of the tree. ID-3 and enhanced

methods such as Classification and Regression Trees (CART) (Breiman, et al., 1984) and C4.5 (Quinlan, 1992) are widely used and can handle high-dimensional feature spaces, ignoring irrelevant features, and, when the trees are pruned to eliminate spurious branches, are relatively insensitive to noise in the data. The resulting ID-3 decision tree induces a hard partition on the real-valued attribute space with its leaves.

Consider an agent that perceives the world through a perceptual vector of continuous attributes  $P$ , and can continuously control the value of its action parameters through an action vector  $A$ , to yield a binary outcome  $O$ . The agent attempts to learn the mapping  $O = f(P, A)$  over some sub-domain of possible  $P$  and  $A$  values. This function can then be used to maximize its reward by selecting actions  $A$  that maximize the probability of  $O = 1$  given what is currently perceived in  $P$ . Therefore, rather than have an input space that consists solely of action attributes, we use an extended representation which consists of a combined perception-action space  $(P, A) = ((p_1, \dots, p_l), (a_1, \dots, a_m))$ , where  $p_i$  and  $a_i$  are the individual perception and action attributes for an agent.

At each step in the execution of the overall algorithm, we select an action as described below, execute it, ascertain its outcome, and then compute a classification tree  $T$  using ID-3 with real-valued splits, based on the distribution of binary outcomes  $O$  in the combined perception-action space. This cycle is repeated for the duration of the task.

### 2.1. Indirect Prediction

To select actions in the combined perception-action space, the system uses the tree  $T$  as an indirect predictor (Salganicoff, 1993) along with the IE method. First, the system senses the perceptual vector  $P$ . Let  $F$  be the set of leaves of the ID-3 tree that have perceptual interval values that intersect the currently sensed attribute values of  $P$ . The set  $F$  represents a partitioning or binning of the action space by the current decision tree subject to the perceived attributes,  $P$ , of the current object. As we vary  $A$  over all possible values with  $P$  fixed, we move over all possible leaves which predict outcomes for different action parameters  $A$  (see Figure 1). Indirect prediction is similar to the idea of using indirect control by taking the reward gradient of the reward function approximation with respect to the action space in order to hill climb to the action with maximum utility (Thrun, 1992). However in our case, the function representation is in terms of a tree decomposition, and is not differentiable, therefore we must do a partial-match search to maximize the probability of rewards over the feasible leaves of the tree.

Given this partitioning into bins in which different actions lead to different outcomes (all conditional on the perceptual vector  $P$ ), we can directly apply the IE algorithm to choose among the possible actions. The statistics necessary for computing the confidence intervals (the number of success and failures contained in a leaf partition) are already available because they are needed to form the tree using the ID-3 algorithm.

For each leaf  $l_i$  in  $F$  we compute the upper-bound probability of success according to the binomial confidence interval formula (Larson & Marx, 1986). To calculate a probability interval that contains the true probability value with confidence  $(1 - \alpha)$  given  $x$  successes (rewarding outcomes) out of  $n$  exemplars in the leaf so that

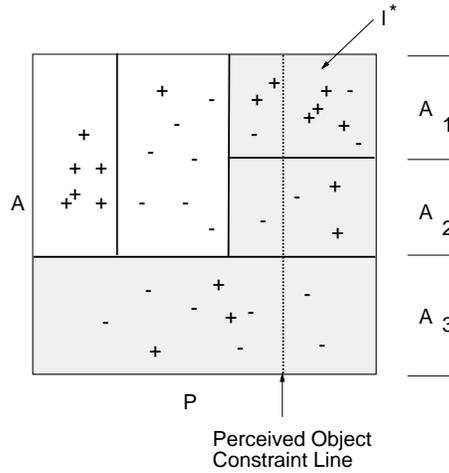


Figure 1. A schematic of the representation. The solid lines represent the ID-3 partitioning of the joint perception-action space  $(P, A)$ . For the purposes of illustration assume a single degree-of freedom action  $A$  and a single sensed value  $P$ , although the approach generalizes to an arbitrary number of dimensions. When a new percept is sensed, this determines the discrete intervals for action values (the dotted line) and their associated statistics of task success (+) and failure (-). The feasible set of leaves  $F$ , is shaded. All feasible leaves are then compared in terms of the upper-bound confidence interval. An action is randomly selected from within the interval corresponding to the leaf with the highest upper-bound confidence interval.

$$P_- \leq P_{success} \leq P_+ \tag{1}$$

where,  $P_-$  and  $P_+$  are the lower and upper-bounds of the probability interval estimate, respectively, one uses:

$$P_{\pm} = \frac{\frac{z_{\frac{\alpha}{2}}^2}{2n} + \frac{x}{n} \pm \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\frac{z_{\frac{\alpha}{2}}^2}{4n} + \frac{x}{n^2} (1 - \frac{x}{n})}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}} \tag{2}$$

where  $z_{\frac{\alpha}{2}}$  is the confidence interval coefficient. The confidence interval coefficient is the deviation from 0 which is exceeded by a random variable with standard normal distribution with probability  $\frac{\alpha}{2}$ . Intuitively, the smaller  $\alpha$  is chosen, the more exploratory the character of the learner, since more experiences are required to drive the confidence intervals down. In the limit, as  $\alpha \rightarrow 0$ , the system becomes completely exploratory and picks all actions as equally good, since the confidence interval upper-bounds approach 1. At the other extreme, as  $\alpha \rightarrow 1$  the system become completely exploitative, since the upper-bound of the estimate approaches the raw empirical probability of the action succeeding.

The selected action leaf  $l^*$  is the leaf in  $F$  with highest upper-bound confidence interval as prescribed in the IE algorithm. Let  $A_{min}, A_{max}$  be the lower and upper action attribute intervals for  $l^*$ . The next action is chosen as random within  $l^*$  using a uniform distribution bounded by the intervals  $A_{min}, A_{max}$ .

### 3. Applying IE-ID3 to the Selection of Grasp Approach Directions

Robotic grasping and retrieval of unknown objects in unstructured environments is an important problem, and has many applications as discussed in the introduction. It is particularly suitable for learning-based approaches since the deployment of the system may occur in various environments. Many of those domains may be difficult to characterize ahead of time, making it difficult to guarantee the existence of specialized domain-specific perceptual and motor strategies that will succeed for all possible scenarios.

In our approach to vision-based grasp learning, we adopt the framework of the two phase model of grasping, consisting of ballistic approach and preshape phases as suggested by the work of (Jeannerod & Biguer, 1981, Jeannerod, 1988). Looking at learning as applied to the both phases leaves many free parameters to be set in both the ballistic approach phase that determines the direction of approach and finger contact locations, as well as in the preshape phase, where preshape category (e.g. pinch) is chosen and parameterized in terms of characteristics such as aperture, inter-finger span and finger stiffness. Similarly, from the perceptual standpoint, many other superquadric shape parameters and other sophisticated shape representations, including taper and bending, can be part of the perceptual representation. Because we had only a parallel jaw gripper available for the experimental setup we simplified our action representation to controlling only the approach direction of the gripper relative to the centroid of the object and held other action parameters fixed. We also simplified the perceptual representation to just the extents of the object's three semi-major axes, since we were most interested in exploring feasibility of the active learning approach using the simplest possible representation (see Figure 2).

