MACHINE VISION AND LEARNING: RESEARCH ISSUES AND DIRECTIONS

by

R. S. Michalski
A. Rosenfeld
Y. Aloimonos

Reports of the Machine Learning and Inference Laboratory, Reports of the Center for Automation Research CAR-TR-739, MLI 94-6, CS-TR-3358, George Mason University, Fairfax, VA, University of Maryland, College Park, MD, October 1994.

MACHINE VISION AND LEARNING:
Research Issues and Directions

A report on the NSF/ARPA Workshop on
Machine Learning and Vision
Harpers Ferry, WV, October 15-17, 1992

MLI 94-6
CAR-TR-739
CS-TR-3358

Prepared by

R.S. Michalski, George Mason University
A. Rosenfeld and Y. Aloimonos, University of Maryland

in collaboration with NSF/ARPA Workshop participants

October 1994

The Workshop was sponsored by the National Science Foundation and the Advanced Research Projects Agency under Grant IRI-92-08974 to George Mason University. The continued support of ARPA and the Air Force Office of Scientific Research under Contract F49620-92-J-0549 to George Mason University is gratefully acknowledged, as is the help of Susan Palmer and Sandy German in preparing this paper. This report is being published by both institutions simultaneously.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

One of the central limitations of current machine vision systems is that they have no or very limited learning capabilities. In recent years, there has been increasing realization among researchers that the incorporation of such capabilities in vision systems is highly desirable, and constitutes an exciting new challenge for interdisciplinary research on vision through learning. Vision systems capable of learning would apply to a wider range of practical problems, would be easier to adapt to new problems, would be more flexible and more robust. Due to the rapid progress in the field of machine learning, implementing such capabilities in vision systems is becoming increasingly feasible.

In reflection of these developments, there has been growing interaction and collaboration between the machine learning and vision communities. By 1992, these research interests and interactions had reached a point at which current research activities and future perspectives needed to be discussed and analyzed. To this end, the NSF/ARPA Workshop on Machine Learning and Vision was organized by George Mason University in collaboration with the University of Maryland, and was held on October 15–17, 1992 in Harper's Ferry, WV. The workshop brought together researchers in vision and learning to discuss the possibilities of cross-fertilizing the two fields and implementing learning capabilities in vision systems. It was attended by 45 participants representing universities (21), industrial and governmental laboratories (16), and sponsoring agencies (8).

This report is based on the discussions at this Workshop and on a preliminary report written by the participants (Michalski et al., 1993). Part 1 of the report provides research overviews—one on machine vision (by A. Rosenfeld) and one on machine learning (by R. S. Michalski). Part 2 discusses potential roles of learning in task-independent and task-oriented (purposive) vision systems; the framework of this discussion was developed by Y. Aloimonos. Much of the material in Part 2 is based on group discussions at the workshop, which were coordinated and initially summarized by J. Shavlik and T. Poggio (object recognition), T. Dean and T. Kanade (navigation), and R. Bajcsy and T. Mitchell (sensory-motor control). P. Pachowicz made substantial contributions to the editing of the preliminary report and the preparation of the supplemental bibliography for this report.

Overviews

The goal of machine vision is to derive descriptions of a scene, given one or more images of that scene. Traditionally, theoretical work on machine vision has pursued a *task-independent* approach which attempts to derive "complete", "general purpose" descriptions using methods that are applicable to broad classes of scenes. An important conceptual advance during the past few years is the development of the *purposive* approach, which seeks partial descriptions that are meaningful in connection with specific tasks being performed by a vision-based system operating in a specific domain.

Machine learning is the branch of artificial intelligence devoted to the study of computational processes by which a system can construct representations of knowledge or skills, using information provided by the external world and its own previous knowledge. The types of representations that need to be constructed and the methods for constructing them can vary greatly, and this leads to a tremendous diversity of possible learning approaches and techniques.

For example, learning may involve building general object descriptions from specific observations, acquiring problem solving methods on the basis of examples of solutions, improving algorithms through practice or experimentation, constructing control heuristics from experience, creating solutions to new problems by analogy with solutions to similar problems, discovering statistical or logical regularities or relationships among entities, and so on.

Candidate descriptions or algorithms must satisfy certain constraints that can be usually expressed in a functional form. Therefore, many learning problems can be characterized as determining a complete description of a function on the basis of examples of input-output pairs that satisfy the function. Learning problems can then be classified according to the nature of the range and domain of this function. In the purposive approach, the search for the desired description can be restricted to the relevant subsets of the range and domain, or the function can be decomposed into simpler functions.

Learning and vision

Children (and animals) "learn" vision—i.e., acquire skills that make use of visual input—merely by being immersed in their environment; they can learn from unordered visual experiences and without a designated teacher. A long-term challenge to researchers in machine learning and machine vision is to eventually give vision systems this ability.

Vision systems can make use of learning in many different ways. For example, learning can be employed in developing or improving various components of "general-purpose" (task-independent) vision systems, for building general descriptions of visual objects from object examples, or for discovering constraints on a class of scenes from observations and prior knowledge. It also provides a general approach to designing task-oriented ("purposive") vision systems, using task-independent learning methods. In particular, it can be used for controlling the dynamical processes used by vision-based autonomous systems ("agents").

In task-independent approaches to learning in vision, many interesting theoretical questions arise—for example, how to determine relevant features, how to choose or search for the most appropriate representation spaces, what languages to use for building object descriptions in those spaces, what are the complexities of different learning tasks, how to learn models for shapes, textures and motions, and how to synergistically integrate these models.

The purposive approach to learning in vision can be well illustrated by the many roles of learning in vision-based navigation. Any active agent requires an ability to navigate, or more generally, to coordinate vision and action, in order to operate effectively in its environment. Navigation problems arise at many levels, from local (steering, tracking, obstacle avoidance) to global (route planning). The problems vary in difficulty depending on how strongly constrained the agent's actions are, how unstructured and dynamic the environment is, and how sophisticated the agent's observations are. Navigation tasks can be classified in accordance with how much inference they require, how much time is available, how much supplemental (non-sensory) knowledge is available, and how complex the required behavior is.

Current research on learning vision systems has been exploring a spectrum of approaches and problems. A large portion of this research has dealt with neural net applications, for example, to road navigation, object detection in various types of images (visible, range, sonar, radar, etc.), or learning control functions. Advantages of these methods include their generality and their ability to learn continuous transformations. Disadvantages include the difficulty of incorporating prior knowledge (especially relational knowledge), the difficulty of learning complex structural knowledge, slow learning rates, and lack of comprehensibility of the learned knowledge. While symbolic learning methods suffer much less from the latter problems, they have been primarily oriented toward learning discrete concepts and transformations and applied mostly in areas other than computer vision. In computer vision, they may be particularly useful for feature extraction, new feature generation, learning visual surface descriptions (e.g., textures), learning complex shape descriptions, acquisition of structural or relational models of objects, construction and updating of model bases, "conceptual" scene segmentation, and others. Applications of symbolic approaches to vision problems remain an insufficiently explored but potentially very fruitful domain of future research. Because of this, and because of the existence of extensive literature on the applications of neural net and closely related approaches to vision, this report gives special emphasis to symbolic learning methods.

Although many examples of the application of learning methods to vision-related tasks have been reported in the literature, there is a lack of principles for deciding which approaches and methods to use, and for guiding the design of learning vision systems to perform specific tasks in specific situations. Such principles are essential for developing systems that are well-suited for given tasks, can deal with a wide range of vision problems, and can work in environments whose characteristics vary significantly and change over time.

Recommendations

- Machine learning offers significant opportunities to extend current vision techniques, particularly in the area of task-oriented vision. Selected projects, oriented toward important potential applications, should be established with short-term and long-term goals.
- Promising research opportunities for the use of learning approaches in vision systems include:
  - Automated selection or determination of sensors, features, algorithms and modules that are most appropriate to given classes of scenes and given tasks; identification of "key" characteristics of given object classes and automated synthesis of new features.
  - Acquisition of general object descriptions from object samples; specific problems include learning 2D or 3D shape models, learning surface descriptions from samples, and decomposing objects or scenes into semantically meaningful components ("conceptual segmentation").
  - Coordination of the vision modules that perform common or similar visual computations.
  - Control of the image acquisition process in active vision systems, in order to acquire the data most useful for the given task.
  - Identification of contexts and tracking of changes in the environment, so that the system can adapt appropriately to variations in the class of scenes.
  - Task decomposition and planning—reducing vision tasks to combinations of simpler subtasks, so that sequences of subgoals can be selected that result in a desired goal.
- Collaboration between the machine learning and vision communities, involving extended visits, shared testbeds, benchmarks, and competitions, should be encouraged.

1. RESEARCH OVERVIEWS

1.1. Perspectives on Machine Vision

1.1.1. Introduction

The goal of machine vision is to derive descriptive information about a scene by an analysis of images of the scene (Figure 1); obviously, the nature of the information to be derived depends on the goals of the system that needs the information. Vision algorithms can serve as computational models for biological visual processes, and they also have many practical uses; but this overview treats machine vision as a subject in its own right. Vision problems are often ill-defined, ill-posed, or computationally intractable; nevertheless, successes have been achieved in many specific areas. It is suggested that by considering the vision system of an organism or robot ("agent") as an embodiment of the agent's relationship to its environment, useful solutions to many vision problems can be obtained.

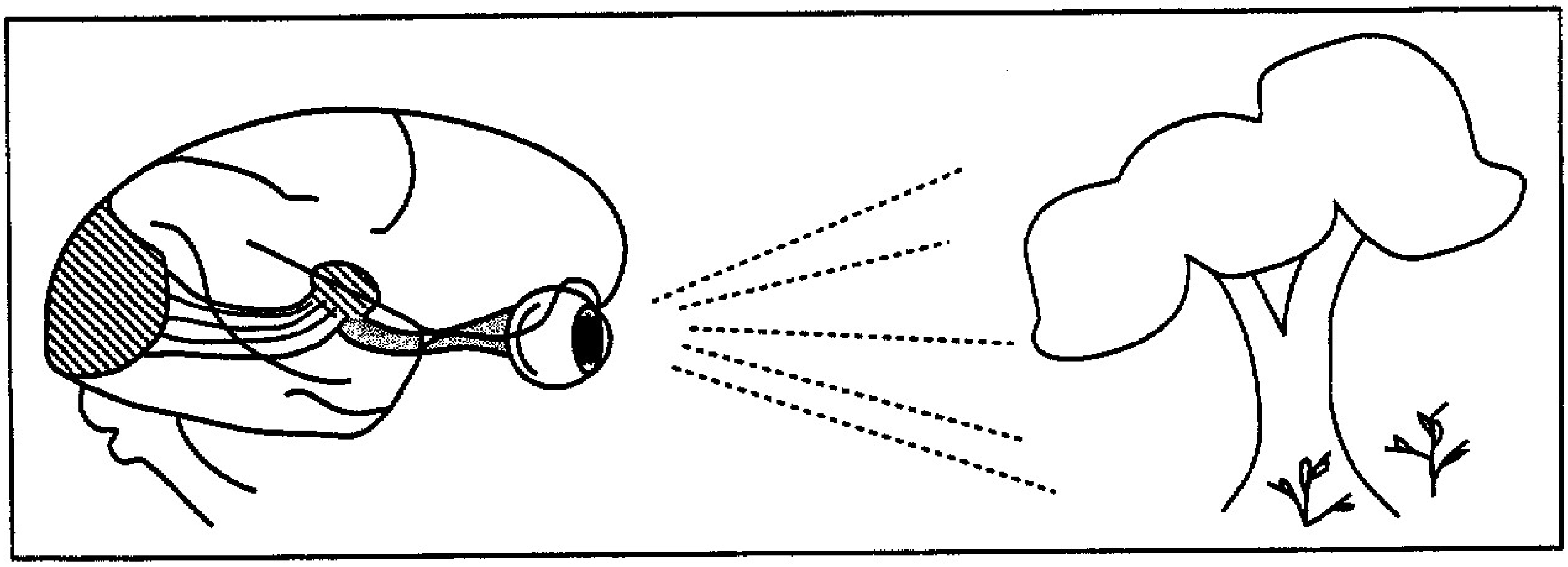

Figure 1: Neither objects nor properties of objects (such as shape, movement, color, etc.) exist in our brains as such. When we see, computations are performed inside our heads that make us understand the visible world (objects and their properties). It is difficult to explain how this is done; many theories of visual perception have been formulated. This figure shows visual areas in the brain. The eye forms an optical image of the scene on the retina, an array of photoreceptors which discretely sample the image. Other layers of cells associated with the retina perform various types of local computations on the sampled image. The processed image is transmitted in parallel along the optic nerve, through the lateral geniculate body, to the striate visual cortex, where additional types of local processing are performed; these processes are sensitive to the presence of local features such as spots, bars, and edges in the image. The outputs of these processes are transmitted to other areas of the cortex, particularly to the posterior parietal cortex and the inferior temporal cortex, where global properties of the image appear to be analyzed.

The most common class of images used in machine vision systems are optical images whose brightness at a point is determined by the amount of light received by the sensor (e.g., a TV camera) from a given direction. Images can also be formed using other types of radiation (e.g., infrared or ultrasound), or they can measure the distance from the sensor to the nearest surface in the scene in a given direction (e.g., radar or range sensing); but in what follows we will usually assume an ordinary optical image.

An image is input to a digital computer by sampling its brightness at a regularly spaced grid of points, resulting in a digital image array. The elements of the array are called pixels (short for "picture elements"), and their values are called gray levels. Given one or more digital images obtained from a scene, a machine vision system attempts to (partially) describe the scene as consisting of surfaces or objects; this class of tasks will be discussed further in Section 1.1.2.

Animals and humans have impressive abilities to successfully interact with their environments—navigate over and around surfaces, recognize objects, etc.—using vision. This performance constitutes a challenge to machine vision; at the same time, it serves as an existence proof that the goals of machine vision are attainable. Conversely, the algorithms used by machine vision systems to derive information about a scene from images can be regarded as possible computational models for the processes employed by biological visual systems.

Vision techniques for analyzing images have many practical uses. Areas of application include document processing (e.g., character recognition), industrial inspection, medical image analysis, remote sensing, target recognition, and robot guidance. There have been successful applications in all of these areas, but many tasks are beyond current capabilities (e.g., reading unconstrained handwriting). These potential applications provide major incentives for continued research in vision. However, successful performance of specific tasks on the basis of image data is not the primary goal of machine vision; such performance is often possible even without obtaining a correct description of the scene.

Viewed as a subject in its own right, the goal of machine vision is to derive descriptions of a scene, given one or more images of that scene. Vision can thus be regarded as the inverse of computer graphics, in which the goal is to generate (realistic) images of a scene, given a description of the scene. The vision goal is more difficult, since it involves the solution of inverse problems that are highly under-constrained ("ill-posed"). A more serious difficulty is that the problems may not even be well-defined, because many classes of real-world scenes are not mathematically definable. Finally, even well-posed, well-defined vision problems may be computationally intractable. These sources of difficulty will be discussed in Section 1.1.3.

It is difficult to formulate a satisfactory theory of visual perception; such a theory would have to take into account the environment, the visual stimuli, the sensory receptors, the brain, and the effectors (Figure 2). This may be the reason for the large variety of theories of visual perception that have been formulated over the past few hundred years. At the same time, vision is studied by researchers in many different disciplines (mathematics, engineering, psychology, anatomy, etc.) Researchers in different areas ask different questions about vision (Figure 3). Some ask theoretical questions (what could be: What kind of descriptive information can be derived about a scene from an analysis of its images?); this question is the subject of research in physics, mathematics, computer science, etc. Some ask empirical questions (what is: How are the visual systems of existing organisms designed?); this question is studied in zoology, psychology, neuroanatomy, etc. Finally, some ask purposive or normative questions (what should be: How should the visual system of an organism or robot be designed so that it can best perform a set of tasks?); these questions are of interest to engineering disciplines. The three types of questions are related but do not necessarily have the same answers; what exists could be suboptimal, and only a small subset of what is theoretically possible could be of practical relevance to autonomous "seeing" systems.

In this overview we restrict our attention to the theoretical and purposive questions. In the theoretical approach we consider vision in isolation and we study what could be possible; we regard vision as a process that generates general-purpose descriptions of scenes (Figure 4). These descriptions are then given as inputs to cognitive processes, such as reasoning, planning, etc., appropriately modified as necessary to suit the needs of these processes. In the purposive approach, on the other hand, we consider vision as a part of a larger system and we study what should be; we regard vision as a process that generates partial descriptions of a scene that are purposive, i.e. that are meaningful in conjunction with a given task or action (Figure 5). Thus this approach views vision as creating an interface between the world and the cognitive processes (reasoning, planning, etc.) for the purpose of taking an action. An action can be practical (like a motor command), theoretical (like reaching a decision or building a specific representation), or aesthetic.

