

**Active and Exploratory Perception**

**MS-CIS-91-91  
GRASP LAB 288**

**Ruzena Bajcsy  
Mario Campos**

**Department of Computer and Information Science  
School of Engineering and Applied Science  
University of Pennsylvania  
Philadelphia, PA 19104-6389**

**October 1991**

# Active and Exploratory Perception

Ruzena Bajcsy and Mario Campos  
GRASP Laboratory

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

October 28, 1991

## 1 Abstract

The main goal of this paper is to show that there is a natural flow from active perception through exploration to perceptual learning. We have attempted to conceptualize the perceptual process of an organism that has the top-level task of surviving in an unknown environment. During this conceptualization process, four necessary ingredients have emerged for either artificial or biological organisms. First, the sensory apparatus and processing of the organism must be active and flexible. Second, the organism must have exploratory capabilities. Third, the organism must be selective in its data acquisition process. Fourth, the organism must be able to learn. In the section on learning, we have clearly delineated the difference between what must be innate and what must be learned. In order to test our theory, we present the system's architecture that follows from the perceptual task decomposition. The predictions of this theory are that an artificial system can explore and learn about its environment modulo its sensors, manipulators, end effectors and exploratory procedures/attribute extractors. It can describe its world with respect to the built-in alphabet, that is the set of perceptual primitives.

## 2 Introduction and Motivation

That perception is active has been accepted by modern psychologists for at least the past 50 years. Most notably, J. J. Gibson has argued for this point of view [5]. Later, E. Gibson documented it via developmental studies in children [4]. As a most striking example, she cites comparative studies of children with Down's syndrome and normal children. The normal children seek perceptual information while the Down's syndrome children are passive vis-a-vis their environment; either they do not interact or interact less. We have recognized

the activity of perception [2] and have undertaken a research program to formulate the computational and engineering consequences for active machine perception.

Once we have accepted this paradigm, then exploration becomes a very natural task. The exploratory perceptual task, elaborated in section 3.1, becomes a necessity if an organism finds itself in an unknown environment. However, one needs to consider this task in the more general framework of perceptual task modeling, since the exploratory perceptual task is only a special case. In order to put our ideas on concrete footing, in section 5 we present the system architecture that is a test bed for experimenting and testing the ideas of exploration and learning, using vision and haptics. A concrete example of the Weight Exploratory Procedure is presented. When the organism explores, inevitably it is for the purpose of learning about its environment. Hence, learning and/or perceptual development follows from exploratory activity in a natural way.

In section 4. we ask what is innate so that the system can explore and accumulate the knowledge gathered through exploration. In principle, an organism can explore without remembering its experiences, although this is not economical. Since we assume that our system has memory, learning becomes inevitable based on this energy/economy argument. In fact, we use the energy/economy argument to show that an organism always acts with purpose.

We believe that perceptual learning is primarily inductive as opposed to deductive learning. The main point of this paper is to show the natural flow from active perception, through exploration to perceptual learning. The proposed theory is constructive, hence we shall outline the architecture that is necessary for the design of an active and exploratory system. Parts of the system have been implemented in the GRASP laboratory over the years and will be mentioned in reference.

### **3 What Do We Mean By Perception Being Active?**

This means for a perceptual system to actively seek information and not just rely passively on information falling accidentally on the sensor. This also means that the system must be mobile. In biological systems the mobility usually means mobility in space. In man-made systems, the mobility can be in other domains beside space, such as in the frequency or spectral domains. The mobility is closely related to exploration and selectivity. What we mean by this statement is that the data acquisition apparatus, being either contact or non-contact is attached to a mobile platform. This in turn allows the sensory system acquire data about the world from almost any arbitrary position/orientation. module accessibility of objects in this world. The perceptual system simply cannot explore without being mobile. Similarly, the perceptual system cannot select information unless it has a set of possibilities to select from. On the other hand the perceptual system must be selective, or it will suffer from an overflow of information.

¿From the above it follows that an active perceptual system must be mobile, and interact with its environment. During this process there could be two different cases:

1. if the system uses only non-contact sensors, such as audition or vision, this interaction does not alter the environment, it results only in observations which alter the state of the observer;
2. if the system uses contact sensors and manipulators (hands, arms, legs, body) this interaction almost always changes the environment, resulting again in observations that alter the state of the observer.

Typically, however the biological or man-made systems use all their sensory apparatus, contact and non-contact whenever possible.

This implies that manipulation and mobility are intimately tied to the act of perception. In fact we have shown [14] that this connection can be modeled formally by a nondeterministic finite state automaton.

### **3.1 Are Exploratory Procedures Selective?**

Exploratory procedures (EPs) are not selective in the same way that focus of attention is selective. According to Klatzky and Lederman [9, 10], they can be more or less optimal. The optimality is defined with respect to which hand movement best extracts the particular perceptual attribute. Also, EPs can be more or less general. This classification is with respect to the hand movement and the number of attributes it can deliver. The more attributes it extracts, the more general it is. A typical example of a most general EP is the enclosure EP. It results in attributes of hardness, general shape and size, and temperature. On the other hand the most specific and the most optimal EP for identifying the texture of a surface is rubbing. They also have found [8] that if vision is available, then vision is preferred for shape recognition while haptic EPs are preferred for material properties.