This simplified representation was chosen for a number of reasons. Our previous experience with our grasping system showed that running experiments with the our laboratory setup up was a frustrating process, due to the unreliability that is typical of laboratory prototype setups. This frustration motivated us to pick minimal representations of perception and action to see if our active learning approach could succeed with a very small number of grasping attempts during a single session of grasping. It is important to note that this does not imply a fundamental limitation of the learning system we developed, since the decision tree learning algorithms used are designed for higher dimensional spaces and can easily accommodate the addition of richer descriptions of object shape, both for categorical and real-valued attributes (this is further discussed in section 4).

#### 3.1. *Experimental System*

The robotic system which carried out the experiments consists of two PUMA 560 robots: the perceptual robot and the grasper. The pertinent aspects of the system are summarized in Figure 3. The range image provided by the scanner (see Figure 4) is processed using a superquadric data reduction procedure (Solina & Bajcsy, 1990) which fits a superquadric to the dense range data. The superquadric is a necessary data reduction step, as the range image can contain thousands of points. It was chosen as a representation since it yields a compact set of parameters that attempt to approximate the overall geometry of the object to be manipulated; implications of this perceptual representation are addressed in section 4.

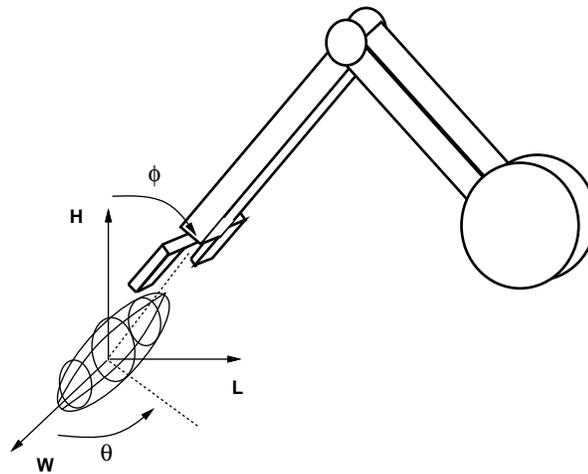
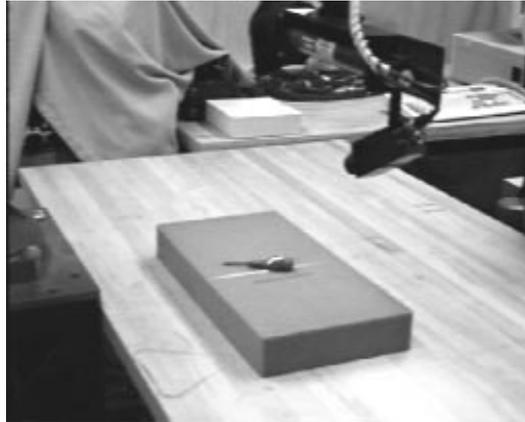


Figure 2. The perceptual and action representation for the grasping system.



Figure 3. The experimental system, (a) consisting of the perceptual robot (left) and the grasping robot (right). The perceptual robot first scans the unknown object to yield a range image. A set of parameters is found which best fits the range data and this is used as a perception input to the learning system. The grasper then attempts to pick up the object based on the percept plus its acquired knowledge about the effectiveness of different grasp approach directions.

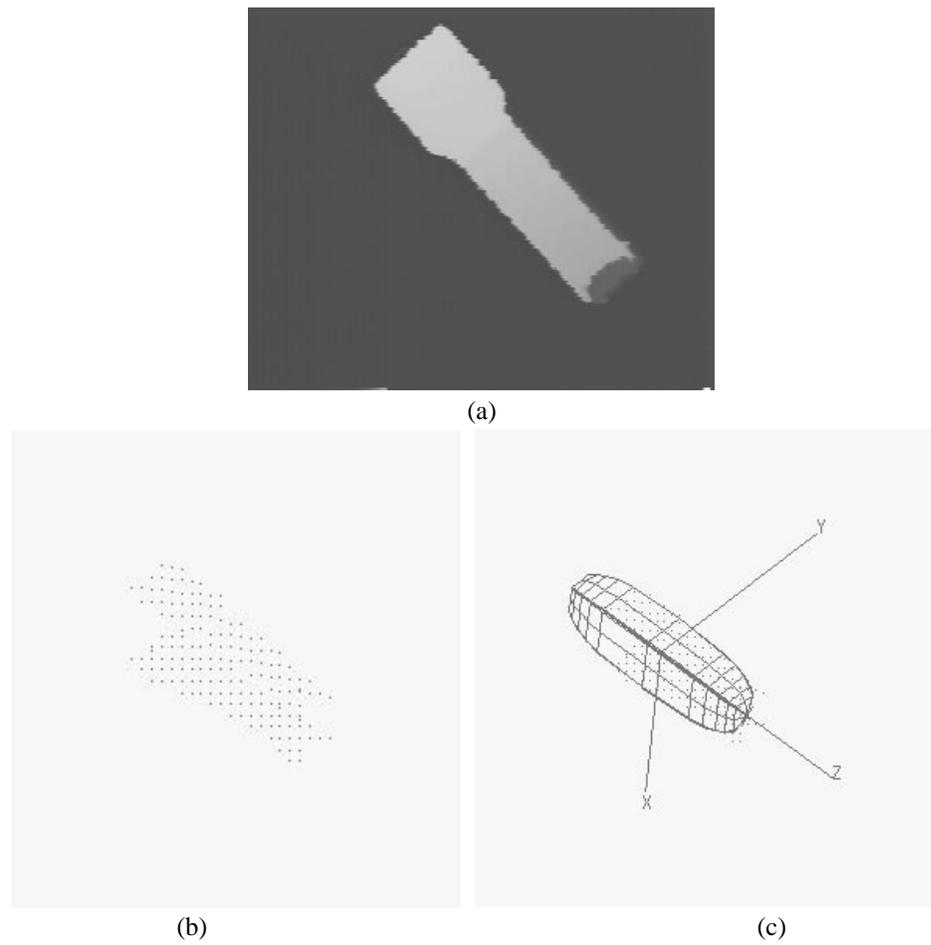


*Figure 4.* The vision system consists of a mobile laser range scanner mounted on the end of the perceptual robot which is swept over the object to be grasped. This yields a sequence of height values which can be combined into a complete range image. This range image is then processed by fitting a parameterized surface (a superquadric) to the height values in the range image. The resulting parameters for the surface are then the input attributes to the learning system along with the chosen action.

Our Laser Range Imaging System consists of two components: The *LOOKER* and the *GUS* processing unit (see Figure 4). The *LOOKER* is composed of a laser stripe generator and SONY XC-39 camera which generates video signal of the images obtained under the illumination of the laser stripe, and the *GUS* unit (Tsikos, 1991) processes the continuous sequence of laser images and generates a static range image of the scene.

The range image is first cleaned by morphologically eroding it in order to remove any spurious outlying measurements and then sub-sampled (see Figures 5 (a,b)). The resulting range points then form the constraints for the numerical optimization to extract the parameters for a surface which best-fits the range points. This yields the superquadric transformation and shape parameters in the image coordinates (see Figure 5 (c)). These values are then transformed into the scanner robot task frame by using the scanner-to-world transformation matrix. The resulting superquadric is put into canonical form, with  $a_x \leq a_y \leq a_z$  and with the  $a_x$ ,  $a_y$  and  $a_z$  values bound to  $h$ ,  $w$  and  $l$  depending on the pose of the object to gravity. It is assumed that the object is placed on a flat surface perpendicular to the gravity vector. An object-centered action frame is chosen so that the  $\theta$  and  $\phi$  are relative to the direction of the  $w$  and  $h$  major axes and aimed at the visual centroid of the object.