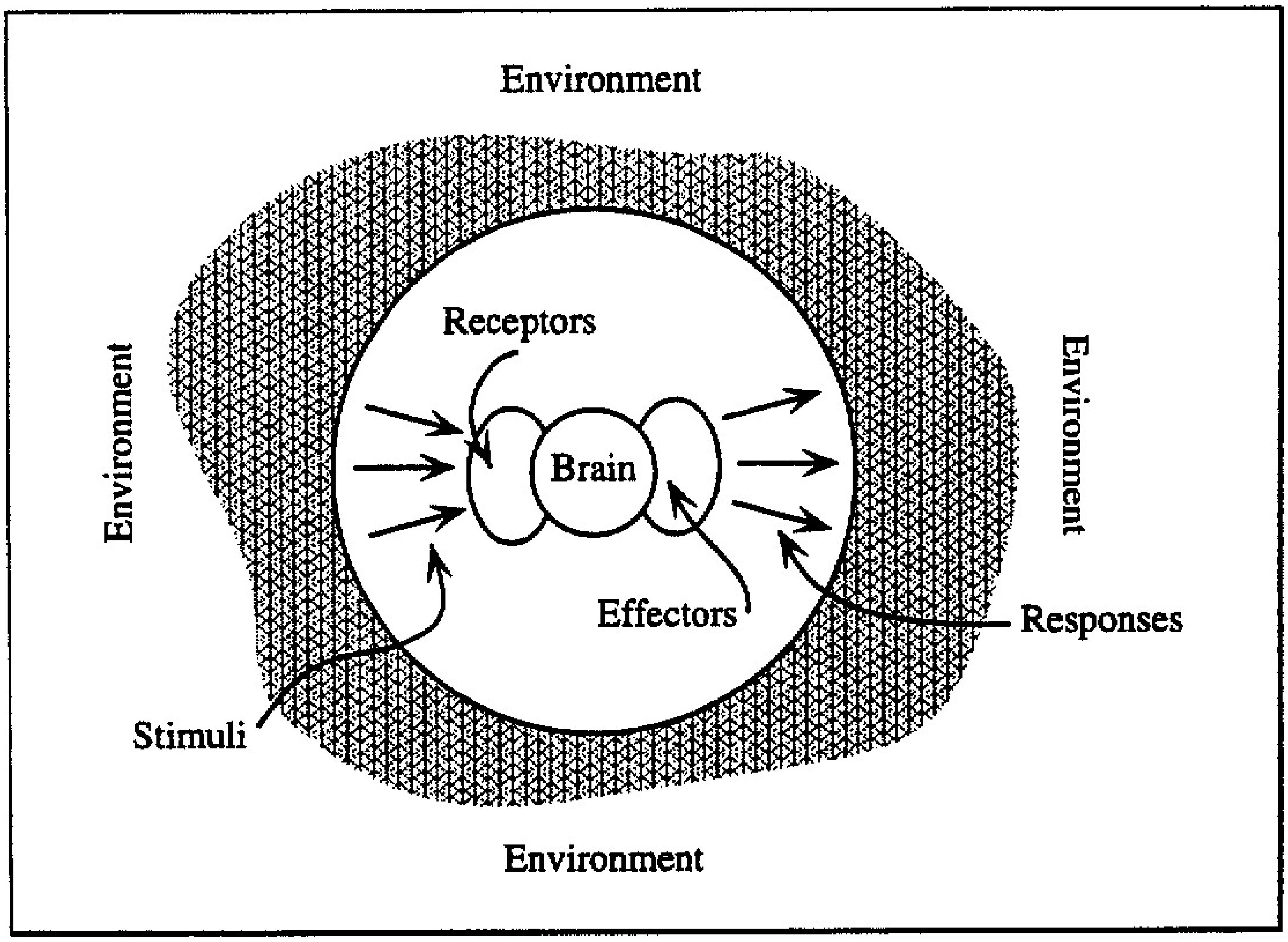

Figure 2: Topics of interest to any theory of visual perception. (1) The environment: the physical world of surfaces and objects, which we assume to exist independently of the perceiver. (2) Incoming stimulation: objects in the world give rise to events, some of which can be detected by perceivers. (3) Receptors (sensory surfaces and peripheral neurons): before a perceiver can respond to stimuli, stimulus energy must be converted into a neural (or computer) code. (4) The brain. (5) Effector systems. (6) Motor responses.

1.1.2. Vision Tasks

If a scene could be completely arbitrary, not very much could be inferred about it by analyzing images. The gray levels of the pixels in an image measure the amounts of light received by the sensor from various directions. Any such set of brightness measurements could arise in infinitely many different ways as a result of light emitted by a set of light sources, transmitted through a sequence of transparent media, and reflected from a sequence of surfaces.

The goal of vision becomes feasible only if restrictions are imposed on the class of possible scenes. These restrictions may be very broad (the class of scenes that can occur in a terrestrial environment, with which an agent might have to deal), or they may be very narrow (e.g., the class of scenes that can occur on a conveyor belt in an industrial facility). The central problem of vision can then be reformulated as follows: given a set of constraints on the allowable scenes, and given a set of images obtained from a scene that satisfies these constraints, derive a description of that scene. It should be pointed out that unless the given constraints are very strong, or the given set of images is large, the scene will not be uniquely determined; the images only provide further constraints on the subclass of scenes that could have given rise to them, so that only partial descriptions of the scene are possible.

Theoretical (what could be)	What kind of descriptive information can be derived about a scene from an analysis of its images?
Empirical (what is)	How are the visual systems of existing organisms designed?
Purposive (what should be)	How should the visual system of an organism or robot be designed so that it can best perform a set of tasks?

Figure 3: Three different questions that we can ask about the processes involved in visual perception. The theoretical question is a subject of research in mathematics and physics while the empirical question is studied in zoology, psychology, neuroanatomy, etc. Purposive or normative questions are of interest to engineering disciplines. The three questions, although related, do not necessarily have the same answer. What exists could be suboptimal and only a small subset of what is theoretically possible could be of practical relevance to autonomous real-time intelligent "seeing" systems.

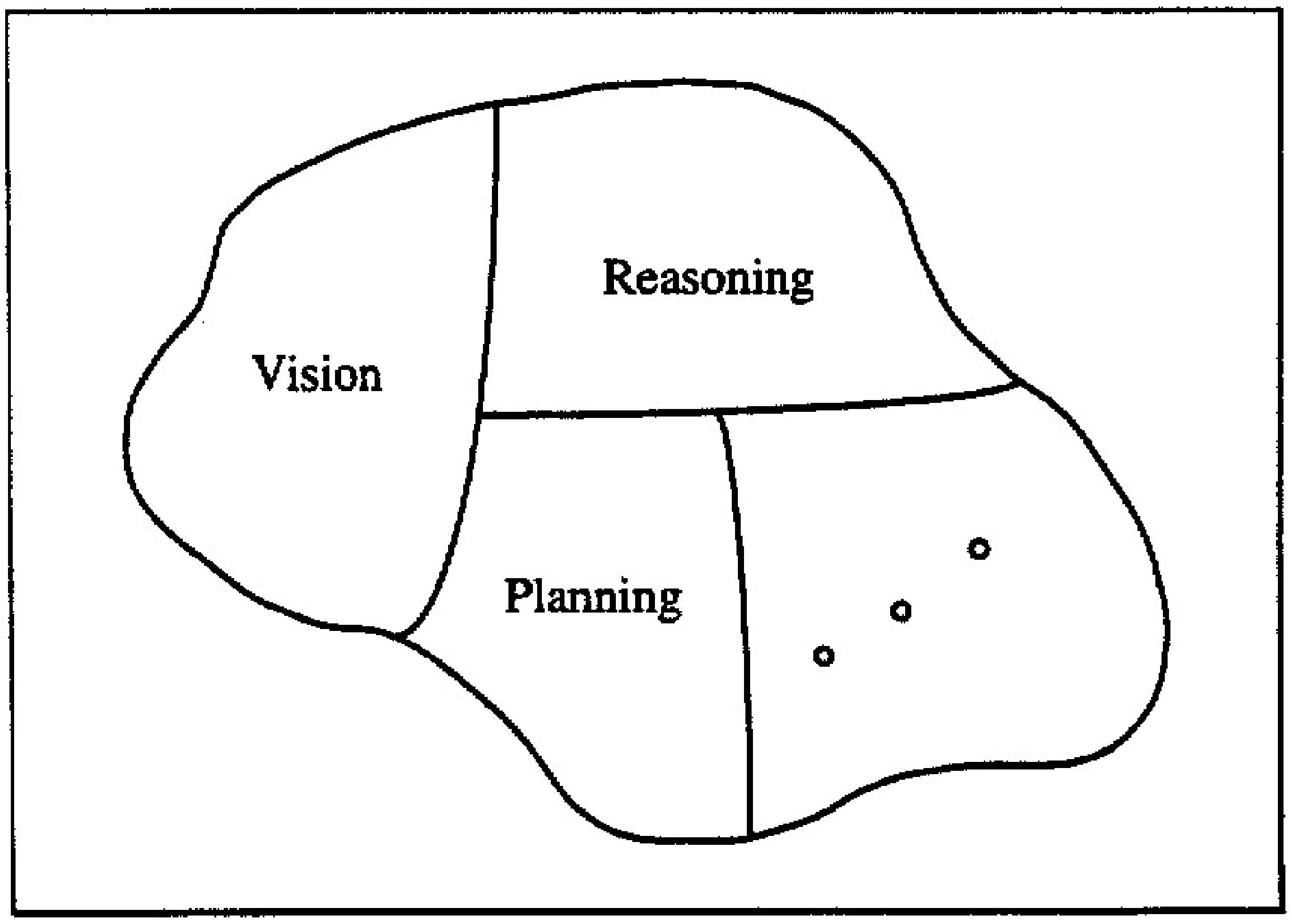

Figure 4: The theoretical approach views vision as a mechanical act whose purpose is to perform recovery of the scene. Vision is studied in isolation and it amounts to deriving general-purpose descriptions of the visible world. These descriptions are then given as inputs to cognitive processes (reasoning, planning, etc.), appropriately modified as necessary to suit the needs of these processes.

Vision tasks vary widely in difficulty, depending on the nature of the constraints that are imposed on the class of allowable scenes and on the nature of the partial descriptions that are desired. The constraints can vary greatly in specificity. At one extreme, they may be of a general nature—for example, that the visible surfaces in the scene are all of some "simple" type (e.g., quadric surfaces with diffuse reflectivities), and that the illumination consists of a single distant light source. Note that the surfaces may be "simple" in a stochastic rather than a deterministic sense; for example, they may be fractal surfaces of given types, or they may be smooth surfaces (e.g., quadric) with spatially stationary variations in reflectivity (i.e., uniformly textured surfaces). At the other extreme, the constraints may be quite specialized—for example, that the scene contains only objects having given geometric ("CAD") descriptions and given optical surface characteristics.

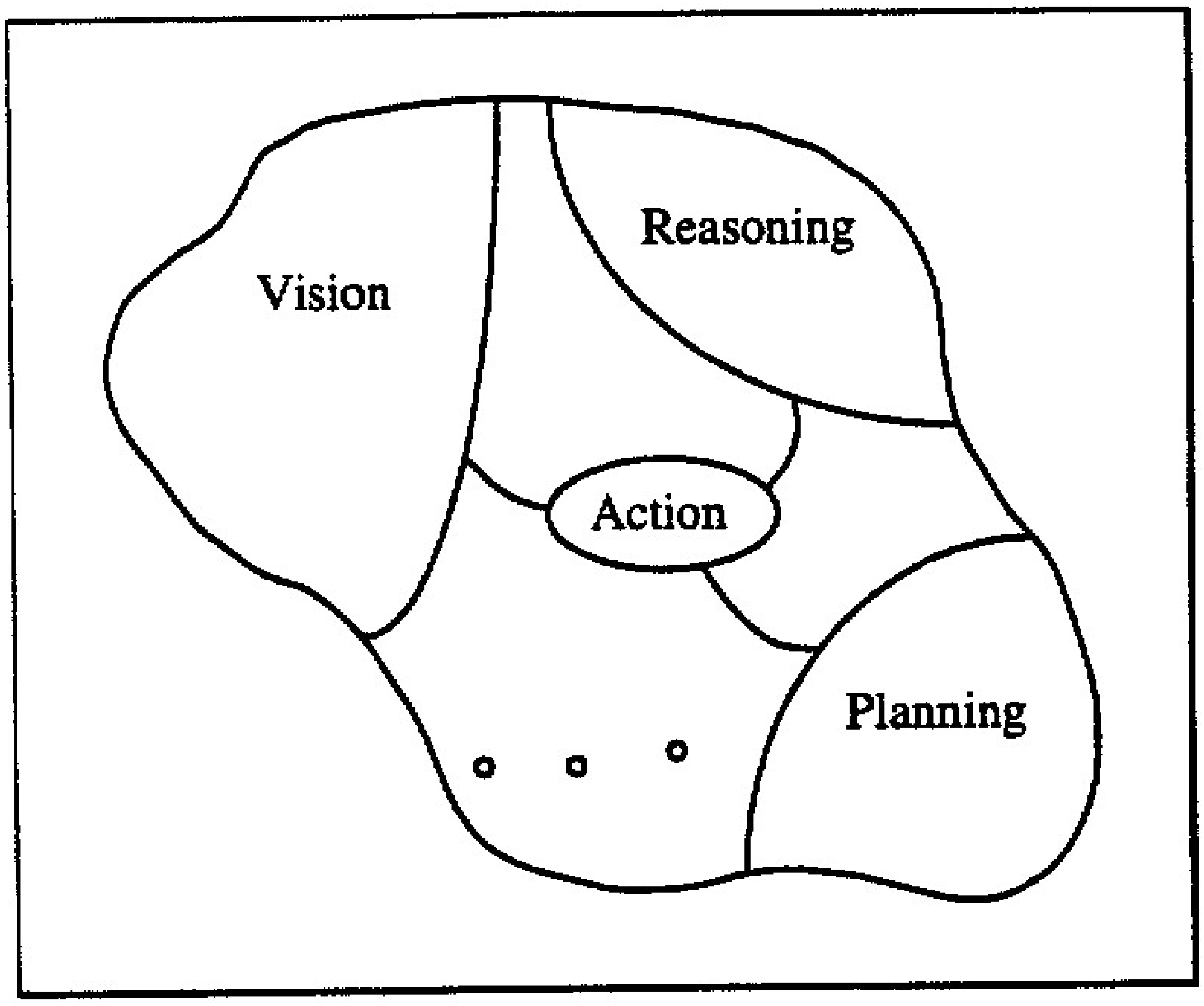

Figure 5: The purposive or normative approach views vision as creating an interface between the world and the cognitive processes (reasoning, planning, etc.) for the purpose of taking an action. Vision is studied as a part of a larger system and creates partial descriptions of the world that are not general-purpose, but that make sense in conjunction with a task or action (or a set of them). An action can be practical (like a motor command), theoretical (like reaching a decision or building a specific representation), or aesthetic.

Similarly, the desired scene descriptions can vary greatly in completeness. In the theoretical approach, a vision task is a recovery task, i.e. the development of a general-purpose description. In the purposive approach, on the other hand, a vision task is the development of a description sufficient for achieving a specific goal; common classes of goals involve navigation and recognition. Recovery tasks call for descriptions that are as complete as possible, but recognition and navigation tasks usually require only partial descriptions—for example, identification and location of objects or surfaces of specific types if they are present in the scene. This classification of vision tasks is illustrated schematically in Figure 6. Note that by definition, recovery tasks require correct descriptions of the scene; but recognition and navigation tasks can often be performed successfully without completely describing even the relevant parts of the scene. For example, obstacles can often be detected, or object types identified, without fully determining their geometries.

In its earliest years (beginning in the mid-1950's), machine vision research was concerned primarily with recognition tasks, and dealt almost entirely with single images of (essentially) two-dimensional scenes: documents, photomicrographs (which show thin "slices" of the subject, because the depth of field of a microscope image is very limited), or high-altitude views of the earth's surface (which can be regarded as essentially flat when seen from sufficiently far away).

The mid-1960's saw the beginnings of research on robot vision; since a robot must deal with solid objects at close-by distances, the three-dimensional nature of the scene cannot be ignored. Research on recovery tasks began in the early 1970's, initially considering only single images of a static scene, but by the mid-1970's beginning to deal with time sequences of images (of a possibly time-varying scene) obtained by a moving sensor.