### **3.2 How Does Selectivity Fit Into All Of This?**

As mentioned previously, selectivity means choices. In the context of EPs, selectivity means choices from a menu of available EPs for a given task. Again from the work of Klatzky and Lederman (see previous references) we learn that people typically use the most general EP such as the enclosure grasp accompanied with lifting, and use the optimal EP only if the task requires it. This selection strategy seems to be supported by the argument of “economy” or “energy,” that is, we optimize the trade-off between the results required by the task and the energy spent on the task. Hence selectivity in the context of Exploratory Procedures is governed by the trade-off between the optimality and generality of the applied procedures for the given task.

### **3.3 Focus of Attention**

One can argue that Focus of Attention is very similarly a selection mechanism for economizing the sensory information gathering, processing and storing effort with respect to a given task.

Naturally, as mentioned earlier, one does not need this mechanism unless the system is mobile and can be active. The control of focus of attention is complex since there are several micro-behaviors, or micro-exploratory procedures that come into play depending on the task. Examples of such micro-EPs are:

- for vision:
  - control of the position and orientation of the head,
  - control of the focus, vergence/divergence,
  - control of the neck/body;
- for touch:
  - control of the position of the arm/body,
  - control of the position/orientation of the wrist,
  - control of the position/orientation of the hand fingers/palm;
- for audition:
  - control of the position/orientation of the head,
  - control of the position/orientation of the neck/body.

What are the rules that govern these micro-behaviors? There are two principles: focus on the expected stimulus (this is the case of tracking/following an object); and focus on the unexpected object/event. The first rule drives the system in normal conditions (task-driven), saying “watch what you are doing.” The second rule is invoked as a protection mechanism (exception-driven), saying “watch out what is happening.” If an unexpected event takes place, the system is interrupted and must decide whether to pay attention to the unexpected event or continue in the previous activity. This decision must be controlled by the cost function, which calculates and compares the risk of the unexpected event to the system with the risk of completion of the currently executed task or lack of it.

## 4 Perceptual Task Modeling

We suggested earlier that the task drives perception on many levels. Hence it should not be surprising that we consider task modeling from the machine perception point of view to be one of the most important research issues. However, task modeling is not independent of the context. Therefore context modeling is just as important as task modeling. Context is nothing more than generalized environment. We believe that the same principles must apply equally to modeling the task and to modeling the environment/context.

## 4.1 A Case for Purposive Perception

This is a good place to pause and ask [1]: “Is it fruitful to distinguish between purposive and non-purposive perception?” We argue that such a distinction is not meaningful, since all perception has purpose.

The purpose of perception is to deliver the necessary sensory information for the task at hand, which can be very concrete (e.g., to answer a specific question) or quite general (e.g., to survive). We argue that any system, biological or man-made, is limited by its finite energy resources, and therefore cannot afford to waste energy on non-purposive activity. When children explore the surrounding world or play, this is not non-purposive. Their purpose is to learn about their environment and their peers for future interaction. This is so even if they are not conscious of the purpose.

## 4.2 A General Consideration for Task Modeling

We follow here the framework introduced by Hager [7], who considers the perceptual task as a trade-off between the sensing and planning that results in optimal task performance. The task modeling structure consists of three components:

1. a transformation from geometric or physical parameter space into the decision space;
2. error sensitivity model describing the effect of decision errors;
3. and a model of information processing cost.

The decision space is that space of features/parameters that is necessary for making decisions with respect to the task. The important point about this particular decision space is that the attributes/features are represented by intervals rather than as points. The limits on intervals are given by the desired or necessary accuracy determined by the task. For example, in the task of grasping, the exact shape parameters are not needed, but the size is important for determining whether the object is grippable with a given hand.

An error sensitivity model describes the effect of a bad or wrong decision on system performance, and thereby introduces the means of deciding which parameter(s) or action is the “best” decision. We use the notion of utility function  $u(p, a)$  or loss function  $l(p, a)$ : if the (real) world is in state  $p$  and the action  $a$  is chosen, the gain to the system is  $u(p, a)$  and the loss is  $l(p, a)$ . The actual choice of  $u$  and  $p$  is very much task/situation dependent, and we are now beginning to experiment in concrete tasks, such as grasping and manipulation.

The cost of information is modeled with a function  $c(p, w)$  that is interpreted as computing the cost of carrying out the commands  $w$  and processing the resulting observations if the true state of nature is  $p$ . Again, we can measure the cost in terms of time/energy spent on a given task. We are using this framework for general tasks, including perceptual tasks.

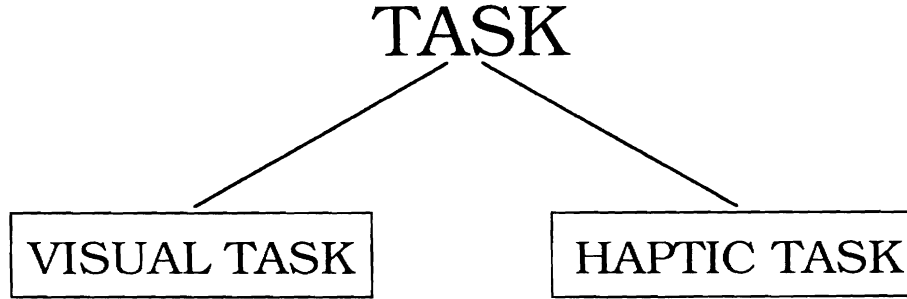


Figure 1: Functional Block Diagram

### 4.3 Modality-Specific Perceptual Tasks

Every perceptual task can be decomposed along the sensory modality axis: visual, auditory, haptic and others. Schematically, this decomposition for the two modalities of vision and haptic is shown in Figure 1.