The success or failure of a grasp is assessed by using the force-torque sensors on the gripper. During the grasp attempt, the system moves the gripper fingers inward from their full open position to an inter-finger distance of 5mm (see Figure 6). After the approach and grasp phase, the robot moves the gripper upwards, hopefully lifting the held object. It then measures the normal forces on the gripper fingers. If the force is below a threshold value, then the fingers are characterized as empty, otherwise if there is a large normal force, the grasp is labeled a success.



*Figure 5.* Acquired 3 dimensional Range Points (a) for the stapler, the sub-sampled and eroded points (b) and the resulting Superquadric fit to the range points (c).



*Figure 6.* The gripper is a simple two finger gripper with force torque sensors to sense grasp force. During the grasp attempt the hand attempts to close the fingers to their minimum aperture. The finger motion is guarded, so that if an above threshold force is detected, the motion stops. After a grasp attempt, the hand is brought upward and the force sensors are used to measure whether the object has been successfully grasped by again monitoring the reaction forces on the fingertips. Compliant foam is placed under the object to prevent damage to the gripper or object in situations where the fingers would be stubbed and possibly damage the object or gripper.

### 3.2. *Experimental Protocol*

In summary, the operation of the system consists of the following sequence:

1. The laser scanner scans the object.
2. Data reduction algorithms refine the data and produce the  $h, w$  and  $l$  of attributes that describe the object (Solina & Bajcsy, 1990).
3. A full decision tree is regenerated based on the database of points using the ID-3 algorithm for real-valued attributes (Quinlan, 1986, Quinlan, 1992)<sup>1</sup>.
4. The tree is pruned ( $\alpha = 0.025$ , see below). In practice, the tree pruning parameter did not seem to have a significant impact on system performance for reasonable values.
5. Interval estimation is used (with confidence level  $\alpha = 0.125$ ) to select the best  $\theta, \phi$  using the approach described in section 2. Here  $P = (h, w, l)$  and  $A = (\theta, \phi)$ .
6. The robot attempts the grasp and assesses the outcome.
7. The new data point  $(P, A, O)$ , of attributes and the outcome, is recorded in the database.
8. Goto step 1.

Pruning is done using Quinlan's (Quinlan, 1992) C4.5 pessimistic binomial upper-bound method. For each leaf the  $(1 - \alpha)$  upper-bound on the estimated error rate is computed



Figure 7. The set of common objects for which learning was attempted.

based on the misclassification incidences for each of the children nodes below a node being evaluated for collapse. If the weighted average of the upper-bound error rates is greater than the pooled upper-bound error rate, then the child nodes are collapsed.

### 3.3. *Experimental Results*

A total of six objects were chosen for grasping: a balsa block, white-board eraser, small flashlight, wire spool, screwdriver and a stapler (see Figure 7). For each of these objects, the protocol of section 3.2 was carried out, starting with an empty learning set for each new item.

Figure 8 shows the average success (or hit) rate over the thirty trials for the balsa block. The hit rate is the windowed proportion of successful grasps to attempts using a sliding window 10 attempts wide over the 30 instances in a learning run. For example, the computed hit rate for trial 10 consists of the proportion of consecutive grasp attempts numbered 10 through 19 that succeeded. It can be seen that the robot rapidly converges on a strategy for selecting approaches that succeeds reliably.

The scatter plot of Figure 8 also illustrates the effect of the exploration strategy. This figure shows the selected azimuth  $\theta$  and elevation  $\phi$  values for the balsa block trials in the range  $0^\circ < \theta < 90^\circ$  and  $0^\circ < \phi < 60^\circ$ . Each point in the  $\theta, \phi$  space is marked by its outcome, 0 for failure or 1 for success. As is evident by the plot, the action parameters are not uniformly distributed in the space, but concentrated in the region which has a small  $\theta$  value, which is the constraint for alignment for this object. Outside of the constraint values there are very few grasping attempts. The exploration algorithm has rapidly focused exploration on the region of success.

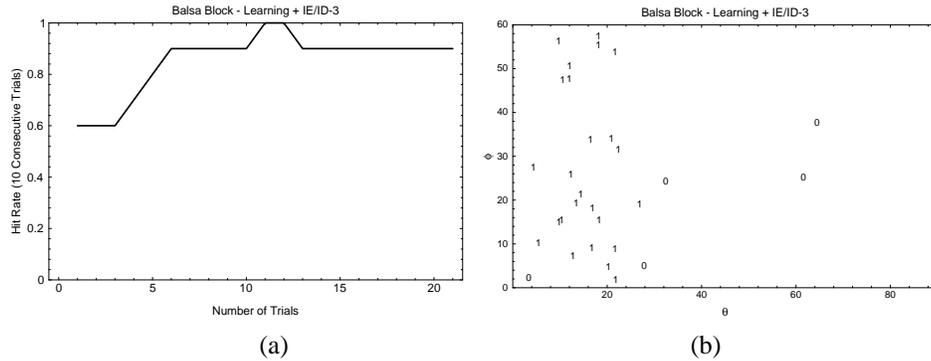


Figure 8. The hit rate for the balsa block over 30 trials and (b) corresponding scatter plot of chosen actions in the action space. The cumulative probability for successful grasping is .83 for the learning condition and .43 for random approach.

In the case of the flashlight, the learning curve increases more slowly (see Figure 9) since the width of the screwdriver is quite close to the maximum width of the gripper and it takes more attempts to find a first successful grasp which can form the basis of a tree partition.

The learning curve for the stapler is rapid (see Figure 10), converging after about 20 attempts. The system succeeds almost at once and maintains a high level of performance. The scatter plot once again shows that the actions are highly concentrated in the favorable part of the action space. Similar learning performance and action distributions were observed for the eraser and screwdriver objects.

The spool (rightmost object in Figure 7) was an interesting case, since it has a significant concavity from the end-pieces attached to the center shaft, which makes it difficult to represent accurately using the convex superquadric representation that we have chosen. Nevertheless, the learning algorithm is able to learn the constraints for grasping this object (see Figure 11), and identifies a region in the action space that succeeds (namely  $\phi < 30$  degrees) by picking an approach direction which avoids approaches that hit the end-cap, which was the main cause of failures in grasping this object. Unfortunately, the system cannot describe concavity with its given shape representation. This may lead to the generalization that other objects with similar parameter values have constraints for approach directions that are overly restrictive when compared to a truly convex object. Another scenario has a convex object with the same extents as the spool presented first in sequence. This will lead to approach constraints which are not restrictive enough, and upon interactions with the spool, the system may err by selecting poor approach directions. Upon further interactions, the system will begin to restrict the approach directions. Unless the representation is extended to include some measure of concavity, this problem cannot be avoided. The only solution is to adopt more sophisticated representation for perceptual parameters into the active learning approach (see section 4).

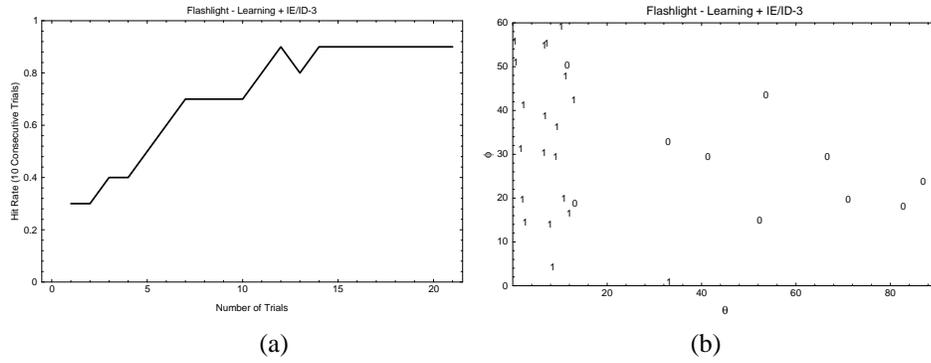


Figure 9. The hit rate for the flashlight over the first 30 trials and (b) the distribution of actions in the action space. The cumulative probability for successful grasping is .67 for the learning condition and .23 for random approach.