CLASSES OF VISION TASKS

	General constraints →	Specialized constraints
Complete information ↓	"GENERAL" RECOVERY	"BIN OF PARTS" RECOVERY
Partial information	NAVIGATION TASKS QUALITATIVE RECOVERY	OBJECT RECOGNITION

Figure 6: A classification of visual problems (tasks) along dimensions of generality vs. specificity in the assumptions made, the constraints employed or the problems considered, and completeness vs. partialness in the amount of information about the scene that needs to be recovered.

Forty years of research have produced theoretical solutions to many vision problems; but many of these solutions are based, explicitly or tacitly, on unrealistic assumptions about the class of allowable scenes, and as a result, they often perform unsatisfactorily when applied to real-world images. As we shall see in the next section, even for static, two-dimensional scenes, many vision problems are ill-posed, ill-defined, or computationally intractable.

Readers interested in more details about vision tasks and techniques may consult any of a large number of textbooks, monographs, and paper collections; we cite here only the texts by Ballard and Brown (1982), and Rosenfeld and Kak (1982), and the monographs by Marr (1982) and by Aloimonos and Shulman (1989).

1.1.3. Sources of Difficulty

1.1.3.1. Ill-posedness

As already mentioned, the gray levels of the pixels in an image represent the amounts of light received by the sensor from various directions. If the scene does not contain transparent objects (other than air, which we will assume to be clear), the light contributing to a given pixel usually comes from a small surface patch in the scene (on the first surface intersected by a line drawn from the sensor in the given direction). This surface patch is illuminated by light sources, as well as by light reflected from other patches. Some fraction of this illumination is reflected toward the sensor and contributes to the pixel; in general, this fraction depends on the orientation of the surface patch relative to the direction(s) of illumination and the direction of the sensor, as well as on the reflectivity of the patch. In short, the gray level of a pixel is the resultant of the illumination, orientation, and reflectivity of a surface patch. If these quantities are unknown, it is not possible in general to recover them from the image.

This example is a very simple illustration of the fact that most vision problems are "ill-posed", i.e., under-constrained; they do not have unique solutions. Even scenes that satisfy constraints usually have more degrees of freedom than the images to which they give rise; thus even when we are given a set of images of a scene, the scene is usually not uniquely determined.

In applied mathematics, a common approach to solving ill-posed problems is to convert them into well-posed problems by imposing additional constraints. A standard method of doing this, known as regularization, makes use of smoothness constraints; it finds the solution that minimizes some measure of non-smoothness (usually defined by a combination of derivatives). Regularization methods were introduced into vision in the mid-1980's, and have been applied to many vision problems. Evidently, however, solutions found by regularization often do not represent the actual scene; for example, the actual scene may be piecewise smooth, but may also have discontinuities, and a regularized solution tends to smooth over these discontinuities. To handle this problem, more general approaches have been proposed which allow discontinuities, but which minimize the complexity of these discontinuities—e.g., minimize the total length and total absolute curvature of the borders between smooth regions. In effect, these approaches find solutions that have minimum-length descriptions (since the borders can be described by encoding them using chain codes). However, the actual scene is not necessarily the same as the scene (consistent with the images) that has the simplest description. Evidently, not all scenes of a given class are equally likely; but the likelihood of a scene depends on the physical processes that give rise to the class of scenes, not on the simplicity of a description of its image.

1.1.3.2. Ill-definedness

It is often assumed in formulating vision problems that the class of allowable scenes is "piecewise simple"—e.g., that the visible surfaces are all smooth (e.g., planar or quadric) and diffusely reflective. This type of assumption seems at first glance to strongly constrain the class of possible scenes (and images), but in fact, the class of images is not constrained at all unless a lower bound is specified on the sizes of the "pieces". If the pieces can be arbitrarily small, each pixel in an image can represent a different piece (or even parts of several pieces), so that the image can be completely arbitrary. For a two-dimensional scene, it suffices to specify a lower bound on the piece sizes; but for a three-dimensional scene, even this does not guarantee a lower bound on the sizes of the image regions that represent the pieces of surface; occlusions and nearly-grazing viewing angles can still give rise to arbitrarily small or arbitrarily thin regions in the image.

Lower bounds on piece sizes are desirable for another very important reason: they make it easier to distinguish between the ideal scene and various types of "noise". In the real world, piecewise simple scenes are an idealization; actual surfaces are not perfectly planar or quadric or perfectly diffuse reflectors, but have fluctuating geometries or reflectivities. (Note that these fluctuations are in the scene itself; in addition, the brightness measurements made by the sensor are noisy, and the digitization process also introduces noise.) If the fluctuations are small relative to the piece sizes, it will usually be possible to avoid confusing them with "real" pieces. (Similarly, the noisy brightness measurements—assuming that they affect the pixels independently—yield pixel-size fluctuations, and digitization noise is also of at most pixel size; hence these types of noise too should usually not be confused with the pieces.) Of course, even if we can avoid confusing noise fluctuations with real scene pieces, their presence can still interfere with correct estimation of the geometries and photometries of the pieces.

Most analyses of vision problems (e.g., for piecewise simple ideal scenes) do not attempt to formulate realistic models for the "noise" in the scene; they usually assume that the noise in the image (which is the net result of the scene noise, the sensor noise, and the digitization noise) is Gaussian and affects each pixel independently. Examination of images of most types of real scenes shows that this is not a realistic assumption; thus the applicability of the resulting analyses to real-world images is questionable.

The problem of ill-definedness becomes even more serious if one attempts to deal with scenes containing classes of objects that do not have simple mathematical definitions—for example, dogs, bushes, chairs, alphanumeric characters, etc. Recognition of such objects is not a well-defined vision task, even though humans can recognize them very reliably.

1.1.3.3. Intractability

Even well-defined vision problems are not always easy to solve; in fact, they may be computationally intractable. An image can be partitioned in combinatorially many ways into regions that could correspond to simple surfaces in the scene; finding the correct (i.e., the most likely) partition may thus involve combinatorial search. For example, even for scenes consisting of wireframe polyhedral objects, the problem of deciding whether a set of straight lines in an image could represent such a scene is NP-complete. Even identifying a subset of image features that represent a single object of a given type is exponential in the complexity of the object, if more than one object can be present in the scene, or if the features can be due to noise.

Parallel processing is widely used to speed up vision computations; it is also used very extensively and successfully in biological visual systems. Very efficient speedup can be achieved through parallelism in the early stages of the vision process, which involve simple operations on the image(s); but little is known about how to efficiently speed up the later, potentially combinatorial stages. Practical vision systems must operate in "real time" using limited computational resources; as a result, they are usually forced to use suboptimal techniques, so that there is no guarantee of correct performance.

In principle, the computations performed by a vision system should be chosen to yield maximal expected gain of information about the scene at minimal expected computational cost. Unfortunately, even for well-defined vision tasks, it is not easy to estimate the expected gain and cost. Vision systems therefore usually perform standardized types of computations that are not necessarily optimal for the given scene domain or vision task; this results in both inefficiency and poor performance.

1.1.4. Recipes for Success

Define Your Domain (What is the system's perceptual world?)

Well-defined vision problems should involve classes of scenes in which both the ideal scene and the noise can be mathematically (and probabilistically) characterized. For example, in scenes that contain only known types of man-made objects, the allowable geometric and optical characteristics of the visible surfaces can be known to any needed degree of accuracy. If the objects are "clean", and the characteristics of the sensor are known, the noise in the images can also be described very accurately. In such situations, the scene descriptions that are consistent with the images are generally less ambiguous (so that the problem of determining these descriptions is relatively well-posed) because of the restricted nature of the class of allowable scenes. If, in addition, the number of objects that can be present is limited, the complexity of the scene description task and the computational cost of recognizing the objects are greatly reduced. Referring back to Figure 2, these remarks imply that in studying vision we have to consider the environment in which the visual system needs to operate. If the environment is very complex, we can attempt to determine those aspects of the environment which it is necessary for the system to recognize—in other words, we can attempt to define the system's perceptual world.

Pick Your Problem (What are the system's tasks?)

Even for specialized scene domains, deriving complete scene descriptions from images—the general recovery problem—can still be a very difficult task. However, there is no reason to insist on complete recovery. The images (further) constrain the class of possible scenes; the task of the vision system is to determine these constraints. This yields a partial description of the scene, and for some purposes this description may be sufficient. In fact, in many situations only specific kinds of partial descriptions of the scene are needed, and such descriptions can often be derived inexpensively and reliably. A partial description may require only the detection of a specific type of object or surface, if it is present, or it may require only partial ("qualitative") characterizations of the objects that are present (e.g., are their surfaces planar or curved?).

Improve Your Inputs

Vision tasks that are very difficult to perform when given only a single image of the scene generally become much easier when additional images are available. These images could come from different sensors (e.g., we can use optical sensors that detect energy in different spectral bands; we can use imaging sensors of other types such as microwave or thermal infrared; or we can use range sensors that directly measure the distances to the visible surface points in the scene). Alternatively, we can use more than one sensor of the same type—for example, stereo vision systems use two or more cameras. Even if we use only a single sensor, we can adjust its parameters—for example, its position, orientation, focal length, etc.—to obtain multiple images; control of sensor parameters in a vision system is known as *active vision*. It has been shown that by using the active vision approach, ill-posed vision problems can become well-posed, and their solutions can be greatly simplified. These improvements are all at the sensor level; similarly, one can improve the inputs to the higher levels of the vision process by extracting multiple types of information from the image data using different types of operators.

Take Your Time

Since the early days of machine vision, the power of general-purpose computational resources has improved by many orders of magnitude, as regards both processing speed and memory capacity. This, combined with the availability of special-purpose parallel hardware, both analog and digital (VLSI), has greatly expanded the range of tractable vision tasks. The availability of increasingly powerful computing resources allows the vision system designer much greater freedom to adopt an attitude of "take your time": use vision algorithms that are as complex as necessary, and use as much input or intermediate data as necessary, without being overly concerned with current computational costs. With no end yet in sight as regards expected improvements in computing power, the time required to solve given vision problems will continue to decrease. Conversely, it will become possible to solve problems of increased complexity and problems that have wider domains of applicability.

1.2. Perspectives on Machine Learning

1.2.1. Introduction

Machine learning is concerned with the development of computational models of processes by which a system can acquire or improve its knowledge or skills. Because the types of knowledge or skills that may be acquired and the methods for doing this can vary very greatly, there is a tremendous diversity of possible learning approaches and techniques. Learning may involve, for example, building general object descriptions from specific observations, acquiring problem solving methods on the basis of examples of solutions, improving algorithms through practice or experimentation, constructing control heuristics from experience, creating solutions to new problems by analogy with solutions to similar problems, discovering statistical regularities or logical relationships among entities, and so on.

The underlying theme of all these processes is that a learning system constructs or improves some type of knowledge structure or knowledge representation. The great variety of views, methods, and approaches that have been developed in the field of machine learning differ in their assumptions about

—what is known a priori to the learner,
—what and how input information is provided to the system,
—what type of knowledge the learning system is trying to acquire,
—how this knowledge is represented,
—what inferential or computational mechanisms are used to acquire it, and
—what criteria and methods are used to evaluate the results of learning.

The study of learning and its relationship to other aspects of intelligence has puzzled philosophers and scientists for a long time. Consider, for example, the following sentence about knowing (Wittgenstein, 1958): "The grammar of the word 'knows' is evidently closely related to that of 'can', 'is able to'. But it is also closely related to that of 'understands'." Since learning includes such processes as acquiring knowledge, the ability to perform some acts, or understanding of some facts, it is clear that all these concepts are intimately related. The question of how learning is accomplished has divided philosophers into two opposing parties, the empiricists and the rationalists. Similar divisions exists among the scientists who study computational approaches to learning processes. There is a large subfield of research concerned with developing and experimenting with learning methods and systems ("Experimental Machine Learning"), and there is a subfield concerned with mathematically analyzing formal properties of various learning algorithms ("Computational Theory of Learning" or COLT). These two subfields have held, until very recently, separate workshops and conferences. To encourage interaction, in 1994, the annual conferences in these two subfields were held in the same location, during partially overlapping periods (Cohen and Hirsh, 1994; Warmuth, 1994). There exists also a large amount of literature in statistics and in psychology devoted to learning problems.

In various studies of learning, a lot of effort has been devoted to ways of formally representing learning behavior from the "black box" (stimulus-response) viewpoint. Two theoretical approaches can be distinguished: statistical and deterministic. The statistical approach describes the behavior in probabilistic terms, aiming at developing statistical models of learning behaviors. It characterizes a learning behavior by a probability distribution, $p(x)$, where x ranges over the space $D = S \times R$, the Cartesian product of the stimulus space and the response space. After learning is accomplished, the system's decisions are generated on the basis of the acquired probability distribution. A related model involves representing the conditional probability $p(r|s)$, defined as $p(r|s) = p(s,r)/p(s)$. Although the statistical approach is theoretically appealing and very general, it has a pragmatic limitation, namely, that real applications rarely provide enough data to determine desirable probability distributions and construct statistically significant relationships. An interesting survey of efforts to view learning as a problem of statistical estimation is in (Landelius, 1993).

The deterministic approach describes learning processes in terms of algorithms, symbolic descriptions or mathematical transformations. It tackles issues that are central for implementing learning processes, such as how to construct and manipulate structures representing the knowledge or skill to be acquired, how to derive them from a given set of facts, how to utilize prior knowledge, and how to model different types of learning, such as inductive learning, knowledge-based deductive learning ("explanation-based learning") or analogy-based learning. The evaluation of the results obtained from applying deterministic algorithms to specific problems and the characterization of the formal properties of algorithms usually involve methods of statistics. Ultimately, the deterministic approach will have to incorporate statistical ideas and techniques in order to provide estimates of the plausibility of the developed models, solutions and results.

This section provides an overview of general ideas and basic methodologies of machine learning, concentrating primarily on methods representing the deterministic approach, developed mainly by the artificial intelligence community. The review does not try to prejudge which methodologies may ultimately play central roles in machine vision. Instead, it attempts to present a broad characterization of major approaches, including those that at present seem unlikely to be particularly applicable to vision. We begin by presenting a general view of learning processes.

1.2.2. Basic Components of Learning Processes

Any form of learning can be characterized as a process of acquiring a model of some real or abstract entity . This model may be in the form of declarative knowledge, procedural knowledge (algorithms, skills), or a combination of both. A learning process can be generally represented by the functional diagram in Figure 7. **Input** stands for information that the learner receives from an external source. **Interface** transforms this input into a form needed by the learning system, or generates requests for new information. The input transformation may involve a mapping of the input into a new representation space, e.g., by applying transformations that enhance relevant information and disregard information irrelevant to the learning goal. **Memory** contains knowledge components that affect the learning process. One of these is the **learning goal**, which specifies criteria for evaluating knowledge to be acquired and controlling the learner's attention mechanism. Another component is the learner's **prior knowledge** relevant to the learning goal (**background knowledge**).

The input is processed by **Inference Mechanism**, which applies "knowledge operators" to the input and background knowledge to generate desired knowledge. These operators make modifications or transformations of the knowledge structures residing in **Memory** and/or those obtained as input. Classes of such operators include various forms of generalization, specialization, abstraction, concretion, explanation, prediction, selection, generation, and others (Michalski, 1994). Depending on the knowledge representation employed, these operators can be implemented in many different ways.

The resulting knowledge undergoes evaluation in the **Evaluation** module in accordance with the learning goal. If the goal is satisfied, the learning process stops; otherwise, it continues.

The modules in Figure 7 may be implemented in learning systems in many different ways. For example, in an artificial neural net, **Memory** consists of the structure of the network and the settings of the weights of the interconnections between the network units. The learning goal is defined by the designer and embedded in the way the network operates. The function of **Interface** is also usually performed by a human expert. **Inference Mechanism** makes modifications of the weights according to some algorithm (e.g., backpropagation). These modifications produce a new state of the network that represents the system's new knowledge.