The goal of each task is to deliver appropriate (with respect to sensory capabilities) perceptual properties. It is well known that vision is at its best in delivering two- or three-dimensional positional information and its spatial derivatives, such as shape, spatial relationships, patterns/texture and size. If time derivatives are being considered, then we obtain a new capability of motion and change detection via vision. However, the essential requirement for the visual sense to function is light and an illuminated surface.

On the other hand, an auditory sensor receives a sound waveform as a function of time. The perceptual properties from the sound are the direction of the source and the frequency spectrum of the source. The auditory and visual sensors are the primary non-contact sensors of biological systems. The haptic system is comprised of tactile, temperature, force and position sensors. The perceptual attributes this system delivers are material properties, such as hardness, thermoconductivity, surface roughness, elasticity vs plasticity, part mobility and weight. In the overall decomposition, the success of different perceptual tasks in a given environment will depend upon the choice of available modalities.

Visual tasks could be further classified into visual search (find, recognize, or identify a certain object) and visual tracking (visually follow a certain object). The typical task would be a combination of the two.

The haptic task is to extract geometric, material and kinematic properties of objects. It can be subdivided with respect to differences between the hand/object relationship and the properties it delivers. With respect to the hand/object relationship, the hand can be in contact with object but no enclosure is necessary; This is the case when the system is exploring only thermoconductivity, hardness and surface roughness. However if the task is to extract gross shape/size, part mobility, and weight, then the hand not only must enclose the object, i.e. grasp, but also must be able to manipulate it.

## 4.4 Exploratory vs. Verification Perceptual Tasks

There is another aspect of perceptual task classification. That is along the dimension of how much a priori knowledge is available. Consider these two extreme cases:

1. the system has no a priori knowledge about the environment/world, such as the case of a newborn baby.
2. the system knows everything about the environment, as well as about the object that it is supposed to find.

In the first case, we have a typical perceptual exploratory task. This will be described more concretely, in section 5.

In the second case, the perceptual task is only to verify the expected parameters about the environment and the object. Hence we shall call this a perceptual verification task. The only remaining question is how much of the verification procedure should be performed and how much the system can just infer from partial sensing. The amount of verification will be controlled by two criteria: the reliability and cost of the measurements, and the importance or accuracy of the success of the task.

## 5 A Robotic Perceptual Exploratory System

As mentioned before, exploration is essential to identification of unknown objects in an unstructured or partially defined environment. During this exploratory process, a robotic system needs to estimate, or recover fundamental object and environment attributes.

In this section we shall concentrate on the recovery of material and kinematic properties, while we assume that vision is available in the form of  $2\frac{1}{2}D$  range images. We shall use the global representation provided by superquadrics as defined in the work of Solina [13] and Gupta [6].

The importance of knowing the material composition of objects is fundamental to the issue of manipulation. How can a robot non-destructively grasp objects if it does not know how much force to safely apply? There is a quite large body of research on grasping and grasping stability issues, but in general it is assumed that the object being grasped is rigid, and usually made of a material that would withstand all grasping forces. However, in the case of exploration, this assumption cannot be made, since objects exist in all sorts of geometric shapes and made of a large set of materials. If a robotic system is ever going to succeed outdoors, it has to have the capability of first identifying some of the object attributes such as some of the object's material properties to only then be able to safely grasp it.

While the concept of Exploratory Procedures is quite general, in a robotic system they must be precisely defined in such away that they will equip the underlying robotic system with the capability of identification of material and kinematic properties of the unknown objects.

For the most part we assume that the objects are solid and manipulable, even though this is not a necessary requirement. Clearly, if the object is not manipulable, some attributes will not be able to be recovered, more specifically weight and part mobility. Even these two, in some sense, could be determined if the objects could be pushed and poked. Liquids, viscous mixtures, soils and biological materials are outside the scope of the present work.

## 5.1 The System Architecture

As we mentioned in section 4.3, the perceptual task is subdivided into *haptic* and *visual*. Similarly we differentiate between *manipulatory* and *haptic* tasks in that we consider manipulatory those actions which involve grasping an object and moving it about and/or around, whereas when the focus is on extracting data and other information we consider the action as being haptic.

Similarly the system architecture must mirror conceptual model for perceptual task. Hence we have integrated a robotic system to perform the *haptic task* and the *visual task*. In order to accomplish each of the above subtasks, we partition the exploration control into three sub-controllers: the vision control module, the haptic control module and the exploration control module. The exploration control module is responsible for starting the exploration, and to set the proper precedence between the other sub-controllers. The general architecture is shown in figure 2.

Klatzky and Lederman have identified five fundamental modules for the haptic task: *motoric*, *sensorial*, *property*, *exploratory procedures* and *object* modules. Our robotic architecture, however, requires a somewhat different partitioning. The object Module in their work corresponds to the *haptic task description* in ours. Our *haptic properties and EPs* are very similar to theirs. Their motoric module can be mapped into two parts of our system: one is the *robot arm controller* and the other is the *end-effector controller*. Their sensor module is, in comprised in our case, of the *force/tactile* and *position sensor* modules. The organization of these modules are shown in Figure 3 b.

As the *haptic task*, the *visual task* is also composed of *visual properties*, which are extracted by *visual exploratory procedures*. These visual EPs control the position of the head, neck, the focus and vergence of the eyes, opening and closing of aperture (iris), similar to the implementation of Krotkov [11]. The visual task also determines what resolution/detail, as well how many views and how much data should be acquired and what features need to be extracted. Interaction between the visual task and the haptic task in this implementation at the physical level is via ethernet.