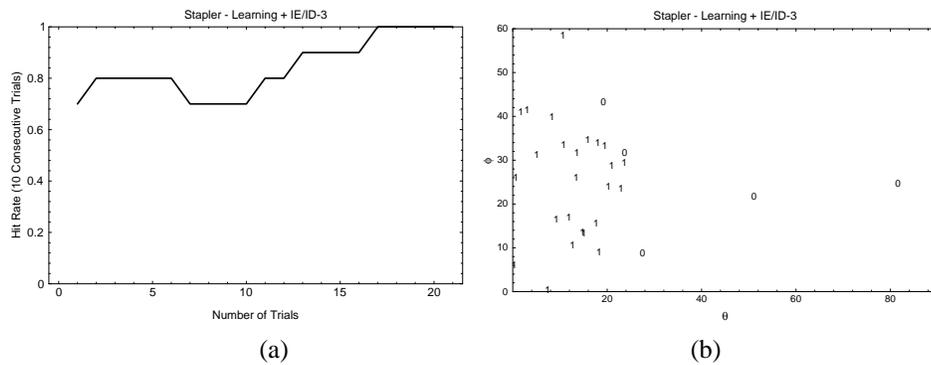


Figure 10. The hit rate for the stapler over the first 30 trials and (b) the distribution of actions in the action space. The cumulative probability for successful grasping is .83 for the learning condition and .4 for random approach.

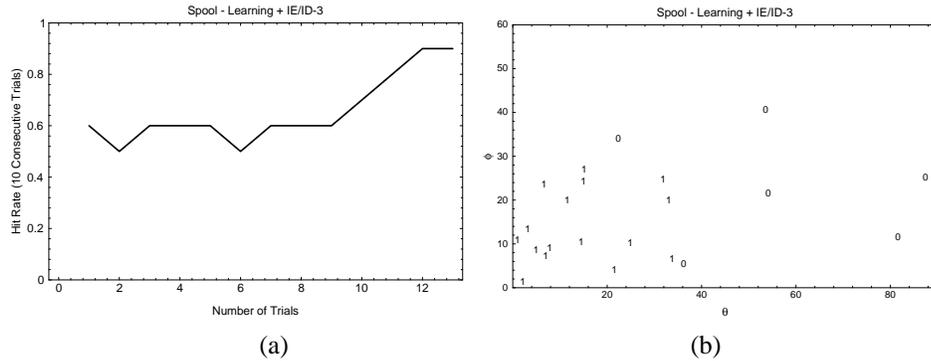


Figure 11. The hit rate for the spool over the first 25 trials and the distribution of actions in the action space. The cumulative success probability is .8 for the IE-ID3 learning condition and .27 for the random condition.

### 3.4. Simulation Studies

A simulation study was undertaken to explore the scaling properties as the number of objects presented increased. In the simulation the gripper was modeled by taking the necessary coordinate transformations on the gripper of rotation( $\theta, z$ ) followed by rotation( $\phi, x$ ) and determining if the projection of the object in the two dimensional plane orthogonal to the approach direction was smaller than the maximum span of the fingers. If the gripper cleared the object, it was taken to be graspable.

The objects for the simulation had dimensions that were selected from a uniform distribution of 35-90mm and the maximum gripper width was taken as 70mm. Additionally, to make the task more difficult, objects were generated so that only one of the aspects of the object,  $w$  or  $l$  would clear the gripper. For the aspect that did clear the gripper, a minimum of 5mm clearance was guaranteed. The allowable approach directions were in the range  $0^\circ < \theta < 90^\circ$  and  $0^\circ < \phi < 90^\circ$ .

For the IE-ID3 algorithm, the learning protocol was to give a randomly generated object from the above distribution to the learner and then have it attempt to grasp the object 10 times, or until it succeeded 5 times in a row, whichever came first. Each trial consisted of running ID-3 on the exemplars generated so far and then selecting an action using the IE algorithm on the resulting tree. For the simulation curves, the horizontal axis represents the interaction number as they occur in each independent execution epoch, and the vertical axis represents the estimated hit rate (the probability of a grasp attempt succeeding) based on the average success rate at each trail interaction in the epoch. The estimated hit rate (success) is computed by counting up all successful attempts for the given trial numbers across all epochs and dividing by the total number of epoch runs.

For comparison, open loop (OL) alternating learning was performed in which each cycle consisted of first executing a random action to acquire the next point in the exploration set. Then ID-3 was run on that exemplar plus any previously acquired exemplars and a test action was executed to form a trial for the performance curve. The test action was chosen by

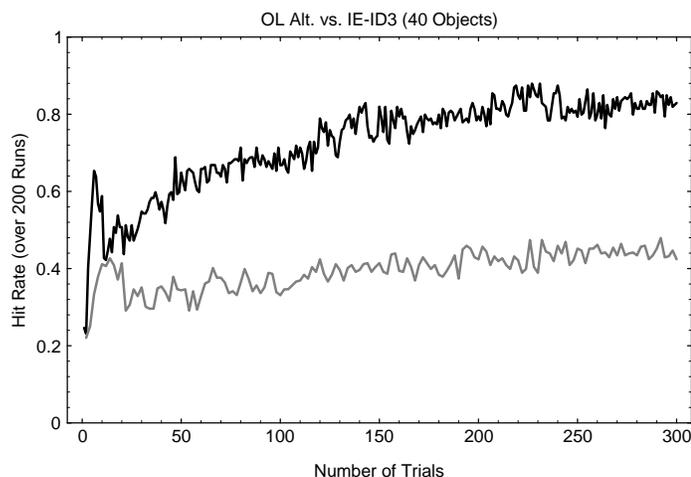


Figure 12. Comparison of system performance over first 300 trials (averaged over 200 runs). The IE-ID3 driven learner (black) exhibits much better performance than the open loop learner (gray) which suffers because it always alternates between the random training action mode and test.

searching for the leaf with the highest raw probability of success given the current perceptual vector constraint  $P$ . The OL plot depicts the average of the exploration and exploitation outcomes. As Figure 12 shows, the IE-ID3 based learner (the black curve) performed almost twice as well the OL alternating strategy. This large difference in performance was primarily due to the fact that in the OL case 50% of the grasp attempts were randomly chosen because of the random selection used in OL.

Figure 13 depicts the cumulative average number of successes for three conditions of learning: IE-ID3, OL and random. Again we can see that the performance of the IE-ID3 mode learning far exceeds the alternating OL explore/exploit paradigm again because of the wasteful random search of the open loop learner. This can be seen in Figure 14 which shows the performance curve (black) and exploration curve (light grey). Since the performance on the exploration is low, the combined performance of the alternating explore/exploit strategy is quite low; as expected, random exploration is expensive. IE-ID3 learning performs much better since it begins to exploit knowledge and abandon exploration as soon as is judged possible. Subsequent actions are guided to focus on more promising parts of the action space. Figures 15(a) and (b) give representative projections of the distribution of actions for the IE-ID3 and OL conditions along the  $\theta$  and  $l$  axes and show that, as in the actual experiments, the data is concentrated in the high success regions for the IE-ID3 condition, as expected.