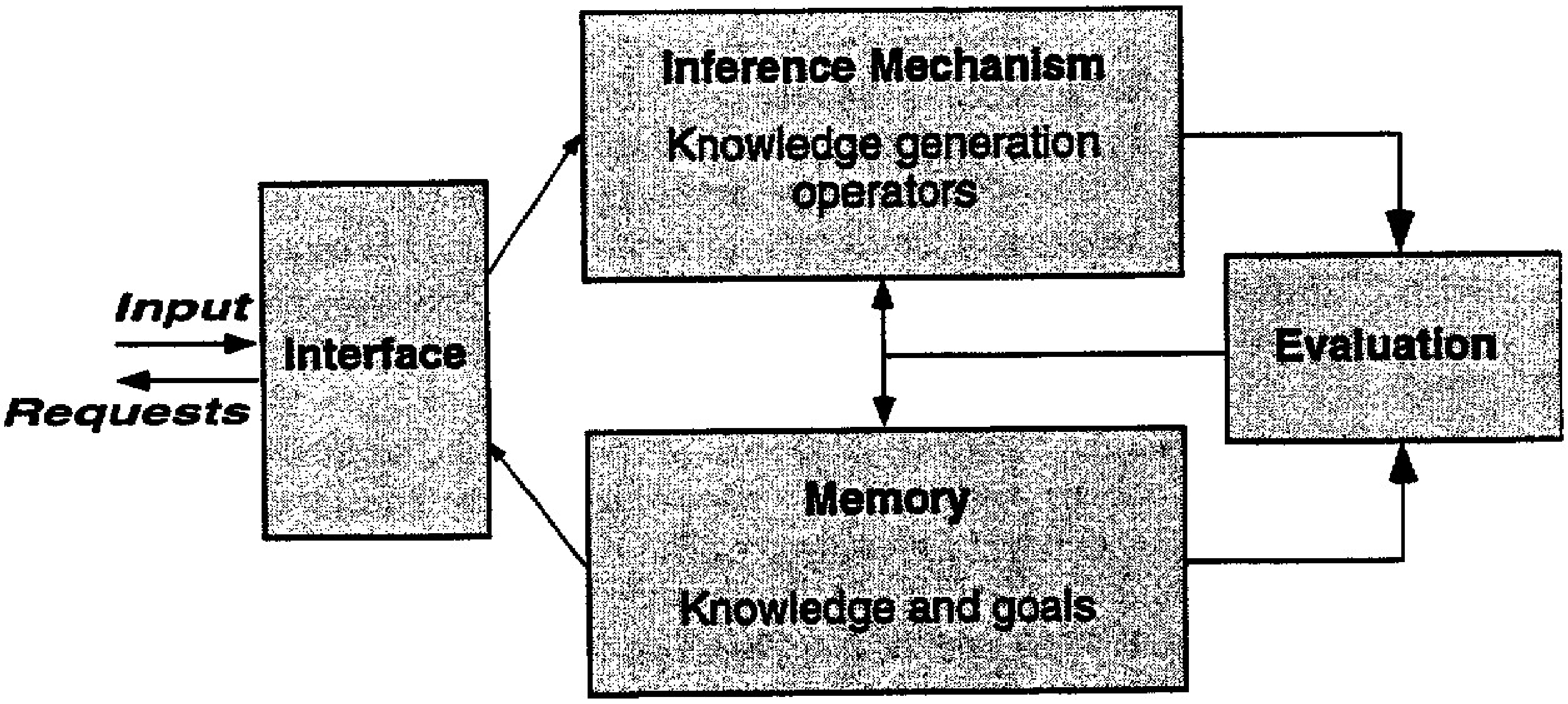

Figure 7: A functional diagram of a learning system. Different types of learning differ in terms of the knowledge operators used by the Inference Mechanism. In inductive learning, the Inference Mechanism inductively generalizes the input. In deductive learning, it draws deductive inferences from the input and from background knowledge, In learning by analogy, it derives new knowledge structures by modifying preexisting structures that represent knowledge similar to the desired. Different learning systems may use very different computational and representational mechanisms for accomplishing these functions.

In symbolic inductive learning, **Inference Mechanism** creates symbolic knowledge structures (e.g., rules, trees, grammars, etc.) representing a hypothesis induced from the input and current background knowledge (such a process can be described as applying "inductive generalization rules"; Michalski, 1983). **Evaluation** evaluates the resulting hypotheses and selects the "best" one according to criteria reflecting the learning goal. In many existing systems, the function of **Interface** is performed by a user; in more recent systems, **Interface** performs complex multistep transformations of the initial knowledge representation space (e.g., Wnek and Michalski, 1994a,b).

As mentioned above, both machine vision and machine learning systems aim at creating descriptions of entities. Thus, in an abstract sense, the two fields have similar goals. In a pragmatic sense, however, they are very different and have followed different paths.

a) Most machine learning methods assume (notable exceptions include artificial neural nets and adaptive control systems) that the inputs are preprocessed symbolic quantities. Machine vision systems, on the other hand, deal specifically with the interpretation of visual sensory signals. Consequently, if such machine learning systems are applied to the problems requiring understanding of a visual scene, they need vision systems to generate symbolic inputs for them.

b) In machine learning, knowledge representations are often symbolic structures, such as decision rules, decision trees, semantic networks, etc. (except, of course, for systems that learn numerical parameters of given mathematical expressions, or connection weights in artificial neural nets). In contrast, vision systems much more frequently employ geometric models or sets of equations.

c) Machine learning concerns an automated acquisition of new or better knowledge, as well as the acquisition of skills and control procedures. While both types of problems have parallels in machine vision (the first relates to building models of objects or scenes from images, and the second to controlling parameters of the image acquisition system, e.g., a camera, as in active vision), most research in machine vision has been primarily concerned with the first class of problems.

One of the advantages of symbolic learning systems is that the knowledge they acquire is usually easy to explain and relate to human knowledge. This is, obviously, an important factor for knowledge-based systems. The above feature is missing in "subsymbolic" learning systems, such as artificial neural nets, in which learned knowledge resides in new values of the weights, and it is difficult to translate it to a form comprehensible by humans. In many machine vision applications, however, this feature may not be important.

Evidently, systems that can handle continuous numerical attributes and transformations, such as artificial neural nets and other connectionist systems, are readily applicable to many vision problems. It is much less clear how symbolic learning systems can be utilized in machine vision. At higher levels of processing, however, visual information (combined with other sensory information) may be converted into a form amenable to symbolic reasoning; at these levels, symbolic learning becomes applicable.

Questions arise as to what machine learning approaches are most appropriate to what kind of vision problems, and how they can be effectively applied to these problems. As a starting point in dealing with these questions, the next section provides a classification of existing machine learning methodologies and attempts to briefly characterize their applicability to vision problems.

1.2.3. A Classification of Machine Learning Methodologies

Over the years, machine learning research has developed a number of methodologies. Each methodology is oriented toward a somewhat different learning task and often uses a different computational or representational mechanism. A *learning task* is specified by the type of input information (i.e., the information provided to the learner through its senses), the learning goal (which defines the knowledge to be acquired through learning), and the background knowledge (i.e., the learner's prior knowledge relevant to the learning goal). In order to explore the applicability of machine learning to vision, the following paragraphs summarize major learning methodologies. Many of these methodologies are clearly applicable to various vision problems.

- **Symbolic learning from examples**

Inducing general concept descriptions from examples ("supervised learning"). Descriptions can be *attributional* (expressed in terms of attributes—continuous or discrete) or *structural* (expressed in terms of attributes and relations). A learning process can be *empirical* (with little background knowledge) or *constructive* (with enough knowledge to be able to generate additional problem-oriented attributes and concepts). The descriptions can be in the form of decision trees, decision rules, semantic networks, frames, grammars, etc. Most methods make few assumptions about the knowledge to be learned, and can be directly applied to many problems. Symbolic methods are particularly useful for domains in which it is important that the learned knowledge be comprehensible by a human expert. Their main limitation is that they are primarily oriented toward learning problems characterized by symbolic variables. Symbolic learning methods, especially those using decision tree or decision rule representations, have been widely applied to classification problems in many different domains, including vision.

- **Connectionist (artificial neural net) learning**

Inducing transformations representing desired input-output behaviors by determining appropriate weights in connectionist systems. The most active research in this area involves representations based on artificial neural networks. Systems of this type can be applied to a wide range of vision problems; they seem to be particularly useful for learning continuous transformations. Their disadvantages are that learning processes tend to require many iterations and thus are slow, and that the learned knowledge is in a form incomprehensible to humans. It is also difficult to incorporate background knowledge into such systems or perform explicit forms of inference or knowledge transformations.

- **Genetic algorithm-based learning**

Iteratively modifying knowledge structures by random or partially random operators, and selecting best performing structures (according to some performance measure) for the next iteration. These methods are particularly useful for searching highly unstructured problem spaces. They are relatively slow and are not desirable when relevant background knowledge is readily available and can be used to guide the learning process.

- **Case-based learning**

Storing past cases (situation-decision pairs, solutions to past problems, etc.) as representatives of concepts or problem solutions. New concept instances are recognized or new problems are solved by matching them with the most similar past cases. The matching may involve complex transformations and inference. These methods are relatively easily to implement, but are not useful when it is important to seek general problem solutions, or to relate concept descriptions to each other, determine their differences, etc.

- **Quantitative and/or qualitative discovery**

Discovering equations characterizing a collection of data. These methods differ from conventional curve fitting or regression analysis algorithms in that they make weaker assumptions about the underlying form of the equations, and utilize heuristics in the process of equation discovery. More advanced methods can also formulate symbolic rules or conditions characterizing the ranges of applicability of the created equations.

- **Explanation-based learning**

Deductively deriving effective ("operational") concept descriptions (or control rules) from abstract ones, using an input example to guide the process. The methods trade the cost of obtaining examples and conducting inductive inference for the cost of handcrafting abstract concept descriptions and drawing deductive inferences from them. Unlike inductive learning methods, explanation-based methods produce descriptions which are as valid as the abstract descriptions.

- **Reinforcement learning**

Learning a mapping from situations to actions so that a reward function is maximized. The difference between the observed and desired behavior drives the modification of the system's parameters. These methods are particularly relevant to sensor-based control (e.g., for navigation).

- **Statistical learning**

A class of methods that characterize learning processes by probability distributions in a priori given representation spaces, such as stimulus-response space or feature space. The probability distributions are estimated on the basis of given facts or observations. Two approaches to doing this stand out. One is to let the system model the average mapping from stimuli to responses, and then apply some form of distribution with this mapping as its mean value. Another approach is to let the system estimate an analytic form of the distribution of the decisions, and calculate the specific distribution from the stimulus-response pairs.

- **Clustering ("unsupervised learning")**

Organizing a collection of entities (objects, observations, etc.) into clusters or classes, or a hierarchy of such clusters. The most common approach is to use some measure of similarity (or distance) between entities, and seek clusters for which intracluster similarities are high and intercluster similarities are low. Another approach, called "conceptual clustering," was developed by AI researchers. Instead of using an a priori given measure of similarity, it employs "conceptual cohesiveness," which uses a measure of fit between clusters and concepts that can be used to characterize the clusters. In contrast to conventional clustering, conceptual clustering produces not only clusters but also generalized symbolic descriptions of the clusters.

- **Learning by analogy**

Learning a new concept (or solving a new problem) by adapting and modifying the description (or solution) of a previously learned similar concept (or problem solution). Learning by analogy is related to case based learning. The difference is that analogical learning employs generalized knowledge structures rather than cases.

- **Abductive learning**

Creating explanations of given facts/solutions, etc. by tracing backward domain-dependent implicative relations.

- **Multistrategy learning**

Integrating multiple inferential strategies (e.g., inductive learning with explanation-based learning) and/or multiple computational strategies (e.g., symbolic learning with genetic algorithm based learning) in a learning process; this is one of the newest and most challenging research directions in machine learning.

As mentioned earlier, in recent years computational studies of learning have split into two related but distinct subfields: experimental machine learning, whose primary concern is to develop methods and implement effective learning systems, and computational theory of learning, whose primary concern is to study formal properties of various learning algorithms (for example, to determine the convergence of inductive learning algorithms, the "learnability" of different types of descriptions, the relationship between the number of training examples and the error rate of the learned descriptions, etc.).

1.2.4. Types of Learning Problems

While the above classification gives a sense of what types of learning methodologies have been developed and for what purposes, it does not give much insight into the nature and diversity of learning problems. Therefore, below we attempt to give a general classification of major types of learning problems. To this end, we assume that learning problems can be viewed as the determination of a complete description of a function:

$$f: D_1 \times D_2 \times D_3 \times \cdots \times D_n \rightarrow D^1 \times D^2 \times D^3 \times \ldots D^m$$

where $D_i, i = 1,2,\ldots,n$ and $D^j, j = 1,2,\ldots,m$ are domains (value sets) of input and output variables, respectively, on the basis of limited samples of input-output pairs and the learner's background knowledge.

The domains D_i and D^j can be discrete, continuous, structured, or relational. A structured domain is a partially ordered set, e.g., a generalization hierarchy (an "is-a" hierarchy of concepts ordered by the generalization relation); a relational domain is a set of well-formed sentences in a language suitable for describing relations, e.g., first-order predicate logic. The domains of the input variables represent legal value sets of "descriptors" (attributes or relations) that are used to characterize entities that the system is learning about. The domains of the output variables are sets of values that can be assigned to any entity. Different types of learning problems make different assumptions about the sets D_i and D^j, and the background knowledge available about the function *f*.

If the input domains are discrete, continuous or structured sets (i.e., not relational), they can be viewed as legal value sets of certain attributes (zero-argument relations), in which case we have *attributional* learning. If the domains are relational, then we have *relational* (or *structural*) learning. For example, learning a decision tree, a numerical equation, or a set of weights in an artificial neural net, is a form of attributional learning; learning a predicate logic description of a scene is relational learning. Different types of learning problems are characterized in the following subsections.

1.2.4.1. Single Concept Learning: D_i are any sets; D is binary ($m = 1$)

The input domains can be any sets, i.e., discrete, continuous, structured or relational; the output domain is binary. The values of the output domain can be interpreted as "belongs to the concept" and "does not belong to the concept".

Single concept learning problems can be of two types:

1. Learning from examples and counterexamples ("positive" and "negative" examples)

   The system is given both datapoints that belong to the function and datapoints that do not belong to the function, called positive and negative examples, respectively.

2. Learning from positive examples only

The system is given only datapoints that belong to the function (positive examples). For example: Given a set of datapoints, determine a simple equation that approximates these datapoints with some accuracy ε. Problems of curve fitting, function interpolation or equation discovery belong to this category.

1.2.4.2. Multiple Concept Learning: $D_i, i = 1,2,\ldots,n$ are any sets

The input domains can be any sets, as in single concept learning. The output domain is a set of concepts from a certain class. Depending on the output domain D, learning problems can be of two types:

1. Learning disjoint concepts: D is an unordered discrete set

   The output domain D is a set of names of concepts belonging to a given class.

2. Learning overlapping concepts: D^j are binary sets $(m > 1)$.

   Each output domain is associated with one concept to be learned. For example: Given a set of entities, learn to assign to each of them various characteristics; a given object can be assigned many such characteristics.

1.2.4.3. Learning Continuous Transformations: D_i and D are continuous sets $(m = 1)$

The input and output domains are continuous sets—the domains of input and output variables characterizing a transformation. For example: Learn a transformation that maps a set of state variables into a control variable (e.g., in visuo-motor control).

1.2.4.4. Learning Ranked Concepts: $D_i, i = 1,2,\ldots,n$ are any sets, D is an ordered discrete set

The input domains can be any sets; the output domain is a linearly ordered discrete set. For example: Learn to rank entities by magnitude, complexity, etc.

1.2.4.5. Learning Structures: $D_i, i = 1,2,\ldots,n$ are any sets, D is a structured set

The input domains can be any sets; the output domain is a structured set, e.g., a classification hierarchy.

1.2.4.6. Learning Sequences or Procedures: D_i are any sets, D^j are discrete sets

The input domains can be any sets; the output domains are relational sets. Each output domain D^j is a set of entities (or operators) that can appear or can be executed at the j^{th} position of a sequence (or procedure). For example: Learn a control procedure for a visuo-motor function.

1.2.4.7. Learning Relational Descriptions: D_i are any sets, D is a relational set

This large class of learning problems concerns building relational descriptions of structured entities (e.g., visual scenes).

1.2.4.8. Learning and Estimation

Most research in machine learning has been concerned with empirical inductive concept learning from examples (types 1.2.4.1 and 1.2.4.2 above). Some of these methods have achieved a relatively high level of sophistication and have already demonstrated their usefulness for selected problems of machine vision. As mentioned above, in this class of problems the output domain of functions to be learned is a discrete set. A different class of problems are those of learning a continuous mapping from an input numerical space to an output numerical space (type 1.2.4.3 above). Such problems are usually viewed as problems of estimating a system that transforms inputs to outputs, given a set of examples of input-output pairs. Classical frameworks for this problem are provided by the theory of approximation and the theory of optimization.