The visual task, similarly to the haptic task, is also subdivided into modules, which is shown in Figure 3 a.

## 5.2 Object Attributes

As well established in perception, objects are described by attributes. Our classification differs slightly from the one proposed by psychologists. While they classify object attributes

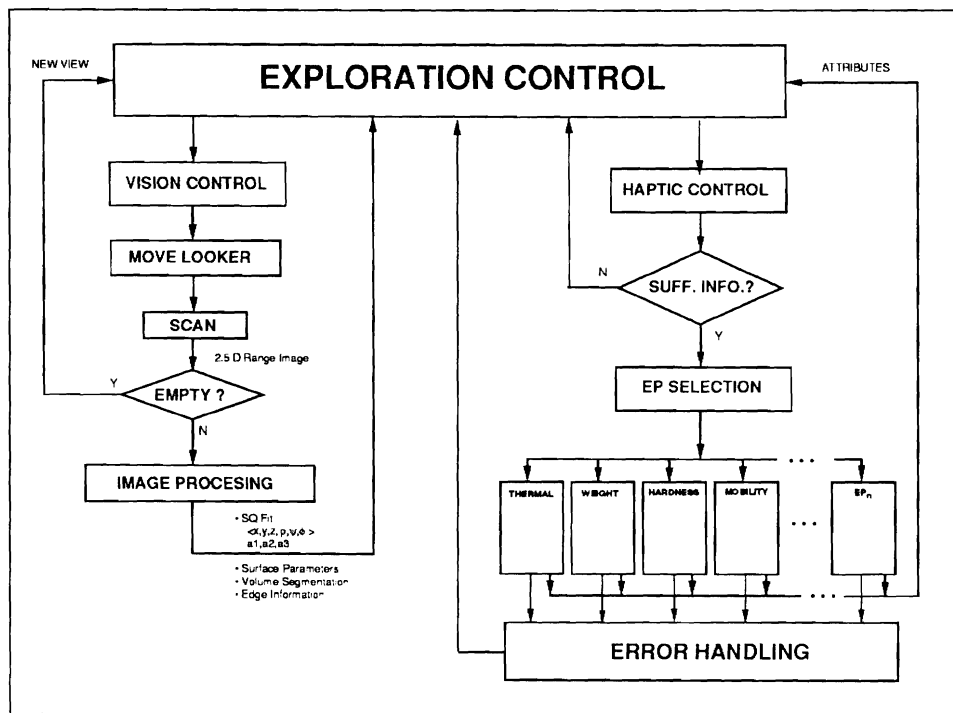


Figure 2: The Architecture

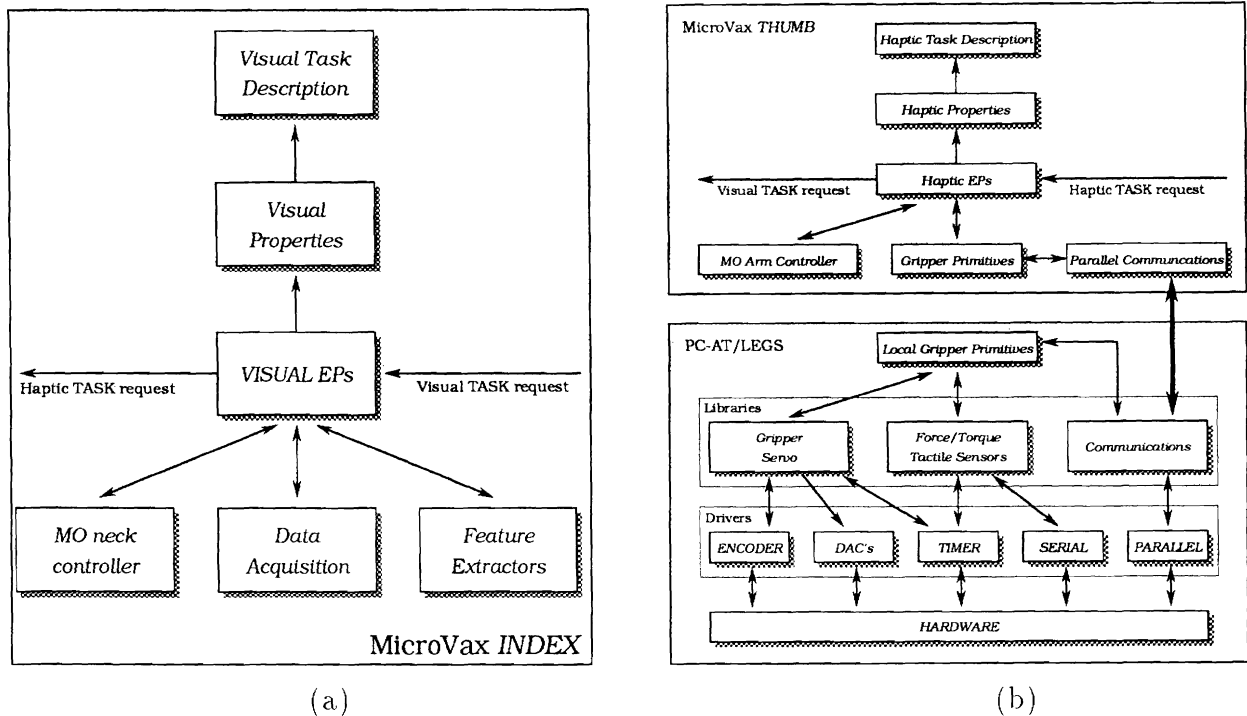


Figure 3: The Visual and the Haptic Tasks

into three categories, namely, *substance properties*, *structural properties* and *functional properties*, we chose to classify object attributes into the following categories:

1. Material Properties
2. Geometric Properties
3. Kinematic properties

Objects can be composed of a diversity of materials. They can be homogeneous or heterogeneous. Depending on the material an object is made of, that material will impart to the object very specific characteristics. An object also has an inherent structure, that we call here its geometric attribute. Together with the material properties, the geometry defines a *structural* quality to an object. For example, take a sheet of cardboard. It has its own characteristic material composition. When "shaped" into different objects, however, its structural characteristics are going to vary accordingly. A sheet is pliable, foldable, not very hard. But when formed into a box, the new object becomes rigid, firm, and in a subjective sense, hard. The material of the box and of the sheet is the same, therefore so are the material properties. But by changing the geometric configuration, a new object arises, with the same material properties of a cardboard sheet, but with different structural properties. Clearly, these structural properties are related to the moments of inertia of the object.

### 5.2.1 Material Properties

Various materials respond differently to the same stimuli and constraints. The quality of the response to imposed stimuli and constraints is determined by the nature of the material. This quality is defined as the *material property*. In most cases it is not enough to *how or qualitative* a material responds to a given stimulus, but also *how much* it responds to that stimulus. Most of the time in our work we will be seeking to obtain answers to both how and how much of a given response. Material properties can be further classified into two main groups: *Mechanical* and *Physical properties*.

1. **Mechanical Properties** Mechanical properties are related on how a material responds to an applied force or stress (stress is defined as the applied force divided by the cross sectional area on which the force acts). Among the mechanical properties we have:
  - Hardness
  - Brittleness
  - Compliance
  - Elasticity
  - Plasticity
  - Viscosity and Creep
  - Ductility
  - Impact
2. **Physical Properties** Physical properties depend on both the structure and processing of the material:
  - Density (mass)
  - Chemical
  - Electrical
  - Magnetic
  - Thermal
  - Optical

### 5.2.2 Kinematic Properties

Kinematic is the branch of Dynamics which deals with motion of physical bodies isolated from the forces associated with a given motion. It is concerned with relative displacements of rigid bodies. We will be looking for mobility within a rigid body, and as such we classify mobility under kinematic properties.

**Mobility** – In terms of object attributes, we will be dealing with part motion, or intrinsic mobility. More specifically, we will be analyzing the presence of the degrees of freedom, and possibly determining the lower pairs associated with those degrees of freedom.

Since mobility is essentially a task for two hands (obviously with dexterity one would be able apply forces using different fingers). Therefore, in our set up, since we do not have dual arm manipulation, we use a stationary vise with which we will hold one of the object’s extremity. The other extremity will be free so that it maybe grasped by the end effector, in our case the gripper.

### 5.2.3 Weight Properties

Weight is a fundamental property of a given object. It is almost natural for anyone to heft an object in order to have a feeling of its “solidity”. The mass as well as its distribution within an object help us identify more accurately the material its made of.

There are some visual cues that may lead us to hypothesize on the weight of an object. Its geometric and surface texture characteristics may indicate something about its weight. To our visual system a solid block of a given material and another hollow block of the same material would be considered as being the same. By picking up the block one can easily and quickly detect the hollow from the solid one. Also. it would be impossible to our haptic system to determine that with just a simple static contact with the object. By moving the objects, however, we are able to compare the work performed and from that discern between the two solids. By unsupported holding we feel directly the force created by the acceleration of gravity on the mass we are lifting. By pushing it perpendicularly to the gravitational field we are assessing, indirectly, the effect of gravity on that object. since the pushing force is proportional to the normal force multiplied by the value of the coulomb friction.

Weight is a manifestation of the mass density physical property of the material a body is composed of *and* of the gravitational field it is submitted to. In outer-space, objects are *weightless* since there is no significant gravitational field to creat a force on the object. For this reason, in the absence of gravity, unsupported holding would be of no use in determining an objects mass characteristics.

A similar situation happens under water. In this case. however, gravitational fields present, but now the weight perceived is modified because of the *buoyant force* acting on a the body. This counter-acting force is related to the material’s mass density by the well known Archimedes principle.

## 5.3 One Concrete Example – The Weight EP

In this subsection we demonstrate our methodology for design of of Exploratory procedures and their application.

¿From vision (complies with our earlier assumptions) we were able to obtain the centroid, which is the geometric center of the body. However, the only instance that the centroid and the center mass will coincide is when the object is perfectly homogeneous. By the utilization of an exploration routine that will grasp the object “about” the centroid, we are

able to determine the mass. If the grasping occurred about the centroid, and the object is homogeneous, then the moments about a point at the center of the gripper fingers will be zero. This implies that the object is either homogeneous or the distribution of mass within it is uniform. If however that does not happen, we can further define the center of gravity by holding the object at one of its extremities. The moments at the gripper fingers should be non zero. By noting the moments and measuring the weight, the center of gravity can be easily computed.

The importance of knowing the center of gravity stems from the fact that in grasping an object, moments about the gripper fingers should be minimized, in order to avoid manipulatory instabilities.