We would expect that along with better on-line performance, the classification accuracy for the active learner would be better than the open loop, since it would concentrate its data gathering in relevant regions of the input space, leading to better approximations of task-relevant concept decision boundaries by the learner. In order to test this hypothesis,

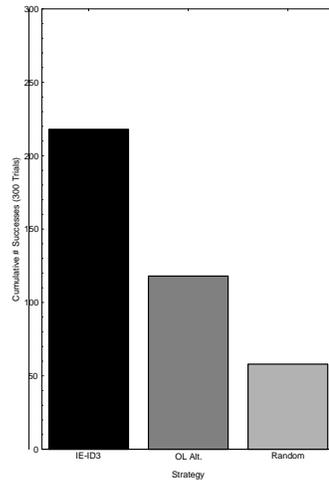


Figure 13. Comparison of the cumulative number of grasp successes for the IE-ID3, alternating explore and exploit open loop (OL) and random conditions over 300 trials.

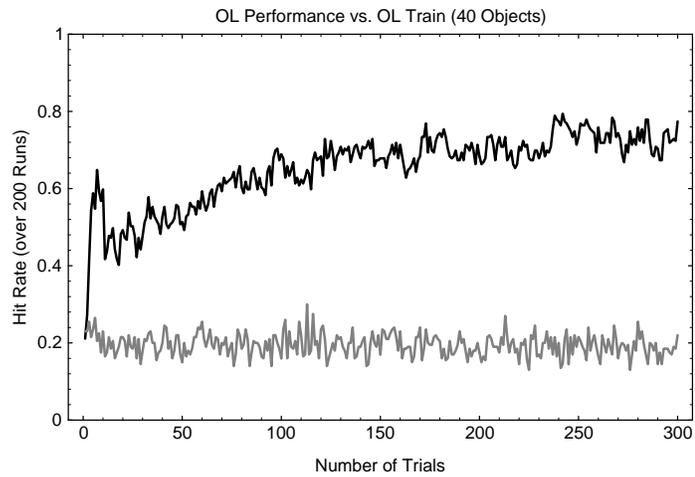


Figure 14. The exploration and exploitation performance of the open loop learner. Since these two strategies alternate, the combined performance is penalized by the random exploitation.

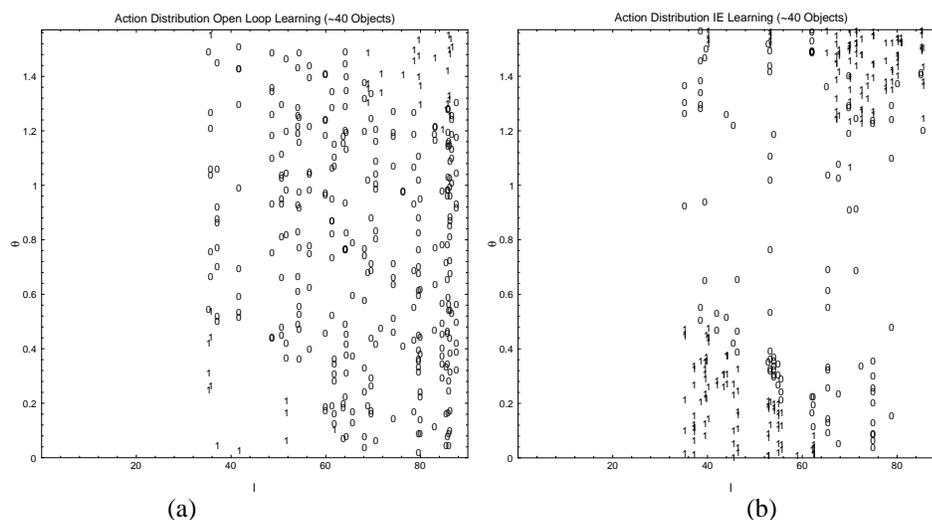


Figure 15. The projection of the location of actions ( $\theta$ ) vs. the length aspect of the object ( $l$ ). We can see that the exemplars generated by the active learning case (b) are much more concentrated in the success regions than the open loop case (a).

we ran the OL strategy and measured the probability of successful classification during the performance cases of each cycle. We also formed an equivalent independent test set for IE. Although a test set is technically unnecessary while using the IE-ID3 method, for the purposes of comparison it is useful to examine the use of raw probabilities for action selection of both methods. Figure 16 shows that the IE-ID3 formed classification is more accurate. In another test, IE-ID3 versus OL were run over 250 interactions using the same protocol as above. Looking at the average number of correct classifications for the 250 interactions for IE-ID3 ( $\mu_{OL} = 151.121$  and  $\mu_{IE} = 175.12$ ) showed that IE-ID3 lead to significantly higher number of correct classifications. The 99% confidence interval for the difference between the means  $\mu_{IE} - \mu_{OL}$  was  $[18.35, 29.4657]$  using a Student's-t small sample confidence interval test. Thus, the simulation results indicate that IE-ID3 does somewhat better than OL from a classification standpoint, but is markedly better from a *performance* standpoint, which is the key issue for real-world efficiency.

#### 4. Discussion

We adopted the framework of the two phase model of grasping, and used learning to determine free parameters for the ballistic approach phase that determines the direction of approach of the robot hand. Our results show that the IE algorithm provides an efficient mechanism for focusing the data-gathering activity of an ID-3-based learner on successful areas of the action space. The method learns to grasp new objects very rapidly, usually within about 20 attempts for this simplified problem. The scatter plots for both the real

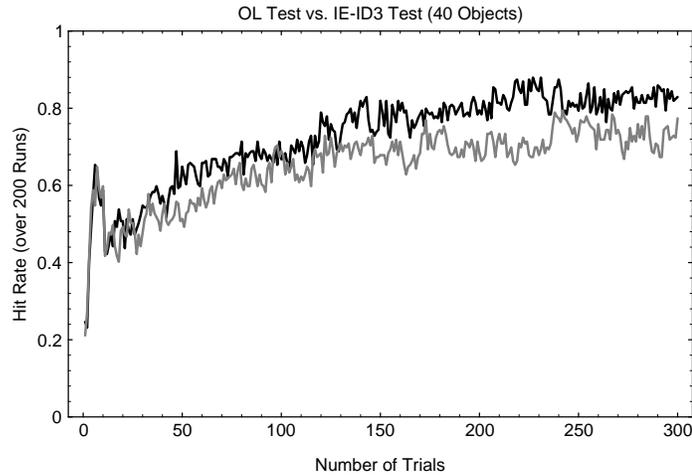


Figure 16. Comparison of the Classification Rate for IE-ID3 (black) Test vs OL. Test. (Grey)

learning trials and the simulations illustrate the fact that actions are concentrated in the high success areas of the attribute space.

#### 4.1. Extending the Approach to More Complex Objects and Effectors

Our work represents a first step, where we are controlling only the approach direction of the gripper relative to the centroid of the object, and holding other parameters fixed. Our active learning approach to grasping performed well with a small parameter set, but further trials with richer perceptual descriptions are needed to verify to the approach with the more complex representation needed to represent general real world objects. In particular as more complex objects with multiple parts and other protuberances are presented for grasping, a global deformation approach such as the superquadric will smooth out these possibly important features (see the discussion of performance on the spool test object in section 3.3). Objects that might be graspable from handles, but too wide to clear the gripper might be found ungraspable using the global deformation approach of superquadrics, leading to decreased performance.