1.2.5. The Role of Representations

As in all of AI, the determination of an appropriate representation is one of the central problems in machine learning. Most of the current machine learning methods assume that the search for the solution (desirable knowledge, hypothesis) is done in the same representation space in which the training examples are presented. In many practical problems, however, this assumption is too strong. If the representation space in inadequate, it may be difficult or impossible to learn the correct description or transformation. (For example, in vision problems, the original representation space is an array of pixels, while the descriptions to be learned are usually formulated in terms of higher level attributes, representing conceptual components of images of objects, which should be invariant to illumination, pose, and similar factors.) Thus a key problem is how to transform the original representation space into a new space that is more relevant to the learning problem at hand.

An approach to such problems has been proposed under the name of "constructive induction." A constructive induction system conducts a double search, first for the most appropriate knowledge representation space, and second for the "best" hypothesis in this space. Methods of constructive induction can be divided into three basic categories:

- *Hypothesis-driven* methods, which look for patterns in the intermediate hypotheses in order to determine desirable transformations of the representation space (e.g., what new attributes to create, which attributes are irrelevant). A very simple method of hypothesis-driven induction is to take as a new dimension (attribute) the most important part of the description obtained at each iteration of the method (Wnek and Michalski, 1994b).
- *Data-driven* methods, which analyze input data to determine desirable modifications of the original representation space; the created new attributes may represent mathematical or logical combinations of the original attributes, or attributes created in previous iterations of the method (Bloedorn and Michalski, 1991; Bloedorn, Wnek and Michalski, 1993).
- *Knowledge-based* methods, which utilize expert-provided rules and transformations to derive higher-level attributes and determine desirable transformations of the representation space.

Automatically determining an appropriate representation space for learning is a very important, and still relatively underdeveloped, direction in machine learning research (Fawcett, 1994).

Readers interested in more details on machine learning methods and methodologies may consult a series of four books on Machine Learning [Vol. I and II (Michalski, Carbonell and Mitchell, 1983 and 1986), Vol. III (Kodratoff and Michalski, 1990), and Vol. IV (Michalski and

Tecuci, 1994)], or a number of other texts, for example, Carbonell (1990).

1.2.6. Recipes for Success

The four recipes for success in solving vision problems, discussed in Section 1.1.4, apply to learning problems as well. If we regard the goal of learning as determining a mapping on the basis of examples of input-output pairs, there are several basic ways to help this process. One is to apply whatever prior knowledge (approximate form, constraints, etc.) about the learning problem is available in order to chose the most appropriate learning methodology and to properly set up the parameters of the learning system. If the type of function to be learned is known (e.g., a polynomial function, a Boolean function, a multivalued logic function, a complex non-linear continuous transformation, a set of decision rules, a structural description), then the choice of the methodology and the parameters is greatly simplified. Each methodology has its limitations and is most appropriate for a certain class of learning problems. Another way to help the process is to determine the most adequate representation space for learning. Formally, this means mapping the given input domain A into a new domain A' represented in a space that has more relevant dimensions (attributes, relations). Also, some dimensions may be simplified by quantizing them to larger units, which reduces the search space ("dimension abstraction"). Methods of "constructive induction" are specifically oriented toward these issues, and may be useful here. It may also not be necessary to learn a complete and/or very specific form of f. Depending on the learning task, it may be sufficient to determine only a partial or more abstract description of f. Learning a partial description means reducing attention to a subset of A; learning a more abstract description means reformulating A and B into a more abstract form. Finally, not all input-output pairs are (always) possible; hence, the search for f can often be done in smaller spaces, involving only subsets of A and B. Thus, we can reformulate the four recipes for success of Section 1.1.4 as follows:

- **Define your domain:** Learning is particularly useful in problem domains where algorithmic solutions are unavailable or difficult to obtain, but where it is relatively easy to give examples of desirable solutions. Many problems of visual object recognition and navigation fall into this category. For example, it is much easier to point to a desk, and tell the system "this is a desk", than to define the "desk" concept generally and program the definition into the system. Learning problems can vary greatly in difficulty, depending on the complexity of the objects to be learned and the context in which they are learned; e.g., learning the concept of a triangle, presented in isolation to the system, vs. learning the concept of a desk in a cluttered office scene, or the concept of a landmark or obstacle in a navigational situation. Existing learning methods have strong limitations, and it is important to choose a learning method appropriately for the given class of problems. Section 1.2.2. attempted to shed some light on this issue. Available learning methods are particularly useful in domains where relevant attributes are known and can be relatively easily measured, the concepts to be learned can be easily represented in one of the known representational systems (e.g., as a set of attributional decision rules, a low degree polynomial, etc.), and it is not too difficult to obtain sets of reliable training examples.

- **Pick your problem:** Given a problem, the desirable role of learning in it must be defined. If part of the problem has a well-defined algorithmic solution, then programming the solution is usually much easier than applying machine learning. It is necessary to determine which part of the problem needs to be solved by learning and determine the most appropriate learning methodology for it. The learning problem may often be decomposed into simpler problems. The set of tasks that the learning of the mapping will facilitate must be defined. When this has been done, it becomes possible to concentrate on the parts of f that are relevant to these tasks, and to decompose f into simpler mappings, which are easier to learn. A crucial step in applying a learning method is the determination of the appropriate representation space and knowledge representation language. A representation space is defined by descriptors—attributes and relations used for characterizing objects or events. Learning attributional descriptions (which use only variables or attributes, but not relations) is generally easier than learning relational descriptions. If the original problem is stated in terms of low level attributes that may be only indirectly relevant to the problem at hand (e.g., image pixels), new attributes (generally, descriptors) can be sought that are more relevant. More formally, to learn $f: A \to B$, it is easier to learn $f': A' \to B$, where A' is the input domain transformed into a more adequate knowledge representation space. The representation language defines the connectives and operators that can be applied to create a representation of f (mathematical operators, logical operators, etc.). Different problems require different representation languages. Moreover, it may be possible that only subsets of A and B are relevant to the task at hand. Learning the restriction of f to these subsets may be an easier problem.

- **Improve your inputs:** One way to improve the inputs is to represent them in the most relevant representation space (as discussed above). However, even when the representation space is well chosen, it is necessary to provide the learning system with a sufficient number of training examples. It is also important to choose examples carefully, so that they adequately characterize the concepts or transformations to be learned (e.g., "near hits" and "near misses" are often very helpful). This emphasizes the role of the teacher in the process. Finally, the learning process is much easier when the examples are reliable, that is, they do not have classification errors ("classification noise") or measurement errors ("measurement noise"). In some applications it is difficult to avoid such noise. In such situations, one should choose learning methods that are less noise sensitive (e.g., model-driven methods rather than data-driven methods).

- **Take your time**: (See Section 1.1.4).

2. INTEGRATING MACHINE LEARNING AND MACHINE VISION

2.1. Why Vision Systems Need Learning

The incorporation of learning capabilities into vision systems can be motivated in several ways:

- The world often changes unpredictably; therefore it is impossible, in principle, to preprogram all the knowledge necessary for understanding images into vision systems in advance.
- Handcrafting the knowledge needed for analyzing images into vision systems is very complicated and time-consuming; learning provides a major mechanism for simplifying this costly and difficult process.
- In biological vision systems, many aspects of perception are genetically preprogrammed, but many are learned. Similarly, it seems desirable that machine vision systems should be able to acquire some capabilities through learning.

Real-world vision systems must use real data, real sensors/manipulators, and must be judged by real performance metrics. Therefore, such systems must reflect the reality of the external world and the real agent's perceptual and action systems, rather than just the designer's aspirations or hunches about these systems. Reality may include many aspects that cannot be captured by the designer, either because of limitations of the analytic model of the system, or because of errors in sensing due to noise, or because of inadequacies in algorithms. Different algorithms may have differing reliabilities in different contexts; their appropriate attributes should be emphasized according to context during recognition and control. Learning can capture regularities found in real data and biases present in algorithms, and take the relevant aspects into account whether or not these were known to the designer.

As pointed out in Section 1.1.4, vision tasks are generally easier in restricted scene domains; many vision problems have no answers until one poses them in a context, relative to a population of images. But in order to take advantage of this principle, a vision system must know about the restrictions on the domain. It may not be possible to provide all of this information to the system in advance; indeed, the constraints satisfied by real-world classes of scenes are not always easy to formulate. However, the system can learn more about these constraints in the course of examining different scenes. It can then, in principle, modify itself (by adjusting the parameters of the operations it performs, or parameters that control how these operations are applied) to increase its efficiency and improve its performance. It is impractical to manually program a system to properly handle all possible combinations of features; instead, the system can learn how to perform properly from the statistics of the population it encounters, i.e. from the context in which the vision task is embedded. The information that the system acquires by learning about the scene domain can take many different forms, involving a wide variety of properties of the domain (qualitative, numeric, structural, etc.); thus the system can make use of many different learning techniques.

In addition to "long-term" learning about the class of scenes that it can encounter, a vision system can also benefit from "short-term" learning about the specific scene instance with which it is currently dealing. Obviously, this instance is not known in advance (unless the vision task is simply one of verification). When the system begins to analyze images of the scene, the nature of the scene becomes increasingly constrained; as the system learns about these constraints, it can tune or "focus" the operations that it performs on the images. This process can even occur in real time, e.g. if the system is controlling its sensors to acquire new images, or is searching for particular features in an image.

Many researchers feel that machine vision research has gone about as far as possible using domain-independent and context-independent visual features. Although there is certainly room for improvement in algorithms for extraction of specific features such as edges, significant progress now depends on the development of systems that can combine these lower-level features in reliable and parsimonious ways. Unfortunately, constructing combinations of visual modules that solve a given task is generally ad hoc and does not lead to solutions of great generality. The solutions obtained depend on the particular task, user, and context, and are therefore not of great scientific interest.

It seems that learning offers the possibility of obtaining *general solutions* for *task-oriented* vision problems. While useful high-level visual features (e.g., a feature used for recognizing the shoulder of a road) may necessarily be specific to a given task, the methods of learning descriptions in terms of these features can be *task-independent.* Because learning methods can be task-independent, the scientific impact of developing such methods will be far greater than the impact of manually developing descriptions in terms of features specific to any single task.

2.2. Roles of Learning in Vision

Learning is important for vision systems whether we consider them from a task-independent or a purposive viewpoint. In both cases, a rich set of problems emerges (Figure 8).

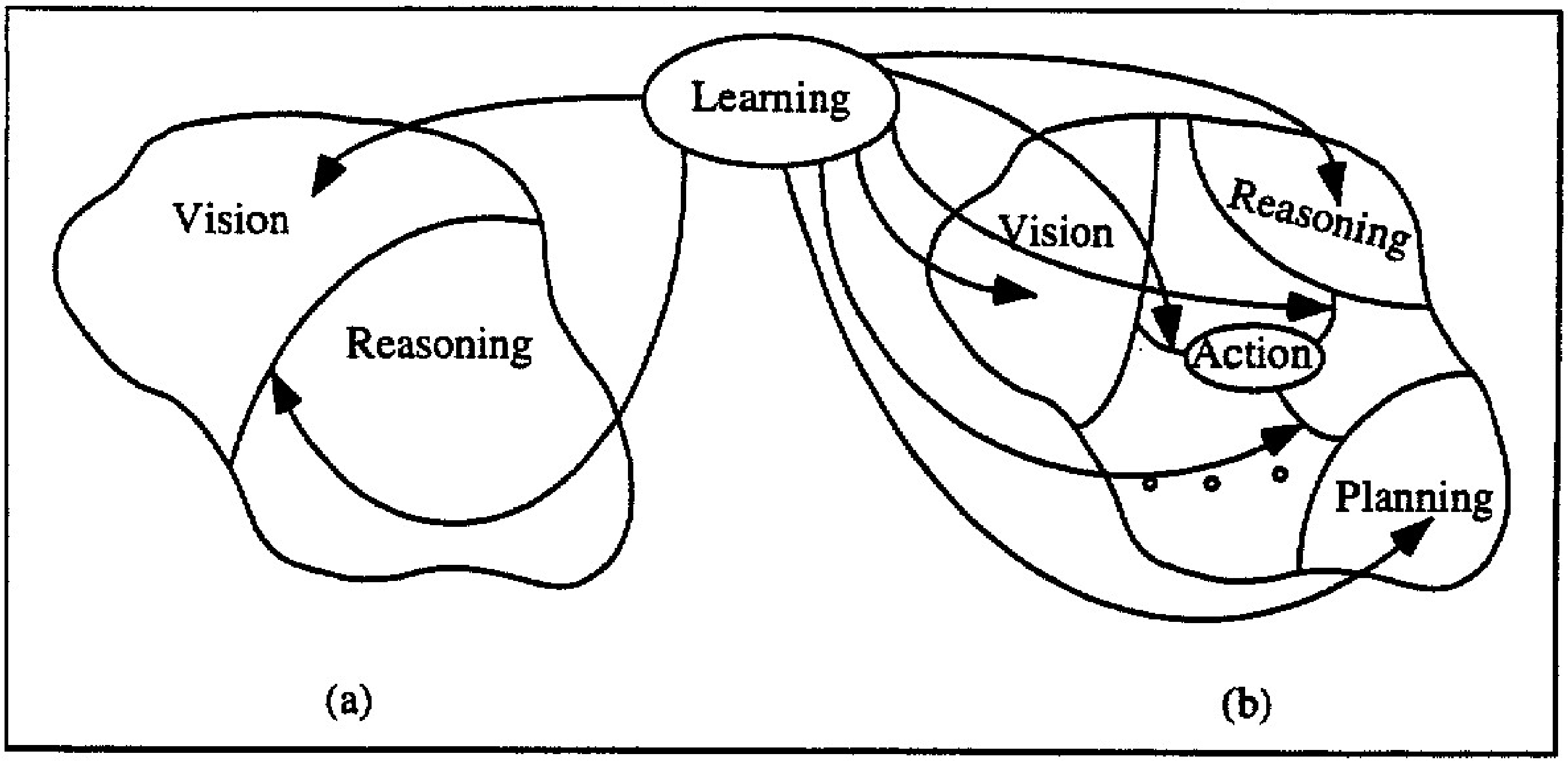

Figure 8: The study of vision and learning from task-independent and purposive viewpoints.

(a) Studying vision in isolation as a process of general-purpose recovery allows learning techniques to be used in two places: the recovery algorithms themselves and the interface between general purpose scene descriptions and other cognitive processes (e.g., reasoning).

(b) Studying vision as a part of a larger system that performs actions allows learning techniques to be used in several places: the algorithms that perform partial recovery, the reasoning and planning processes, and the interfaces between purposive scene descriptions (partial descriptions that make sense in conjunction with a given task (purpose) or set of tasks), reasoning, planning, and action.

In the task-independent approach, a vision system generates general-purpose descriptions of a scene, using various types of representations, and may use these descriptions to recognize objects and events in space-time, by comparing them to models that characterize these objects and events, as described in Figure 9. The system proceeds from images, after appropriate processing, to the development of a general description of the scene. Models of objects or events that exist in the system's memory are appropriately instantiated in the prediction process. The system's goal is to perform matching of predictions to descriptions at one or several levels. The output of the system is given as input to other cognitive processes, such as reasoning. Learning techniques can clearly be used in various parts of the system (see Figure 9). Various aspects of the uses of learning in task-independent vision, as regards the learning of descriptions, representations, and models, will be discussed in Section 2.3.

The purposive approach deals with autonomous systems that are capable of performing various tasks with the aid of visual inputs. These systems must respond "appropriately" to changes in their environment; thus, on a high level, they can be described as evolving or dynamical systems. Such a system can be represented as a function from states and control signals to new states; both the states and the control variables may be functions of time.