Following we list an algorithm of the exploratory procedure to determine the mass.

```
double weight_ep()
{
    rs_homog
transform;
    sq_struct
sq;
    double
f,
t,
CG,
weight,
dz;

    sq = scan_object(transform);
    pick_up_object(transform,sq);
    lift_object(dz);
    get_forces_and_torques(f,t);
    weight = trans_ft(f,t);
    CG1 = estimate_cg(sq);
    lower_the_object(-dz);
    release_object();
    sq = scan_object(transform);
    move_to_obj_point(sq);
    pick_up_object(transform,sq);
    lift_object(dz);
    get_forces_and_torques(f,t);
    CG2 = estimate_cg(sq);
    lower_the_object(-dz);
    release_object();
    if (CG == sq->centroid)
HOMOGENEOUS;
```

```

    else
HETEROGENEOUS;
    return(weight);
}

```

This algorithm assumes that the object can actually be lifted, which in some cases it may not be true. Other error conditions, such as this, would be checked in the actual implementation.

### 5.3.1 Preliminary Results

The following results were obtained by an implementation of the algorithm mentioned in the previous section. The objects utilized had the exact same dimensions of  $78mm \times 39mm \times 26mm$ . For each object we realized 100 measurements. The results of these measurements can be seen in Figure 4. The standard deviation on the measurements were in the order of 9 grams.

In conclusion, we have shown a design of exploratory procedure of weight. This EP subsumes ability to grasp the object. While for computing the weight it is not essential whether one grasps the object at the mass center or not, the grasp at the mass center is the most stable one. Hence it is desirable to grasp at the mass center whenever it is physically possible. In view of this effort with the information about centroid obtained from vision as a side product one obtains the mass distribution of the object as described above.

## 6 Learning As A Natural Consequence of Exploration

If we accept that there is a whole scale between the perceptual exploratory task and the perceptual verification task with respect to the amount of a priori knowledge available to the system, then the system must have capabilities to collect and organize the knowledge about the environment and the objects. In other words, one can state the arguments, why a system—either artificial or biological—must have learning capabilities:

1. bounded amount of memory;
2. bounded time/energy of access to/from memory.

The fundamental issue is what is innate and what is learned. We hypothesize that exploration is a necessary prerequisite to perceptual development and learning. Hence the question is transformed into what is innate and what is learned for exploration.

### 6.1 Innate Assumptions

We begin with some given hardware/anatomic configuration. We postulate that the sensors, joints, muscles, actuators (input/output devices), memory and some basic processing capabilities (an instruction set) are built-in and innate. Naturally, we are fully aware that the