Fortunately other superquadric shape parameters and more sophisticated shape representations including taper and bending parameters for superquadrics could also be used within this framework. In particular, the multi-component aspect polyhedron described used by Stansfield (Stansfield, 1990) seems a particularly suitable candidate. This representation supports symbolic 3-D attributes of object patches such as curved or flatness properties, as well as 3-D positions and extents in space and other predicates that express the spatial relations between patches for a given view. These views from different directions of the object can be combined to form an aspect polyhedron. The aspect polyhedron permits

the representation of objects with multiple parts, rather than the single global deformations permitted by superquadrics. It allows a more sophisticated representation of both of perception and action, since grasp sites can be associated with segmented surface patches on the object. In order to apply learning to this representation, the different slots of the multi-part perceptual representation could be bound to the input attributes of the decision tree (similar to the input bindings to the input nodes in (Stansfield, 1992)) but using the same protocol and active learning approach described in this work instead of an off-line back-propagation approach. A similar approach could be taken with Bard's hierarchical ellipsoid representation (Bard, Troccaz & Vercelli, 1991), if it were augmented with spatial-relation predicates between the component ellipsoids.

The IE-ID3 approach should scale well if these more sophisticated representations are adopted since all of its components are known to scale well. The ID-3/CART family of classification tree algorithms have proven effective in applications with much higher dimensionalities than what we have used here, and support both real-valued and categorical variables, both of which may be present in grasping tasks. Future implementations should include symbolic values, such as the grasp preshape (e.g. pinch, hook, cylindrical, spherical etc.) for more complex dextrous hands, along with real-valued parameters for those grasp categories such as finger stiffness and extent of preshape. Additionally, ID-3 and related techniques such as CART and newer non-axis parallel techniques (Murthy, et al., 1993) are very well-suited to rejecting irrelevant attributes because of the information gain splitting heuristic utilized, and so should scale well to more complex representations. Another advantage of using classification trees is that although the action space is continuous, using the indirect indexing method with the ID-3 partitioning affords a natural and meaningful partition of that space. The resulting intervals provide a range of feasible action which can be intersected with other high-level constraints, such as approach limits that might be due to obstacles around the object.

Further phases of grasping that follow the preshape and approach would also benefit from active learning. These include the contact phase and lifting phases (Howe, et al., 1990), re-grasping, transport phases and release or insertion for assembly (Gullapalli, Barto & Grupen, 1994), each of which is indicated by specific sensory markers, such as tactile forces, that invoke differing control modes. An interesting problem is to learn reliable markers from this stream of sensory information that would indicate state transitions, as well as the correct control parameters that would hold during these states, as has been demonstrated in some initial work by (Pook & Ballard, 1993) using hidden Markov models.

One possible drawback of the approach is that it is possible that other potentially successful areas of the perception/action space might be missed, since the method could greedily identify a good region and cease exploration in other regions. For example, an initial split based on the trials seen so far might yield a partitioning which covers only a subset of the true success region. Subsequent attempts would then be focused in that restricted region. There is no way to guarantee not missing a small region with higher success probability unless one is willing to exhaustively explore the entire perception-action space. Generally, the space is sparse in success for non-trivial problems, and it is not worth the cost of doing so.

Another possibility is that the IE algorithm may get stuck in actions which are sub-optimal due to pathological data orderings (Kaelbling, 1990). Consider the case two actions,  $a_1$  and  $a_2$ , with  $P_{success}(a_1) > P_{success}(a_2)$ . When sampling  $a_1$  and  $a_2$  sequentially using the IE strategy it is possible that a pathological ordering of outcomes may occur leading to  $P_+(a_2) > P_+(a_1)$  and  $P_+(a_1) < P_{success}(a_2)$  at the same time. In this case the algorithm will be stuck in a vicious cycle:  $a_2$  will always be selected, since its upper-bound will converge to  $P_{success}(a_2)$  and  $a_1$  upper-bound will never rise since it is never selected since  $a_2$  has the best upper-bound. Sticking may also occur in Bayesian-optimal systems; it is an unavoidable consequence of the exploration/exploitation tradeoff. Fortunately, sticking becomes less probable as the difference in the underlying probability of success between actions increases. Therefore, for practical purposes, if the sticking phenomenon occurs, it is likely that the efficacies of the different alternatives will be small.

#### 4.2. Comparison to Other Methods

Active learning approaches have taken two overall approaches. Knowledge-directed approaches select actions which maximize overall knowledge gain about a domain model, independent of a particular task. Task-directed approaches concentrate on actions that yield a maximum of information about the domain that is relevant to a particular task in a domain, rather than a general domain model. For example, Cohn has proposed two methods for control of experimentation to maximize overall information gain. Selective sampling (Atlas, et al, 1990) attempts to identify the region of uncertainty between the most general and most specific concepts that are consistent with the training set. The distribution of query points is then biased to favor those points that lie within the region of uncertainty. Actions can also be chosen to be maximally informative in terms of minimizing the expected generalization error over the entire domain, for example, using optimal experimental design (Cohn, 1994). Whether a task-directed or knowledge-gain directed gain approach is taken for learning is a function of the costs of experiments and the expected amount of the domain that will be visited by the learning agent during its lifetime, and how desirable it is to front-load learning of the entire domain, versus learning it on an as needed basis as new tasks are encountered.

The system presented is the first grasp learning system to incorporate active learning to intelligently select manipulatory actions. Previously developed grasp learning systems adopt random action search, or rely of domain specific heuristics. For example in Dunn & Segen (1988) if no similar object is found in the data base, a randomized trial and error search for successful grasps is undertaken. In Kamon et al. (1994) if no match is found, a domain dependent heuristic is used to pick a grasp with a random component. Similarly in Tan (1993), although consideration is given towards picking good sensing actions, manipulatory actions are chosen randomly if no classification is predicted. Bennett's explanation based system (Bennett, 1991) uses domain knowledge to choose its next grasping attempt, which requires explicit human provided knowledge regarding the general characteristics of the sensing and motor apparatus of the robot, limiting its flexibility and true autonomy. Our approach represents a domain-independent and principled technique for selecting actions.

Our work can be compared Tan's Cost Sensitive Learning (CSL-ID3) (Tan, 1993) at several levels. Both CS-ID3 and IE-ID3 are similarly concerned with decreasing the cost of learning, but differ philosophically on how they decrease this cost. If one examines the overall costs encountered by a learning system during its lifetime, these costs can arise either from measuring the attributes of each case processed, or due to errors that occur by acting on the wrong classification output by the learning system (executing the wrong action). Consider a hypothetical system that attempts to learn to grasp and retrieve Ming Dynasty vases. Certainly, we would want the system to converge as quickly as possible, since each dropped vase has very high penalty. Secondly, if certain perceptual features, such as the color of the vase, were rapidly computable and relevant to deciding where to grasp the vase, we should prefer those over difficult to sense and unreliable object features. If the cost of failure is high in comparison to sensing costs, then we want the system improve its classification accuracy on encountered cases as rapidly as possible. This can be done by intelligent selection of the next action, as in IE-ID3 and other approaches (Thrun & Moller, 1992, Moore, 1990, Schneider, 1993, Sutton, 1990, Atlas, et al, 1990, Cohn, 1994) so as to improve the classification accuracy as much as possible per each fixed cost exemplar. Until it reaches convergence CS-ID3 seems to require many more interactions with the object than a standard ID-3 and encounters many more classification errors than a standard ID-3 approach ((Tan, 1993),pp. 27), which can be problematic in tasks with high misclassification costs. On the other hand, if misclassification costs are small in comparison to sensing costs, the CSL-ID3 approach of decreasing average sensing costs makes sense. In fact, CS-ID3 achieves rough parity with ID-3 regarding total cost until convergence by decreasing the average sensing cost per each case.