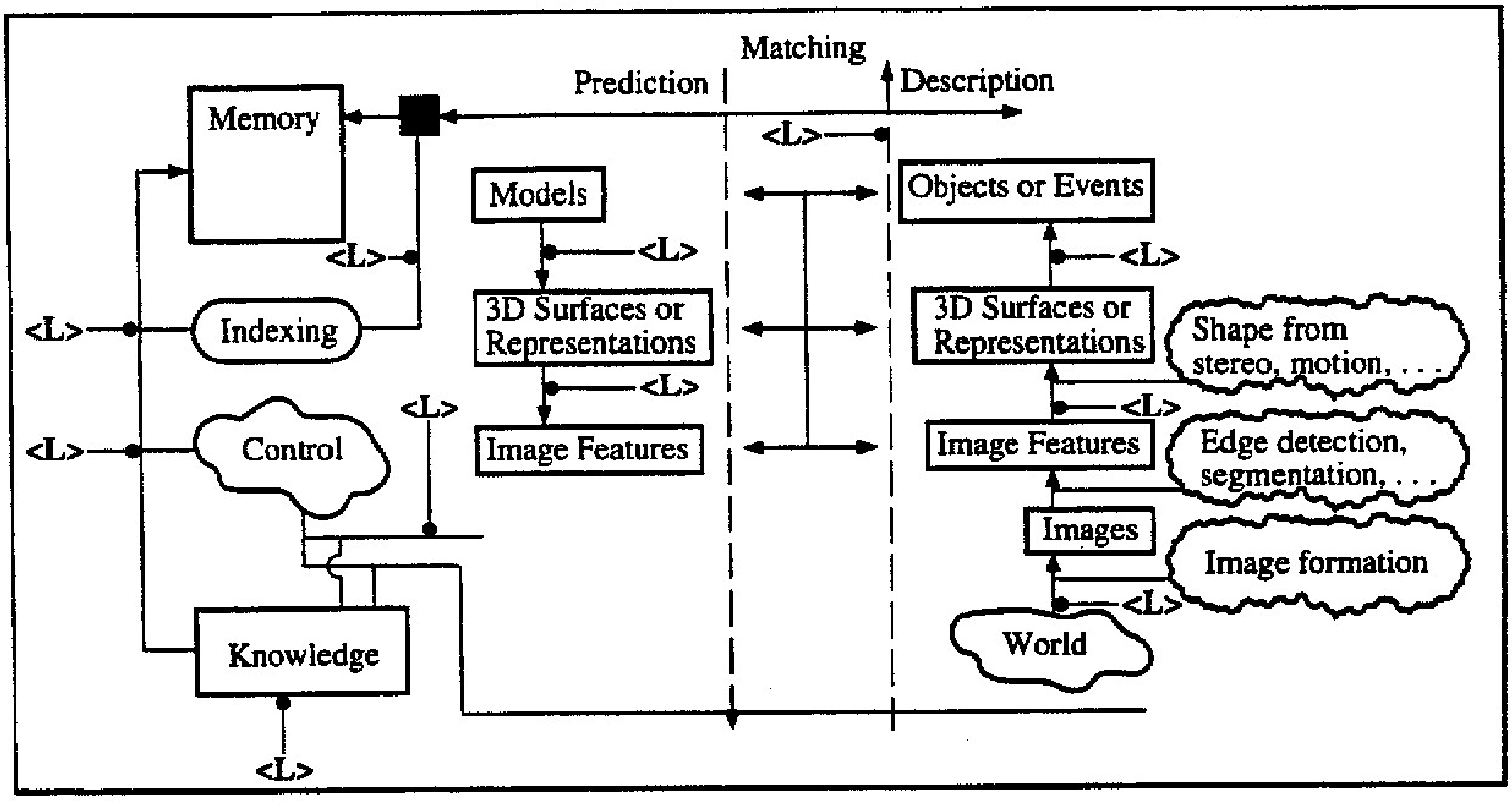

Figure 9: The task-independent approach to studying vision. On the right we proceed to derive from images of the world descriptions of objects or events (bottom-up). On the left we proceed from classes of models to representations of objects or events and their images (top-down, prediction). Vision then entails matching predictions to descriptions at one or several levels. <L>—• indicates stages at which learning techniques could be applied.

Such systems can be described in many ways—e.g., by a set of differential equations (or difference equations), a stochastic process, a finite automaton, or a set of expressions in a suitable logic. Each of these descriptions gives rise to learning problems of different natures. In the discrete case (see Figure 10), the dynamical system takes as input the current state $x(t)$ (e.g., derived from the visual input) and the control signal $u(t)$ and provides as output the next state. In many such systems, the state is not immediately observable, because some filter g masks out or corrupts the actual state. One approach to controlling such a system is to design an observer or state *estimator* e to obtain an estimate of the state x; for example, this estimator might implement a partial visual recovery process. This estimate is used by a controller or state *regulator* r to compute a control signal to drive the dynamical system. Ideally, observation and control are separable in the sense that if we have an optimal controller and an optimal observer then the control system that results from coupling the two is guaranteed to be optimal.[1] Different agents, i.e., dynamical systems, have different capabilities and different amounts of memory, with simple reactive systems at one end of the spectrum and highly sophisticated and flexible systems, making use of scene descriptions and reasoning processes, at the other end.

Learning can be useful in connection with the algorithms of the state estimator and the interfaces between the estimator e and the regulator r. Equally important, learning can take place across systems. When we consider an autonomous "seeing" agent as a dynamical system, we must also consider various integration and control problems and how they relate to learning. For example, at what level do we put the modules of the system? Are behaviors (sequences of per-

[1] In the case of estimation, optimality might correspond to minimizing some measure of the error in estimating the state. In the case of regulation, optimality might correspond to minimizing some measure of the error between the actual state and some target state or state trajectory.

ceptual recognitions and actions) the appropriate level?

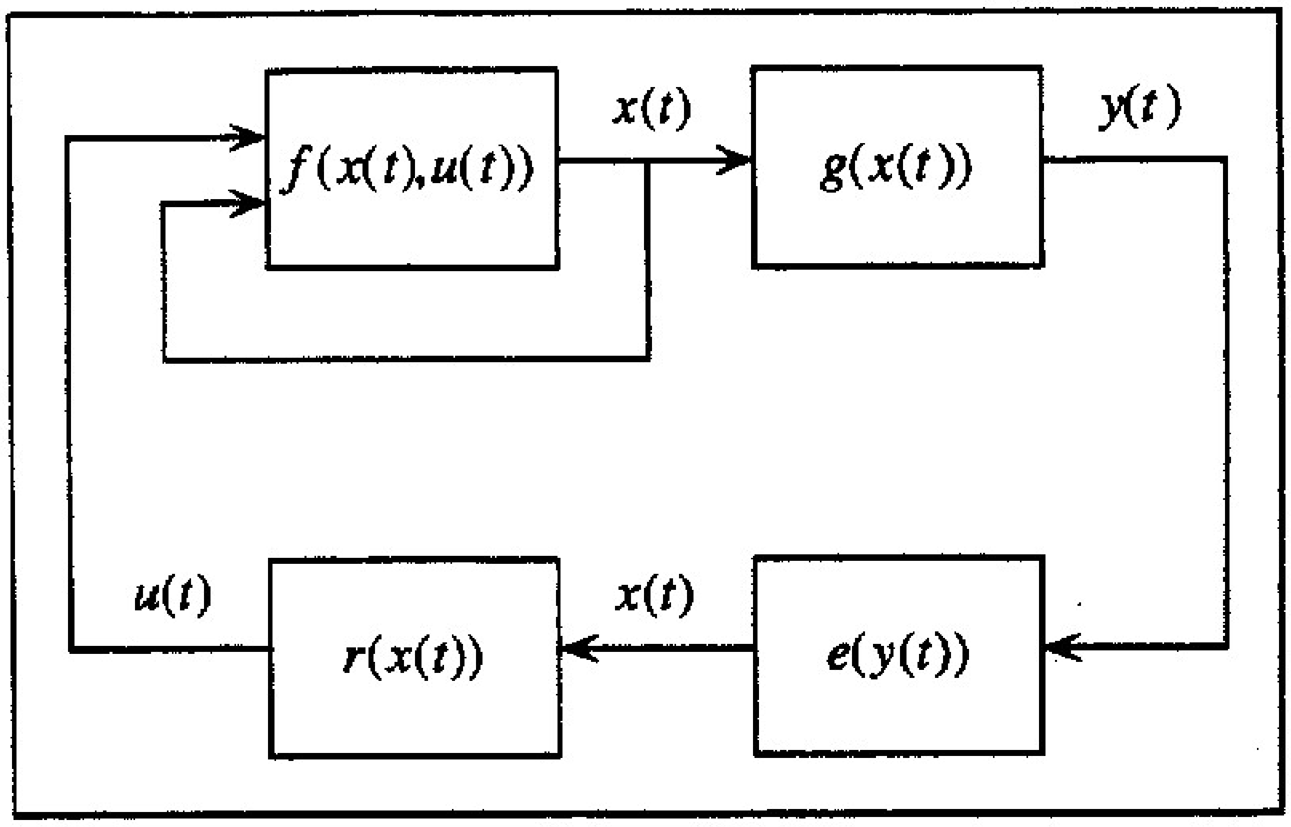

Figure 10: Block diagram of a dynamical system. $f(x(t), u(t))$ represents the system as a function from states x and control signals u to new states.

To theoretically understand how purposive systems might be built, we need to know how to build systems to accomplish simple purposes and how to integrate simple systems to accomplish harder tasks. We need to create a logic and semantics for modality and reasoning by these agents. These problems involve several issues related to learning.

It is useful to analyze certain systems teleologically, and to think teleologically when creating systems. That means it is useful to pretend that robots or animals have purposes or goals, and that they do the best they can to fulfill their goals. Here "best they can" means best given what they know, the knowledge they can acquire, the reasoning abilities they have, the actions they can perform on their environment, etc.

The purposive analysis of robotic or animal behavior is based on such a view. We pretend that agents have beliefs, goals, reasoning capabilities, etc. We pretend that agents can make choices as to what actions they will perform in order to fulfill their goals. It may be that a complex purposive agent can be conveniently viewed as a compound of simpler purposive systems each with its own goals, beliefs, choices, etc. This is another fiction and we are free to divide a complex agent into simpler agents in any way that works. We want to predict the actions of agents, or analyze the actions of agents, or create an agent that can perform certain actions—we accept any purposive analysis that allows us to predict or analyze or create agent behavior in the way we want.

The fact that any actual agent is governed by deterministic or stochastic laws and not by purposive choice is irrelevant; whether complex agents, including humans, actually are conscious and can make conscious choices is irrelevant. We see the purposive language of belief, goal, reasoning, intention, etc., as a convenient specification language that could become a programming language—a kind of nondeterministic logic programming that must ultimately be compiled in machines that are either deterministic or stochastic. To specify an agent is to write a program; learning can be incorporated into the program in many ways.

The description of an agent and a task also determines categories of objects or events that may need to be recognized in order for the agent to carry out the task. It would be an interesting problem to study how these relevant categories could be learned. In addition to the standard kinds of models of objects, defined in terms of geometry and surface characteristics, we might consider functional models, i.e., models that contain information about how an agent can "use" or "interact with" the object under consideration. Learning such models would often be sufficient for object utilization.

We will not attempt in this report to formulate a general theory of purposive vision, since this would require us to construct theoretical frameworks for describing environments, agents, and tasks. Instead, in Section 2.4 we will discuss the rules of vision in an important class of tasks that are common to nearly all purposive agents—namely, the class of navigational tasks.

2.3. Learning in Task-independent Vision

2.3.1. Introduction

From the task-independent standpoint, as illustrated in Figure 9, an intelligent "seeing" system recovers the structure of the scene in its visual field and builds representations that can be compared to class models. The representations may be general descriptions of the scene, or they may involve more specialized descriptions of objects or events. Many important theoretical issues arise in connection with the learning of descriptions, representations, and models, as well as with the complexity of the learning process itself. These issues will be discussed in the following subsections.

2.3.2. Descriptors and Objects

Images contain very large amounts of information. They can be described at different levels of detail and from different points of view. The characteristics of images, and of objects or events that appear in them, include texture, motion, color, occlusion, as well as information obtained from multiple views (such as stereo, for example). Such features, attributes, properties and relations, generally called *descriptors*, define dimensions of the image representation space. A description of an object or event is created by specifying combinations of descriptor values for it. Learning techniques are valuable in this connection, for determining the combinations that characterize a given object. Whatever innate mechanisms are available to existing visual systems to enable them to distinguish between important and unimportant features of the visible world, there is no doubt that descriptions of objects built from these features are learned from examples. For this reason alone the study of visual learning can provide important insights into the structure of intelligent systems.

There are many theoretical problems inherent in recognizing objects visually, even for the restricted case of recognizing rigid objects based primarily on their shapes. First, projection from 3D to 2D means that an object's image can change radically as its pose changes. Second, the interaction of light with the object's surface means that its image can change radically when the distribution of either surface materials or light sources changes. Third, objects are rarely seen in isolation, which means that other objects may occlude them, and that large parts of the image may contain other objects which typically need to be removed in some way for recognition to be possible.

The first two problems suggest the need to find descriptors (input "features")—which we define generally as any measurements or characteristics of an image or its components—that change as little as possible when pose, surface material, or lighting changes. Whether these descriptors should be defined a priori (such as "edges" and filter values), or whether they should, in some sense, be learned, is an important research issue for machine learning.

The third problem suggests the need for some way of dealing with only partial information about an object, and a way of focusing attention on just that part of the image corresponding to a given object. Both of these issues have been extensively studied in the vision community. The issue of partial information appears to be one that standard machine learning algorithms can deal with fairly effectively; many algorithms developed by the vision community, such as Hough transforms and alignment techniques, are designed to deal with this problem. Furthermore, the problem of partial information is not unique to vision, although particular ways in which visual information is incomplete may be specific to vision.

However, the general problems of focusing attention, and of image segmentation, remain largely unsolved. No one has yet found an algorithm that can reliably segment an image into parts that people would generally agree represent coherent parts of the scene. It is very likely that such a generic algorithm does not exist. Specific segmentation techniques have been developed, however, that work well under special conditions. Relative motion, for instance, is a powerful segmentation cue and one that is easy to use. We expect that more reliable segmentation requires the integration of several different cues such as motion, stereo, texture, color, perspective, occlusion, and so on; this provides another important area for learning. Several recent research projects have demonstrated the usefulness of cue integration for image segmentation for the goal of recognition. It should be emphasized that the segmentation process can be iterated several times in a tight loop with a recognition module; we expect, therefore, that perfect segmentation is not necessary and that model-based information is often necessary for segmentation.

2.3.3. Representations

Object representations and models may involve either two-dimensional image descriptions, including texture, contours, 2D shape, image motion, color, etc., or three-dimensional descriptions such as 3D shape, or both. The difficulty of learning how to recognize objects or events depends on the choice of appropriate representations.

Two extreme views about the nature of the representations employed in the recognition of objects have appeared in the literature. One calls for the development of object-centered models of objects and the other for the development of a database consisting of a set of viewer-centered views of the object (2D vs. 3D or higher dimensional models).

Clearly, a very large number of views capturing all possible images of the object for all poses and illuminations would make the problem trivially solvable through a look-up table approach. How many views are really needed at the learning (or model acquisition) phase in order to recognize an object? If the models employed are related to structure (shape), structure-from-motion theorems ensure that under some conditions, such as rigidity, very few views (defined in such a way as to factor out illumination and context) are sufficient to extract full information about the 3D structure of the visible parts of an object. Exploiting the literature on structure from motion under orthographic projection, it is easy to demonstrate that any view of an object is a combination of a small number of views of the same object. These results have been obtained for parallel projections (orthographic and various kinds of paraperspective), but these theorems may generalize to the case of perspective projection under additional assumptions about the environment.

Recent findings on the psychophysics of recognition indicate that object recognition by humans may be accomplished through a relatively simple process involving a comparison of viewer-centered views that, depending on the situation, may or may not contain 3D information. Upon receiving input about an object, the brain compares the novel view to a series of viewer-centered "snapshots" of previously seen objects, stored in memory. Recognition takes place when the brain, through a classification process akin to interpolation, selects the snapshots that most closely resemble the new object.

These conclusions stem from experiments in which subjects were first shown computer-generated images of unfamiliar 3D objects defined as targets. Such targets were either thin wirelike structures or amoeba-like blobs with small projections. The subjects were then presented with single views of either the target, after it had been rotated in various fashions, or a distractor—an object similar to but not identical with the target. The subjects' task was to press, as quickly and as accurately as possible, a yes-button when the displayed object was the same as the target and a no-button otherwise. The study showed that recognition was successful only when the views of the target and the displayed object were sufficiently close for interpolation to take place.

Human performance in recognition turned out to be limited just as it would be if its underlying computational mechanism were memory lookup. The main findings were as follows.

- When subjects had to recognize previously-seen views of objects appearing at arbitrary 3D orientations, some of the views yielded shorter response times and lower error rates than others. This happened even when each view was shown for the same number of times during training.
- Generalization to novel views was severely limited, with performance dropping to chance level at misorientations of only about 40 degrees relative to familiar views.
- Adding binocular disparity to provide the subjects with an additional and reliable cue to the third dimension reduced the mean error rate, but the performance was still far from viewpoint-invariant. Importantly, the availability of depth information did not change the basic feature of generalization to novel views—namely, the increase in the error rate with misorientation relative to a familiar view.

These findings provide evidence against theories which maintain that the input image is compared with a single model of the target, stored in memory in an object-centered fashion and in three-dimensional detail. According to these theories, the brain can rotate this model in any direction and to any degree until it aligns with the new image. The experiments described above showed that recognition of radically different views of the same object is poor, meaning that the brain probably does not have access to rotatable three-dimensional models stored in memory. Thus it is possible that a recognition strategy based on memorizing specific 2D views of objects is a viable alternative to sophisticated techniques employing 3D feature alignment, as well as a better model of human performance in recognition. The findings also suggest that effective object learning can be based on 2D views, rather than on 3D models.