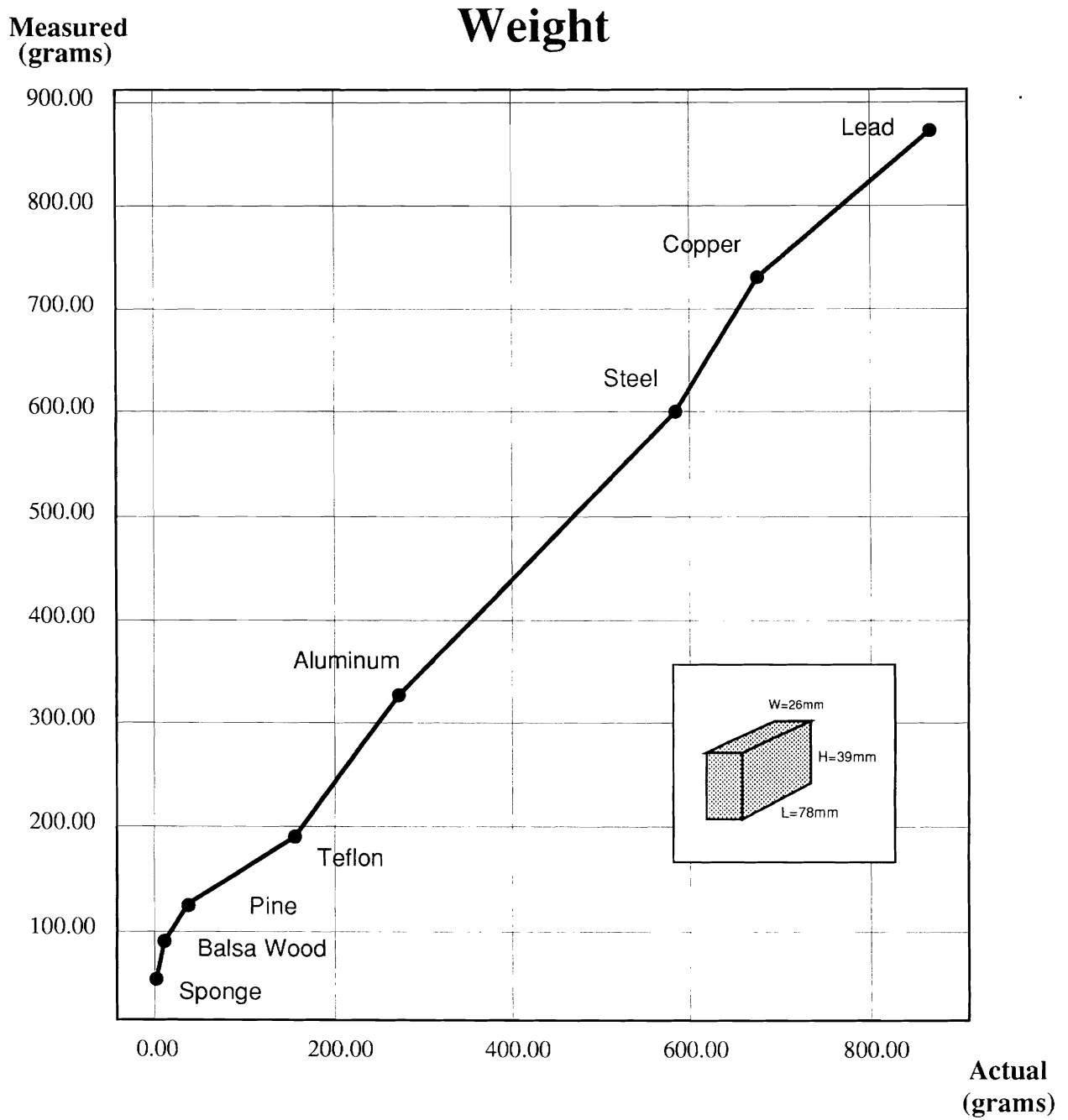


Figure 4: Weight EP - Several objects with same dimensions but of different materials

innate local processing (specialized processing) and the interconnections among the above modules are quite complex and not yet fully understood. However for our purposes—to outline and learn about a skeletal architectural system that would have exploratory capabilities in a simple environment—what is known will suffice.

From psychological and physiological studies we know that primates and humans are born with motor reflexes, such as reaching, flexing their hands, and moving their heads and eyes. These are the basic capabilities that seem to be hard-wired into the system. In humans, body movement and walking develop with maturity, but many animals are born with these capabilities. From the machine perception point of view, we take it as given that all the motor reflexes are innate, as well as the connections between sensors and their respective motor controller. An example is the feedback system between the visual sensor and the head/eye controller.

Furthermore, some data reduction mechanisms must be initially available. Examples of those include an edge detector, a spatial homogeneity detector (a blob detector), and a pattern detector both in space and time. The system must have computational devices such as a thresholding device that can serve as a quantizer, a differentiator, and a comparator.

There must be a capability of an energy/time measure and some prioritization or optimization scheme with respect to the energy required to make decisions, selections, and choices. This decision mechanism controls the organization of different modalities, how to apply them to a given situation, within each modality, how to process/reduce the data, how to combine the information from different sensors and how to use the memory.

We assume that each sensory modality has allocated a certain portion of memory (analogous to the visual, auditory, and somato-sensory cortex), as well as a place where intermodal information is stored. However, there must be a separate innate place-holder in the memory for the task representation. As mentioned before, the initial task is to eat and not be eaten, which translates to exploring the environment in order to identify what can be eaten and what might eat you.

Where are facts about the physics of the world encoded? We postulate that this information is innate and is distributed as follows.

1. There are built-in gravity sensors.
2. The visual and auditory sensors are positioned on the organism to assure well-defined orientation and position of the objects in the outer world and the organism.
3. The contact sensors reinforce the position and orientation information from vision and audition, and also provide force and thermal information.
4. As mentioned earlier, there is evidence of innate computational mechanisms computing derivatives of position information both in space (for surface normals and curvature) and in time (velocity, acceleration) as well as derivative of force or energy in time.

## 6.2 Learning Process

Given the above innate apparatus, how can the organism explore? The task controller begins the execution of the perceptual exploratory task. This in turn is distributed into visual, haptic and other perceptual tasks. The vision modality delivers shape and size information; the haptic modality delivers material properties and supplies grasping reflexes to enclose objects. All this information is stored for the first time in respective memory slots, but is also recorded in the intermodal region if the sensory information comes from the same object. This is the initial phase in the perceptual development of an organism. In the paradigm of active perception in an exploratory mode, the system will continue in its activity, i.e., exploration will continue until the agent knows everything or its energy is exhausted. In the second phase of its activity, there are two possibilities:

1. the experience does not differ from the previous instance (this is case of reinforced learning); or
2. there is a new experience, which the system has to record and register with previous knowledge.

This brings us to the issue of learning and reintroduces the question of what mechanisms must be innate so that the system can begin to learn. We are only concerned here with perceptual learning and perceptual development. In this context, we postulate that this type of learning is primarily inductive as opposed to deductive. As noted by the artificial intelligence community [12], the innate capabilities must include reinforcement and inhibition, search (which includes comparison and recognition of sameness and difference), plus differentiation and association.

With the above tools the system can begin to learn categories of objects. first in a coarse categorization. and later in a refined one. The important point at this stage is the realization that all sensory information is processed and transformed into some parametric representation. This parametric representation is the “alphabet”—or the quantized version of the measured signal—that encodes the percepts. Some examples of this alphabet include size, orientation, shape parameters, hardness, and surface roughness. Initially, the system may have only two values for each parameter; later, the scale is refined. Nevertheless, each parameter is limited by an interval that is determined by the physical capabilities of the sensor/actuator. The size of the interval is learned through experience. In machine perception it can be derived partially from the knowledge about the particular sensor, and partially by calibration, which is the same as experience.

If we accept that the parametric representation is the basic entry into the memory, then the problems of finding a certain object in the scene (a classical top-down problem), and interpreting the object found in a scene (a classical bottom-up problem) are equivalent in the sense that they amount to a search problem. In the first case, we take the template of the object from the memory and look for the appropriate match in the perceptual data. In the second case we take the perceptual data about the object and look for the proper match in the memory in order to make the interpretation. This implies that all the memory can

be modality specific, similar to the visual, auditory, somato-sensory cortex with connections to the intermodal region. Of course, we are not addressing the issues of generalization, abstraction and reasoning.