Although initially developed as a supervised learning system (Tan, 1990) CSL-ID3 can technically be run in either a supervised or unsupervised mode (Tan, 1993), where either a teacher suggests the appropriate grasp for a sensed object, or the robot randomly tries grasps from a finite set of available alternative grasping procedure until a grasp succeeds. However if size of the action set is large, then the random experimentation approach can be unduly expensive and even hopeless. Secondly, if an object is not graspable by any of the preset grasp actions in CS-ID3, the system cannot synthesize a new action that succeeds. The IE-ID3 approach, on the other hand, allows for continuous actions, which do permit truly novel actions to be discovered. Secondly, because of the probabilistic nature of the confidence intervals we use, we can have a measure of reliability of an action, which is not captured the recognition tree generated by CSL-ID3

Since CSL-ID3 achieves a decrease in the on-going sensing costs of the learning agent after it has converged, it can lead to much lower costs during the lifetime of the system. To truly minimize costs during the life of the agent, the sensing and misclassification costs, as well as the expected number of times the gathered knowledge will be used (e.g. the lifetime of the agent) must be taken into account. An interesting avenue of future research would be to combine a technique such as CS-ID3 with IE-ID3 in a multi-objective approach, taking into account both of the above considerations.

### **4.3. *Extending the IE-ID3 Approach to Other Learning Algorithms***

The key to the approach described here is the combination of learning mappings from perception and action to reward with selecting actions which improve the expected reward over all future trials. Above, we used ID-3 to build decision tree mappings and the IE algorithm to select actions. Other learning methods such as radial basis functions, nearest-neighbor approaches and hierarchically gated mixtures of Gaussians could equally well be used to build the mappings since they also can, in principle, provide confidence intervals for the quantities they predict.

One immediate extension would be to apply the IE-ID3 approach toward learning function with continuous outputs. This could be done by employing the real-valued extension to IE as suggested in (Kaelbling, 1990), (pages 74-75), along with using regression trees (Breiman, et al., 1984) in place of classification trees. The real-valued extension for IE-ID3 involves building confidence intervals for the estimate of the true mean reward for the learner, again picking actions with the highest output upper-bound. The statistics for the upper-bound would be available in the nodes of the regression trees.

Additionally, one can use exploration approaches other than IE, for example, those suggested by statistical decision theory (Berger, 1985). These can incorporate priors on the distribution of reward in the attribute space in order to control exploration in a less heuristic fashion than IE, and allow a more formal analysis of the properties of the IE approach. Also, a wealth of results from multi-armed bandit problems (Gittins, 1988) become applicable once the real-valued space is partitioned and can lead to even more efficient solutions to problems (Salganicoff & Ungar, 1995).

### **4.4. *Other Application Areas***

We expect the IE-ID3 method to be broadly applicable in robotics and in many other domains. A variety of other robotic skills have been subjected to learning approaches through parameterization of the perceptual and action spaces, such as throwing (Schneider, 1993), juggling (Aboaf, Atkeson & Reinkensmeyer, 1988) and playing billiards (Moore, 1991a) and the IE-ID3 active learning approach should be applicable those domains as well.

Chemical plant control and optimization provides a second example of a class of problems with the right characteristics for such active learning, particularly IE-ID3. The plants are often nonlinear, and complete, accurate, models are rarely available. Many control actions can be taken - e.g. adjusting the temperatures, pressures, and flow rates of the plant. It is far too expensive to explore the whole range of possible actions (controls), since most of them are suboptimal - and expensive. Thus, it is necessary to trade off exploration of actions against exploitation of the best actions found to date. Similarly, in many batch production environments, such as semiconductor processing, it is difficult or overly expensive to build detailed and accurate models of the production process that may permit a principled enhancement of yield as a function of the variation of the process parameters. By using the IE-ID3 active learning approach an efficient search of the parameter space of the process may be made, even for environments with combinations of real-valued and categorical parameters.

## 5. Conclusion

We have presented a technique for active learning which combines the Interval Estimation (IE) exploration heuristic (Kaelbling, 1990) with a classification tree algorithm, ID-3 (Quinlan, 1986). We demonstrated the success of this technique in selecting parameters for the grasp approach direction during the ballistic phase of grasping. The active learning approach was motivated by the fact that gathering each exemplar (interaction with an object to attempt a grasp) has significant costs, similar to many learning tasks. Thus, minimizing the number of trials before good performance is attained is crucial. The resulting learner avoided random exploration and focused its data acquisition areas with effective actions. Empirical tests on a small set of test objects using a laser ranger scanner and simplified superquadric shape descriptions of the objects showed that the system was able to learn to select successful approach direction rules in a small number of real interactions with the objects. Simulation results for a scenario with many more objects and trials showed that using the active learning exploration strategy dramatically improved performance over an uninformed batch learner. The representation and learning algorithms employed are extendable to more sophisticated perceptual and action representations for grasping. Although the approach is promising, much work remains to be done in verifying the scalability of the approach through experiments and in tackling learning as it applies to other aspects of vision-based robotics and grasping.

## Acknowledgments

This work was supported by Navy Grant N00014-92-J-1647, AFOSR Grants 88-0244, AFOSR 88-0296; Army/DAAL 03-89-C-0031PRI; NSF Grants CISE/CDA 88-22719, IRI 89-06770, and ASC 91 0813 and an NSF HPPC Postdoctoral Associateship for MS (CDA-9211136) at the University of Pennsylvania. Applied Science and Engineering Laboratories funding was provided through the Rehabilitation Engineering Research Center on Rehabilitation Robotics—National Institute on Disabilities and Rehabilitation Grant #H1333E30013 and Nemours Research Programs. We also thank Len Kunin for his help in implementation of the system.

## Notes

1. The ID-3 algorithm with single partitions per attribute has worst case time complexity of  $O(n^{2+\log k})$  (this solves the recurrence relation  $T_n = 2T_{n/2} + kn^2$ ) where  $n$  is the number of exemplars, and  $k$  is the number of attributes. In practice performance is much better, since the above analysis assumes a learning set that results in a full binary tree partitioning of one leaf for each exemplar in the set. Truly incremental versions of ID-3 such as ID-5 (Utgoff, 1988) could have been used, but the non-incremental algorithm executed rapidly enough that it was not a hindrance to regenerate the full tree after each new observation.