2.3.4. Models

Whether the models employed for scene description are 2D (image-like) or of higher dimensions, they have to be learned from examples. An interesting problem would be to study the complexity of acquiring (learning) object models as a function of the dimensionality of the models employed. This becomes more interesting if we put a bound on the available memory. Complexity aspects of visual learning will be discussed further in the next subsection.

Recognizing three-dimensional objects from arbitrary viewpoints is difficult because an object's appearance may vary considerably depending on its pose relative to the observer. There are two possible ways to approach this problem: (1) Find regularities in the set of views that belong to a single object. These regularities allow the fitting of models to some of the features in the image. (2) Store templates of all possible views of the object and compare them with the actual view. In either case, when building a system for 3D object recognition, it does not seem feasible to create the object library manually. Rather, the ability to learn from examples appears to be essential for the achievement of high performance in real-world recognition tasks.

To learn the shapes of objects for the purposes of recognition using visual input would mean to acquire models for the shapes from examples. Thus, the starting point is the selection of the form of the models. Some types of models are of a simple geometric nature (usually polyhedra, quadrics, superquadrics, etc.). Many systems have been constructed that recognize components of mechanical (or other industrial) parts using small numbers of such models. However, for these types of rigid objects, where an instance is a transformed (and possibly occluded) view of a model (template) plus noise, recognition can be achieved without learning by examining possible transformations of the template until an acceptable match is achieved. It would be interesting to determine whether some form of learning could reduce the search space. Other types of transformable models, based on snakes or deformable contours and their variations, have been successfully used in graphics. Still other types of models, which have not attracted as much attention, are structural descriptions, which characterize objects as systems of parts and relations. Machine learning has developed a number of systems for learning structural descriptions from examples stated in terms of symbolic descriptors. They could be applied to vision problems by coupling them with systems that handle lower level descriptions.

The problem of learning object surface characteristics—specifically, learning classes of textures—has been extensively studied. Many standard approaches can be used to learn such classes from samples of texture feature vectors. Recent research has dealt with the development of efficient methods for the inductive learning of texture descriptions, using multilevel symbolic image transformations. Specific tools used include principal axis representations of texture descriptions, and a combination of inductive rule learning with genetic algorithm based rule enhancement. A problem of particular concern in learning texture classes is the need to learn from noisy inputs under varying perceptual conditions (illumination, resolution, pose, etc.).

Very often a moving object can be recognized on the basis of its motion—that is, the motion field produced by the moving object has a set of characteristics (properties, regularities) that allow us to uniquely identify the object (or its class). For example, we may be able to recognize various types of natural objects (water, leaves, animals, etc.) by learning the appropriate motion descriptions.

2.3.5. Complexity

Any investigation of visual learning at a task-independent level must start with the selection of a class from which the concepts to be learned will be taken—for example, a class of shapes, textures, or motions. Whatever the class, the concepts have to be sufficiently expressive to describe what needs to be learned and to capture any fine distinctions that may be present in images of the same concept. At the same time these concepts, whose descriptions need to be learned from examples and must support generalization to novel situations, have to be kept as simple as possible. One can envision a line of research on the complexity of visual learning problems.

Learnability of visual concepts can be formalized as a problem of learning functions. Let Π be the problem of learning functions that belong to a representation space R, with domain X and range Y. A pair $(x,y) \in S = X \times Y$ is called an *example*, and a sequence of examples is called a *sample*. A function f defined for the set of all samples in R solves the visual learning problem, if it represents a hypothesis that is acceptably close to the "target" function in R. Various theoretical frameworks have been proposed in the literature for handling the general problem using tools taken from approximation theory and probability theory.

When facing a visual learning problem, the first step is to determine an appropriate representation space. The representation space is constrained by the fact that input transducers in both biological and artificial systems provide a signal that is spatially discrete. Existing analyses of visual learning have taken this at face value by assuming that the basic unit of representation is the pixel. Results on the non-learnability of visual concepts have been obtained. Particularly, it has been shown that the number of "templates" needed to achieve learning of Boolean template representations is impractically large.

It is possible, however, that pixel-based definitions of visual learning complexity are unrealistic and other scale-independent complexity measures might be more fruitful. For example, one could develop complexity measures based on features such as those in the primal sketch and show that learning visual concepts is a tractable problem. Determining the complexities of various learning problems, such as learning 3D object recognition algorithms, learning visual search techniques, and relating the complexity of a representation or description to recognition algorithms, are important research problems.

2.4. Learning in Purposive Vision

2.4.1. Introduction

The issues discussed in Section 2.3 were related to general aspects of scene and object description, without reference to the purpose for which the scene is being described. This section discusses vision from a purposive viewpoint. It does not attempt to formulate a general theory of task-oriented vision; instead, it discusses a class of tasks that are basic to nearly all purposive vision systems—namely, the class of vision-based navigational tasks.

One of the most fundamental abilities of an agent is the ability to move around in its environment. An agent that cannot move, or at least move its sensors, is severely limited in its ability to sense (which it must do from a fixed viewpoint); and an agent that cannot move any part of its body cannot exercise motor control for mobility or manipulation. Conversely, an agent that can move its body can use sensory data locally to control its motion, and globally to build up representations of its environment (this includes recognition of objects in the environment, such as obstacles or landmarks).

Navigation deals with both local and global aspects of sensor-based motion control. It is inherently a purposive activity; navigational tasks depend on the nature of the environment, the agent, and the agent's goals. As we shall see in this section, learning can play many roles in the performance of such tasks.

2.4.2. Examples of Learning Problems in Navigation

Many different types of learning problems are encountered in vision-based navigation. In this section we use two representative examples: Learning to drive a vehicle on a road (a local task), and learning paths in an environment (a global task).

In the problem of learning to drive a vehicle on a road, the inputs $x(t)$ (see Section 2.2) might correspond to the images seen by a camera mounted on the vehicle and the outputs $u(t)$ to the control actions taken by an expert driver (see Figure 11a). Using these examples, we might learn a function r corresponding to a state regulator $r(x(t)) = u(t)$, agreeing on the training examples and hopefully generalizing to cover situations not encountered in the training data (see Figure 11b).The problem of learning to drive on a road will be discussed further below and in Section 2.4.3. Some of the roles that learning can play in local navigation tasks will be discussed in Section 2.4.4.1.

In many navigation problems, observation and control cannot be easily separated. In driving a vehicle on a busy highway, it is generally impossible to recover the entire state. Instead, the control system must direct its sensory apparatus so as to estimate different state variables (e.g., the position of the vehicle on the road, the presence of traffic in other lanes) depending on the situation. The system attends to different aspects of the environment depending on the task that it is currently trying to perform. This interdependence of action and perception is critical in most real-world navigation problems.

In the case of learning to drive a vehicle, we are assuming that estimating the state of the dynamical system (or at least the relevant aspects) is relatively straightforward. In many problems this assumption is not warranted and it will be necessary to learn a good state estimator, $e(y(t))$. In still other problems, estimation and regulation must be tightly coupled and the problem of control is considerably more complicated.

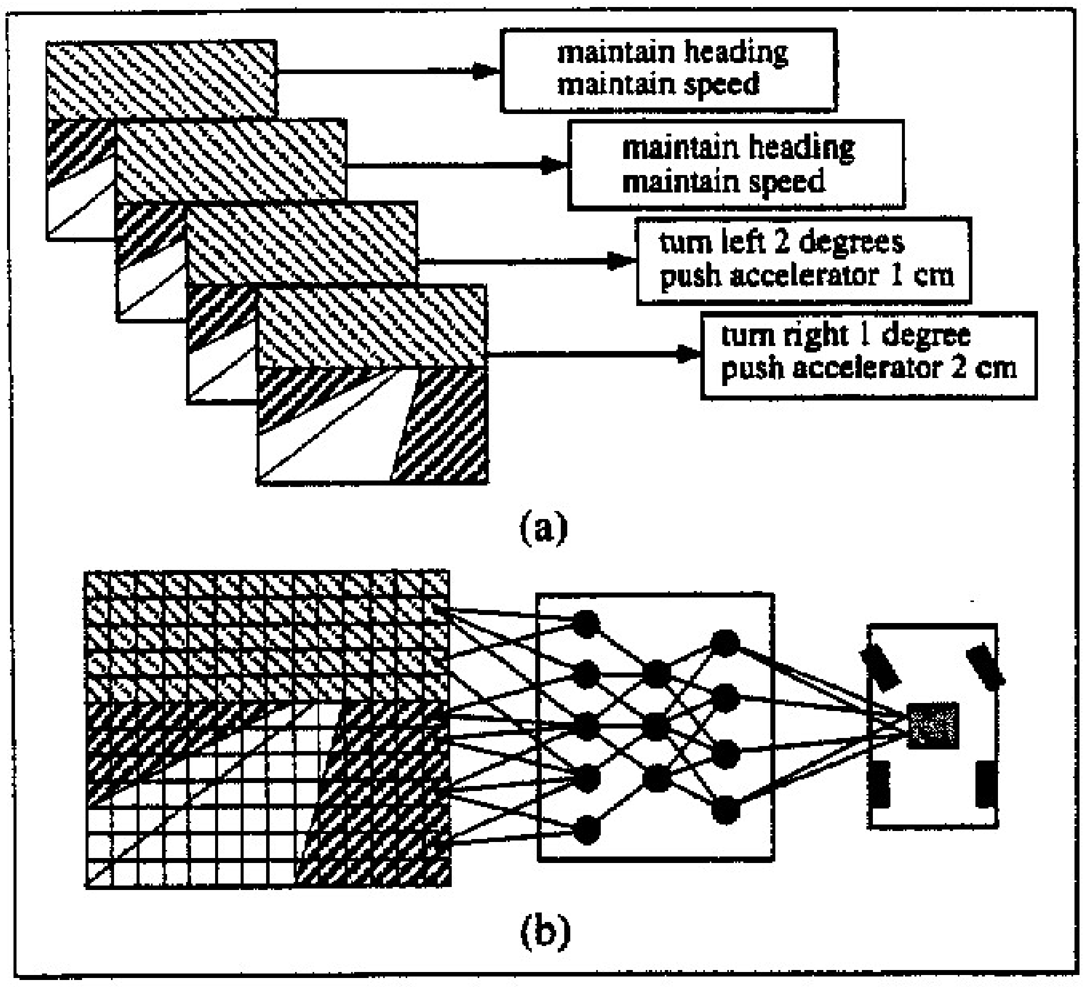

Figure 11. Learning to navigate by observing an expert driver.

There are problems for which it is difficult or impossible to obtain appropriate training data. For these problems it is sometimes possible to obtain a reinforcement signal from the environment (e.g., overloading the drive motors is evidence of negative reinforcement—perhaps the robot is pushing against an immovable object) and then use this signal to learn to improve performance. This is the basic idea behind using learning methods based on stochastic dynamic programming for control problems.

For path planning problems, it is often necessary to predict the consequences of acting. Such prediction may require that we learn the dynamical system, $f(x(t),u(t))$. For some navigation problems, the dynamical system can be described by an annotated map of the environment where the annotations indicate which of a set of navigation routines is most appropriate for getting from one location to another. In most problems of this sort, there is no teacher and so various forms of unsupervised learning must be employed.

Abstractly, the environment can be described as a labeled graph, and the dynamical system can be represented as a finite state machine in which the inputs correspond to navigation procedures and the outputs correspond to the features observable in a given state. Figure 12 shows such a finite state machine in which the states (shown as circles) correspond to locations, the transitions (shown as arrows) correspond to navigation procedures, and the outputs (shown as Boolean values inside the circles) correspond to the features observable when in the location. Possible roles of learning in global navigation tasks, and in the integration of large-scale and local navigation, will be discussed in Section 2.4.4.2.

2.4.3. Classification of Navigation Problems

Now that we have some idea of the sorts of learning problems that arise in navigation, we attempt to classify such problems a bit more systematically. In this section, we provide a set of dimensions to characterize navigation problems that involve learning. We begin with dimensions that serve to characterize the environment and the ability of the vehicle to move about in that environment and observe various features.

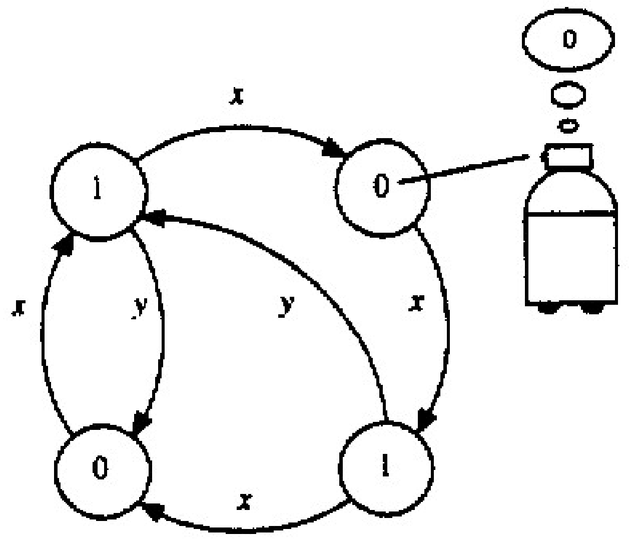

Figure 12. Automaton representation of a dynamical system.

Referring back to Figure 2, in order to classify the capabilities of autonomous "seeing" agents, we should consider both the agent and its environment, in addition to the relationship between the agent and the environment. Different environments differ in fundamental ways; an agent trying to navigate on Mars would almost certainly rely on different navigation methods than if it were operating on an interstate highway. Similarly, the agent's task, e.g., following a road as opposed to getting to some predetermined place, determines what navigation method might be most appropriate. We characterize a given environment as a point in a three dimensional space.

a) CONSTRAINED/UNCONSTRAINED: How constrained are the agent's actions? In some environments, the agent has many possible actions available at any given point, while in other environments only a few options are available. For example, on Mars (a prototypical unconstrained environment) there are essentially an infinite number of actions (e.g., directions) that can be taken at any given time. In contrast, navigating a series of hallways constitutes a constrained environment.

b) STATIC/DYNAMIC: How dynamic is the environment? A dynamic environment is one in which changes occur that are not caused by the agent. For example, shifting illumination patterns and weather conditions make for a dynamic environment. An interstate highway is a dynamic environment to the extent that other cars are present.[2] On the other hand, agents operating alone and indoors must deal with a much more static environment.

c) STRUCTURED/UNSTRUCTURED: The degree to which an environment is structured depends on the distribution, accessibility, and uncertainty implicit in the observable features. Thus a highly structured environment is one in which the observable features uniquely identify the current situation with a high degree of certainty. In contrast, an unstructured environment might be feature-poor, or there might be so many features that a sort of saturation is achieved, where distinct situations are no longer separable.[3]

In Section 2.4.2, we described navigation in terms of a dynamical system corresponding to the vehicle and the environment in which the vehicle is trying to negotiate. In characterizing

[2] The presence of multiple agents generally implies a dynamic environment. We will not, however, emphasize coordination and negotiation involving multiple agents in this document.

[3] While we would like to pose this dimension in a task-independent manner, we acknowledge that there is an aspect of task dependence about it.

navigation problems, it is not clear whether the vehicle and its complement of sensors are part of the problem or part of the solution. While we might be given a vehicle and asked to write software to control it, we might also be asked to design or refine the hardware. In the latter case, the vehicle is definitely part of the problem specification. We make no attempt to characterize all aspects of the vehicle but we do mention one important dimension that affects perception and, hence, impacts on the problems discussed in this section.

d) SOPHISTICATION OF OBSERVATION: How sophisticated are the agent's observational capabilities? There are several aspects: (a) ability to recognize objects in the environment, (b) ability to re-acquire a formerly recognized object and relate it to previous recognition instances, (c) ability to quantify spatial relationships between multiple objects observed simultaneously, (d) ability to recognize motions of other objects. Note that there are a wide range of possibilities in (a–d). For example, reasonable navigation in constrained environments can be done when, for (a), objects are only avoided, not distinguished. Wasps, on the other hand, are very good at both (a) and (b) to the level of recognizing and re-acquiring individual bushes and buildings.