Let us consider a concrete example. How would Learning via Exploration, using the EP of weight work? Initially the system is presented or randomly exposed to a series of object made to a same shape/size but from different materials. As a result of the EP of weight, the system will store a table of different materials associated with different weight, as shown in Figure 8. The association of the particular names is unessential but the recognition of different categories is the crucial point. When the system is again exposed later on to a similar shape/size object based on its measured forces in its wrist can by a simple lookup table identify the different materials. This is the bottom-up instance of search in the memory. The top-down analogy is when the system is asked to find/identify from a given set of similar shape/size objects that object which is made from a given material. While learning via exploration is very basic to learn about the environment, it does not include learning about behaviors. This is in some sense a higher order learning activity, much more complex called learning by mimicking. This entails : observe the pattern of behavior, store it as percepts. Generate the copy of this behavior. Observe yourself, compare/match with the stored pattern. If agreement, it is success and calls for reinforcement. If failure, then correct and repeat. Clearly this is the next research agenda.

## 7 Conclusions

The main goal of this paper has been to show that there is a natural flow from active perception through exploration to perceptual learning. We have tried to conceptualize the perceptual process of an organism that has the top-level task of surviving in an unknown environment. During this conceptualization process, four necessary ingredients have emerged for either artificial or biological organisms. First, the sensory apparatus and processing of the organism must be active and flexible. Second, the organism must have exploratory capabilities to deliver information about new and unknown environments. Third, the organism must be selective in its data acquisition process, guided by the task. Fourth, the organism must be able to learn, i.e., organize the perceived information to reduce the data so that it can be stored and retrieved in a finite time and with finite memory. We also have shown that if an organism has the above capabilities then it can perform more specific tasks, such as finding and manipulating arbitrary objects.

In the section on learning, we have clearly delineated the difference between what must be innate and what must be learned. This delineation allowed us to identify the indexing scheme, which is based on geometric and physical parameters. This indexing scheme transforms the traditional bottom-up and top-down processes to the *same* search problem.

As mentioned, this paper is primarily conceptual. However, it has some experimental results to document the concreteness of this proposed theory. The system's architecture shows the design that follows from the perceptual task decomposition. While we are not discussing in any detail the visual module, it should be clear that vision is an equal player

in our system as is the haptic module. The details about the visual module is the subject of another paper [6]. In this paper we are attempting to project an integrated view of perception, exploration, and ultimately, perceptual learning. The haptic module alone is much larger than we have discussed in this paper. Here by showing the weight EP, which is of average complexity, we have tried to outline a general methodology in concrete terms for the EPs. The whole haptic system is described in [3]. The predictions that this theory makes are that an artificial system can explore and learn about its environment modulo its sensors, manipulators, end effectors and exploratory procedures/attribute extractors. It can describe its world with respect to the built in alphabet, that is, the set of perceptual primitives.

## 8 Acknowledgements

Acknowledgements: Navy Grant N0014-88-K-0630, AFOSR Grants 88-0244, AFOSR 88-0296; Army/DAAL 03-89-C-0031PRI; NSF Grants CISE/CDA 88-22719, IRI 89-06770; and Du Pont Corporation

## References

- [1] Y. Aloimonos. Purposive and qualitative active vision. In *Proc. DARPA Image Understanding Workshop*, pages 816-828, 1990.
- [2] R. Bajcsy. Active perception vs passive perception. In *Third Workshop on Computer Vision: Representation and Control*, Computer Society Press, Bellaire, MI, October 1985.
- [3] Mario Campos and Ruzena Bajcsy. A robotic haptic system architecture. In *ICRA*, pages 338-343. Sacramento, CA, April 1991.
- [4] E. J. Gibson. Exploratory behaviour in the development of perceiving, acting, and the acquiring of knowledge. *Ann. Rev. Psychology*, 39:1-41, 1988.
- [5] J.J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979. Reprinted 1986, Erlbaum.
- [6] Alok Gupta. *Surface and Volumetric Segmentation of Complex 3-D Objects Using Parametric Shape Models*. Technical Report MS-CIS-91-45. Department of Computer and Information Science. University of Pennsylvania. 1991. Ph.D. Dissertation.
- [7] D. Hager. *Task-Directed Sensor Fusion and Planning: A Computational Approach*. *International Series in Engineering and Computer Science*. Kluwer Academic Publishers, 1990.
- [8] R. Klatzky and S. Lederman. Hand Movements: A Window into Haptic Object Recognition. *Cognitive Psychology*, 19:342-368, 1987.

- [9] R. L. Klatzky and S. J. Lederman. Haptic Identification of Objects as a Constraint Satisfaction System. 1991. Unpublished manuscript to be submitted.
- [10] R.L. Klatzky, S.J. Lederman, and C. Reed. Haptic Integration of Object Properties: Texture, Hardness, and Planar Contour. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1):45–47, 1989.
- [11] E. P. Krotkov. *Active Computer Vision by Cooperative Focus and Stereo*. Springer-Verlag, New York, 1989.
- [12] J.R. Quinlan. Determinate literals in inductive logic programming. In *12th International Conference in Artificial Intelligence*, pages 746–750, Aug. 1991.
- [13] F. Solina. *Shape Recovery and Segmentation with Deformable Part Models*. PhD thesis, University of Pennsylvania. Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, Dec 1987. Technical Report MS-CIS-87-111/GRASP LAB 128.
- [14] Constantine J. Tsikos and Ruzena Bajcsy. Segmentation via manipulation. *IEEE Transactions On Robotics And Automation*. 7(3). June 1991.