## References

- Aboaf, E.W., Atkeson, C.G. & Reinkensmeyer, D.J. (1988). Task-level robot learning. In *Proceedings, 1988 IEEE International Conference on Robotics and Automation*, pages 1290–1295. IEEE Press, 1989.
- Atlas, L., Cohn, D., Ladner, R., El-Sharkawi, M.A., Marks II, R.J., Aggoune, M.E. & Park, D.C. (1990). Training connectionist networks with queries and selective sampling. In *Advances in Neural Information Processing Systems 2*, pages 567–573. Morgan Kaufmann.
- Bard, C., Laugier, C., Milesi, C. & Troccaz, J. (in press). Achieving dextrous grasping by integrating planning and vision based sensing. *The International Journal of Robotics Research*, 1995.
- Bard, C., Troccaz, J. & Verceles, G. (1991). Shape analysis and hand preshaping for grasping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 64–69.
- Bennett, S (1991). Planning to address uncertainty: An incremental approach employing learning through experience. In *Proceedings of the DARPA Workshop on Advanced Planning and Control Strategies*, pages 313–324.
- Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. Springer-Verlag, NY.
- Box, G.E.P. & Draper, N.R. (1969). *Evolutionary Operation: a Statistical Method for Process Improvement*. Wiley.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and regression trees*. Wadsworth International, Belmont, CA.
- Cohn, D.A. (1994). Queries and exploration using optimal experiment design. In *Neural Information Processing Systems 6*, Denver, CO. Morgan-Kaufman.
- Cutkosky, M.R. (1989). On grasp choice, grasp models and the design of hands for manufacturing tasks. *IEEE Journal of Robotics and Automation*, 5(3):269–279.
- Dunn, G.B. & Segen, J. (1988). Automatic discovery of robotic grasp configuration. In *Proceedings, 1988 IEEE International Conference on Robotics and Automation*, pages 396–401.
- Gittins, J.C. (1988). *Multi-armed Bandit Allocation Indices*. John Wiley and Sons, Chichester.
- Gullapalli, V., Barto, A.G. & Gruben, R.A. (1994). Learning admittance mappings for force-guided assembly. In *Proceedings, 1994 IEEE International Conference on Robotics and Automation*, pages 2633–2638.
- Howe, R.D., Popp, N., Akella, P., Kao, I. & Cutkosky, M.R. (1990). Grasping, manipulation and control with tactile sensing. In *Proceedings, 1990 IEEE International Conference on Robotics and Automation*, pages 1258–1263.
- Iberall, T. Jackson, J., Labbe, L. & Zampano, R. (1988). Knowledge-based prehension: Capturing human dexterity. In *Proceedings, 1988 IEEE International Conference on Robotics and Automation*.
- Jeannerod, M. & Biguer, B. (1981). *Advances in the analysis of visual behavior*, chapter Visuomotor mechanisms in reaching within extrapersonal space. MIT Press, Boston.
- Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Clarendon Press, Oxford Science Publications, Oxford.
- Kaelbling, L.P. (1990). *Learning in Embedded Systems*. PhD thesis, Stanford University, Stanford, CA. Dept. of Computer Science.
- Kamon, I., Flash, T. & Edelman, S. (1994). Learning to grasp using visual information. Technical Report, The Weizmann Institute of Science, Rehovot, Israel.
- Kang, S. & Ikeuchi, K. (1993). Toward automatic robot instruction from perception-recognizing a grasp from observation. *IEEE Journal of Robotics and Automation*, 9(4):432–443.
- Kuniyoshi, T., Inaba, M. & Inoue, H. (1989). Teaching by showing: Generating robot programs by visual observation of human performance. In *Proceedings of 20th International Symposium on industrial robotics*, pages 119–126.
- Larson, R. & Marx, M. (1986). *An introduction to mathematical statistics*. Prentice Hall, Englewood Cliffs, NJ.
- Liu, H., Iberall, T. & Bekey, G.A. (1989). The multi-dimensional quality of tasks requirements of task requirements for dextrous hand control. In *IEEE International Conference on Robotics and Automation*, pages 452–457.
- Moore, A.W. (1990). Acquisition of dynamic control knowledge for a robotic manipulator. In *Proceedings of the seventh international conference on machine learning*.
- Moore, A.W. (1991a). Fast, robust adaptive control by learning only forward models. In *Advances in Neural Information Processing Systems*. Morgan Kaufman.
- Moore, A.W. (1991b). Variable resolution dynamic programming: Efficiently learning action maps in multivariate real-valued state-spaces. In *8th International Workshop on Machine Learning*. Morgan Kaufman.

- Moreno, C.W. & Yunker, S.B. (1992). Ultramax: Continuous process improvement through sequential optimization. Technical report, Electric Power Research Institute.
- Murthy, S., Kasif, S., Salzberg, S. & Beigel, R. (1993). Oc1: Randomized induction of oblique decision trees. In *Proceedings of the 11th National Conference on Artificial Intelligence*, pages 322–327, Menlo Park, CA, USA. AAAI.
- Napier, J.R. (1956). The prehensile movements of the human hand. *Journal of Bone and Joint Surgery*, 38b(4):902–904.
- Pook, P.K. & Ballard, D.H. (1993). Recognizing teleoperated manipulations. In *Proceedings, 1993 IEEE International Conference on Robotics and Automation*, pages 578–585.
- Quinlan, J.R. (1986). *Learning efficient classification procedures and their application to chess end-games*, pages 463–481. Morgan-Kaufman, Los Altos, CA.
- Quinlan, R. (1992). *C4.5: Programs for Machine Learning*. Morgan-Kaufmann, Los Altos, CA.
- Rumelhart, D.E., Hinton, G.E. & Williams, R. (1986). Learning internal representations by back-propagating errors. *Nature*, 323:533–536.
- Salganicoff, M. (1992). *Learning and Forgetting for Perception-Action: A Projection-Pursuit and Density-Adaptive Approach*. PhD thesis, University of Pennsylvania. Dept. of Computer and Information Science.
- Salganicoff, M. (1993). Density-adaptive learning and forgetting. In *Proceedings of the Tenth International Conference on Machine Learning*. Morgan-Kaufman.
- Salganicoff, M. & Bajcsy, R. (1992). Robot sensorimotor learning in continuous domains. In *1992 IEEE Conference on Robotics and Automation*, Nice, France.
- Salganicoff, M. & Ungar, L.H. (1995). Active exploration and learning in real-valued spaces using multi-armed bandit allocation indices. In *Twelfth International Machine Learning Conference*.
- Schlesinger, G. (1916). *Der Mechanische Aufbau der Kunstlichen Glieder*. Springer, Berlin.
- Schneider, J. (1993). Open loop motor skill learning. In *AAAI Fall Symposium Series: Machine Learning in Vision: What Why and How?*, Raleigh, N.C.
- Solina, F. & Bajcsy, R. (1990). Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–147.
- Stansfield, S. (1990). Knowledge-based robotic grasping. In *Proceedings, 1990 IEEE International Conference on Robotics and Automation*, pages 1270–1275.
- Stansfield, S.A. (1992). Connectionist and neural net implementations of a robotic grasp generator. In *Proceedings of SPIE - Engineering Applications of Artificial Intelligence X: Machine Vision and Robotics*, volume 1708, pages 375–388, Bellingham, WA, USA. Int. Soc. for Optical Engineering.
- Sutton, R.S. (1988). Learning to predict by the methods of temporal differencing. *Machine Learning*, 3:9–44.
- Sutton, R.S. (1990). Integrated architectures for learning planning and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning*. Morgan-Kaufman.
- Tan, M. (1990). A cost-sensitive learning system for sensing and grasping objects. In *Proceedings, 1990 IEEE International Conference on Robotics and Automation*. IEEE Press.
- Tan, M. (1993). Cost-sensitive learning of classification knowledge and its applications in robotics. *Machine Learning*, 13(1):7–33.
- Taylor, M., Blake, A. & Cox, A. (1994). Visually guided grasping in 3D. In *Proceedings, 1994 IEEE International Conference on Robotics and Automation*, pages 761–766.
- Thrun, S. & Moller, K. (1992). Active exploration in dynamic environments. In *Neural Information Processing Systems 4*, pages 531–538. Morgan-Kaufman.
- Thrun, S.B. (1992). The role of exploration in learning control with neural networks. In *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*. Van Nostrand Reinhold, Florence KY.
- Tsikos, C. & Bajcsy, R. (1991). Segmentation via manipulation. *IEEE Transactions on Robotics and Automation*, 7(3):306–319.
- Utgoff, P. (1988). ID-5: An incremental ID-3. In *Proceedings of the Fifth International Workshop on Machine Learning*. Morgan-Kaufman.
- Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD thesis, King's College, Cambridge.