The second defining aspect of a navigation problem is the task that the agent is to accomplish. As we have done for navigation environments, we next characterize navigation tasks along four dimensions.

e) SHALLOW/DEEP INFERENCE: How much inference is required in order to determine what to do next? For certain navigation tasks, fast and cheap reactive behavior may suffice. Consider, for example, a convoy-following problem; gradient descent is an example of a fast and cheap inference technique that, given the appropriate sensory input space, may be expected to attain reasonable performance for this particular task. Thus this is an example of a task that requires only shallow inference. On the other hand, finding the shortest path through some graph is an example of a task that requires some deliberation (e.g., search) and would therefore be considered a deep inference task. A key component of this distinction is whether or not an internal representation is used.

f) RESOURCE CONSIDERATIONS: How much time is available to decide and execute? Real-world navigation tasks are by nature resource-critical. Resource limits may be imposed on different aspects of the task. For example, it might be necessary to limit computation time in order to ensure real-time response. For a different task, obtaining a more efficient solution (in terms of execution-time resources) might be important. In a hostile dynamic environment, doing something quickly may be much more effective than doing the right thing slowly.

g) AVAILABILITY OF SUPPLEMENTARY KNOWLEDGE: How much additional knowledge is available to the system? Types of additional knowledge include maps, guidance from teachers, and a priori specification of landmarks. For example, easy availability of satellite position data clearly affects which solutions are viable. Note that additional knowledge may differ qualitatively from task to task; some feedback may be immediate, (e.g., when the agent hits a wall), while other sources of supplementary knowledge may not be quite so direct.

h) COMPLEXITY OF BEHAVIOR: Does behavior require long or short sequences of primitive actions? Some navigation tasks decompose in such a way as to reduce the overall complexity of the task. Thus some large-scale deep-inference navigation problems might naturally decompose into multiple shallow-inference problems. Other problems may not be decomposable at all, or no reasonable decomposition may be known.

To illustrate these environment/task characterizations, let us consider ALVINN (Pomerleau, 1989), an existing simple visual navigation system developed at Carnegie-Mellon University, to see how it fits into our classification. ALVINN is a neural-network based system that learns to perform a road following task by observing a human driver. The input to an ALVINN network is a low-resolution sensor image; the output is a continuous variable representing steering direction. ALVINN networks learn to steer under specific conditions, such as highway driving, off-road dirt path following, etc. Each individual network, then, could be considered an expert for a particular restricted visual navigation task.

ALVINN's environment is constrained, since the agent is limited to following the road. Choices occur occasionally (e.g., at intersections) and are of limited degree (that's not to say the problem is trivial, since there are infinitely many steering sequences that will keep the vehicle on the road). The environment is dynamic, since each ALVINN net is expected to operate under varying illumination, weather, road and traffic conditions. These aspects of the environment are clearly beyond the agent's control. Finally, ALVINN's environment is a structured one, since relevant features (e.g., lane markings) are always available on the input.

Clearly the visual road-following task, in its simplest form, is a shallow inference task. In fact, each individual neural network serves to directly map inputs into steering responses. Note that it is possible to integrate the different ALVINN networks into a single system by using some meta-level reasoning procedure to arbitrate among experts. The particular arbitration procedure used involves to some degree of inference; thus this version of ALVINN is necessarily deeper than the single network version.

From a resource perspective, ALVINN's task is time critical. ALVINN's design (i.e., the neural network architecture used) reflects this constraint: it guarantees some output within a fixed time frame.

Supplementary knowledge is provided to ALVINN in two forms. First, every ALVINN network is structured according to some fixed architecture that embodies some domain bias. Second, to avoid brittleness, ALVINN's training procedure is constrained according to domain-specific knowledge. For example, each ALVINN network is trained on an equal number of left and right turns. ALVINN is explicitly trained by adding structured noise to the input image in order to ensure insensitivity to infrequent events, such as passing cars or guardrails.

2.4.4. Some Navigational Issues

In this section, we consider two sets of issues. The first set is concerned primarily with acting and sensing in a constrained ("local") spatiotemporal context; an example problem would be steering a vehicle to remain on a roadway and avoid other vehicles. The second set is concerned with building representations to facilitate planning in a larger ("global") spatiotemporal context; an example problem would be building a map to facilitate path planning. This distinction is not always clear-cut, but it provides some useful structure for the following discussion.

2.4.4.1. Local Issues

We define local navigation issues as those pertaining to sensing and acting within a constrained spatiotemporal context. The bounds of the context are determined relative to the physical and information processing capabilities of the navigating system. Local issues are those that take place over a small enough spatiotemporal interval not to need significant cognitive, symbol processing activities, but rather those that produce control and other outputs more or less directly from environmental and other inputs. Local navigation may be reactive (closed loop) or blindly executed (open-loop)—some functions operate as reflexes, some as skilled behaviors.

We assume that local navigation takes place in the context of global navigation. That is, local navigation tasks are always performed in a global context that provides information, representations, particular requests, etc.

The inputs to a local navigation task include visual and nonvisual sensory input (e.g. inertial sensors, dead reckoning or odometry, proximity sensors, range, velocity). Other important inputs come from the global navigational context: visual and control tasks to be accomplished (landmarks to be located or verified, lane change warnings, upcoming turnoffs or intersections, and the like).

Outputs from local navigational tasks include the obvious control outputs (commands to accelerators, steering, brakes) as well as outputs to control active visual capabilities (non-navigational control), and many varieties of information, either solicited or unsolicited, relevant to the global navigation task—for instance, reconstruction of physical reality (time to collision, free space, location of looming scene areas), landmark identification, and the degree of confidence the local task system has in its decisions and reports. In what follows, we provide a catalog of local navigation issues.

a) BEHAVIORS AND TASKS: Given a set of local navigational behaviors, how can we learn what tasks they can be used to perform? Developing a flexible, uniform scheme for describing what behaviors do may be impracticable; some form of category or concept learning might be useful, as well as more quantitative forms of characterizing performance.

b) COMBINING BEHAVIORS: How can we learn to combine behaviors? There is essentially no theory about combining behaviors (aside from some work on discrete event dynamical systems, DEDS), and the obvious questions arise as to the stability, deadlock properties, controllability, predictability, etc. of such interacting systems. It seems that some innate rules or structure might be necessary to assure reasonable behavior, but that within some limits there might be the possibility of learning about interactions with an eye either to avoiding or exploiting the results. Can the coordination take place without sensor input; does the central coordinator perform strictly as a function of the modules or can it use input from the environment in making its choices? Is there in fact a central controller or can control be distributed, as in flocking or schooling behavior (cellular automata models, etc.)? Can reinforcement learning be extended to deal with multiple goals to be satisfied simultaneously?

c) USING SENSORY DATA: How can we learn to use and combine sensory information? Such information can be passive or active, sparse or dense, visual or tactile; clever combination of sensory modalities can lead to fast, dense depth data. Aspects of the algorithm that combines the data might vary (thresholds, calibration), but effective methods of combination could be learned. A related issue is learning to combine sensing with predicting. The global behavior may furnish many sorts of predictions, from symbolic to iconic, about what is likely to be sensed by the local task. The incoming data must be matched to these expectations. When can sensory data be disregarded?

d) USING MEMORY: How can we learn memory management? What should be remembered and what can be forgotten? What are the roles of short-term and long-term memory? Of iconic and symbolic memory? How do we sense events, and reason about events in time?

e) CALIBRATION: How can we learn calibration constants for sensors and effectors? The calibration problem is a major curse of robotics and quantitative vision. Especially in systems that are active, subjected to shock, and by their nature operating in rapidly changing environments, calibration by the usual method of physical measurement or rituals involving precise calibration objects are simply impractical. The hope is that reliable feedback and closing the loop through tasks and attempts at tasks can substitute for off-line calibration procedures.

f) CONTROL: How can we learn to select and adapt control functions? In particular, how can we learn optimal feedback control? Abstractly, reinforcement learning yields optimal control; practically, there may be issues involving the dimensionality of the state space and likewise the length of description of control actions. Learning open-loop control is like learning behaviors, only the task to be learned may be parameterized. Thus optimal control for a class of tasks must be learned, which amounts to learning a family of control trajectories for multiple cooperating effectors. The goal here is to produce behaviors that are faster than those that can be mediated through feedback. An issue is that the controlled plant may interact with a plant whose properties are unknown and must be learned. We need to be able to learn task-level control algorithms in which a parameterized version of a task is given (say following some trajectory in the phase space of position, velocity, and load) and the output is a trajectory in control phase space. Assumptions about the controlled plant may be available but the problem could also include learning about the plant to be controlled and also learning information about the task that is exogenous to the plant (e.g. road conditions) that nevertheless affect performance. In all but trivial cases, the dynamical model will be stochastic.

2.4.4.2. Global Issues

Large-scale or "global" navigation is the problem of traversing a large-scale environment of which only a portion will be within the field of view at any given time. Such a task will therefore involve reference to some type of internal representation of the large-scale space, even if that representation is simply in the form of (re)actions of the agent to situations encountered while in pursuit of specific goals. There are several areas here that provide opportunities for learning.

a) LEARNING REPRESENTATIONS OF LARGE-SCALE SPACE:

Representations of space for the purpose of navigation should allow the agent to make decisions about what to do next; they should be able to handle different types of environments and to support exploration of space at multiple spatial scales. To be useful, representations should establish correspondences between actions and positions, to allow planning and simulation. They should suppress information about the environment that is not relevant to potential navigation goals, but maintain relevant information.

Representations of space may be continuous or discrete, e.g., a Cartesian map (*x-y* coordinates) or a finite automaton, and this choice influences the relevant issues and techniques with respect to learning and sensing. Relevant information may vary significantly in level of abstraction: e.g., local coordinate systems with precise locations of landmarks versus spatial equivalence classes of regions on the basis of landmark visibility. The representation may be purely symbolic and topological, or may include metric information as well. Multiple representations may be needed for different tasks.

Learning spatial representations may take place either by exploration or it may occur during goal-directed activity. This involves the tradeoff between the goal of acquiring new information and the goal of using current information, known in control engineering as the tradeoff between identification and control.

In discrete representations that use spatial equivalence classes according to visible landmarks, there is also a tradeoff between the simplicity of the representation and the ability to discriminate between similar positions. As the number of landmarks increases, the size of each equivalence class decreases, but the number of classes increases. This has implications for the complexity of learning and planning methods that use the resulting representation.

b) LANDMARK ACQUISITION AND RECOGNITION:

Agents position themselves within the spatial map by recognizing landmarks. To the extent that landmarks can be arbitrary objects, landmark recognition is an instance of the general object recognition problem. Landmark recognition is distinct from traditional object recognition in several respects, however:

- Indexing—the spatial map provides expectations of the landmarks to be recognized, simplifying the object indexing problem.
- Context—whereas general object recognition systems often strive to be independent of context, a landmark's context is an integral part of its model.
- Performance criteria—the goal of landmark recognition is to fix the viewer's location with respect to a spatial map, whereas the goal of object recognition is rarely explicit.

Another promising area for learning is landmark acquisition. Autonomous navigation systems must learn models of the landmarks in their environments, both for individual landmarks such as buildings and mountain peaks and for "class" landmarks, such as road intersections and stop signs, that may occur repeatedly. Each landmark should be modeled so as to support recognition, and should include (quantitatively or qualitatively) the landmark's position in the spatial map. Landmarks should be selected according to their distinctiveness, salience, and utility for localization. Landmarks should also be visible from a wide range of viewpoints so that they can be recognized when the location of the viewer is only approximately known.

c) INTEGRATION OF LARGE-SCALE AND LOCAL NAVIGATION:

Typically, a large-scale navigation system identifies goals for local navigation, and is alerted when these goals are met. An example is following a road until a particular landmark is reached. In general, the goals for local navigation will depend on the extent to which the environment constrains the mobility of the agent (e.g., an environment with well-defined roads or corridors is a highly constrained environment). The detection of the "termination" conditions for a particular segment of local navigation requires monitoring the environment. Since the appearance of these conditions is usually predictable from prior visual events ("context"), this process can benefit from active visual investigation of the environment instead of passive detection of the conditions when they appear.

The integration of large-scale and local navigation problems is a complex process which is akin to real-time operating system design. It affords a number of opportunities for learning. For example, while traversing familiar paths, the decomposition of large-scale navigation into a sequence of local navigation steps can be learned. It may also be possible to generalize this process and learn a common set of strategies for similar environments (e.g., traversing road networks may involve a common core set of strategies that are applicable over a wide range of roads).

Since the local navigation process and the monitoring required for global navigation are both based on visual information, the same processing resources may be expected to be required by both processes. It may be possible to learn the active vision strategies needed for handling resource contentions—e.g., learn a familiar sequence of moves to monitor the road while periodically investigating landmarks that appear and move toward the periphery of the field of view. The usefulness of such a strategy depends on the types of visual processing architectures available, and may also have to be adapted to handle variations in the quality and characteristics of the resources over time (e.g., sensor drifts, resource failures, etc.).

3. RECOMMENDATIONS

The state of the art in both machine learning and machine vision is such that combining the two fields promises to produce important scientific and practical results. This claim is supported by recent progress in machine vision on 3D object recognition, successful application of machine learning to a wide range of real-world problems, and concrete suggestions about feasible approaches derived (or soon expected to be derived) from biological and neurophysiological experiments.

Further advances in learning technology and tools are needed for vision applications, in order to handle the complexity and noisiness of vision data and to deal with different task requirements. Major efforts should be applied to task-driven learning, automatic selection or construction of key features, noise-tolerant learning, making effective use of prior, domain-specific knowledge, learning multiple representations, multistrategy learning, and learning under variable perceptual conditions. New tools must be created that will integrate the developed techniques. Finally, the learning systems often will have to deal with environments where object characteristics can change over time.

At the same time, new vision techniques should be developed to extract, manipulate, and combine hybrid characteristics of objects on different levels of the vision hierarchy. Fast techniques for matching image data and class descriptions (i.e., recognition modules) must be developed. High-level scene segmentation and annotation through automated reasoning must be advanced, utilizing learned concept descriptions, new matching/recognition techniques, and concept manipulation techniques.

Continued connection with biology is important for the scientific content of the enterprise. Increased knowledge about how the brain recognizes objects, and the role learning plays in this task, will also provide useful guidelines to those designing practical systems.

Learning to perform vision-based tasks is an extremely rich problem domain, which, most importantly, has many intermediate goals which will yield tangible benefits to the research and applications communities in the long and short run. Some of the short-term research opportunities relating to task-oriented vision are listed below.

1. Learning cost-efficient visual search and surveillance strategies by selection of sensors and algorithms which are most discriminating in a given context. In other words, learning what to sense and where to focus attention and resources at different stages of a task.
2. Combining multiple visual modules for task-oriented vision using learning approaches.
3. Learning to observe by segmenting actions temporally and learning appropriate perceptual actions to maximize the observability of processes by active vision systems.
4. Developing selective forgetting strategies which allow learners to track changes in the environment, especially to allow system performance to re-adapt when system sensors and actuators fail.
5. Learning contexts so that the environment and task can be partitioned into special cases that require different perceptual and action control strategies.
6. Developing effective general-purpose methods for combining teaching, exploration, and exploitation in learning.
7. Using learning algorithms for redundant actuators and sensors to achieve improved robot

path planning in obstacle-filled environments.

8. Task planning using learning by observation for task decomposition, and specialization using reinforcement learning for each sub-action.
9. Developing learning techniques that do not require explicit feature tracking, because they combine the extraction of relevant features with simultaneous learning of action models.

We believe that the following issues are important considerations in planning future efforts in learning and vision:

a) **Machine learning can play an important role in vision**. Machine learning appears to offer significant opportunities to extend current methods for vision, especially in the area of task-oriented vision. We have described a number of initial results, and a variety of suggested roles for machine learning in visuomotor and visual tasks.

b) **Reasonable expectations**: We should certainly not assume that by introducing learning into visual tasks one will solve all vision problems. Many of the short-term goals enumerated above are restricted problems; however, solutions to any of these problems will have important research implications and practical uses. Visual processing for almost any task is extremely complex, which is the reason why progress in this area has been slow. This complexity arises from the multivariate problem of varying illumination, the observer's optics, and the complexity of the environment. Furthermore, the data are spatially and temporally distributed; hence the necessary data selection and reduction mechanisms are very task- and context-dependent. We must therefore continue to study basic analysis of data reduction and selection mechanisms. We must study invariances and spaces which enhance these invariances.

c) **Collaboration between the machine learning and the vision/robotics communities**: Many members of the machine learning community do not have access to the laboratory facilities needed to pursue learning for visuomotor coordination. In particular, the equipment costs and staff expertise necessary for designing and maintaining robotic and vision hardware is out of the reach of most machine learning groups. To facilitate interchanges between these two communities, cross-disciplinary postdoctoral fellowships should be established to allow machine learning researchers to make extended visits to robotics and vision facilities.

d) **Infrastructure development**, in the form of shared testbeds, is strongly recommended. Shared testbeds provide a means of measuring and demonstrating progress; they have proven useful in other fields and are likely to greatly benefit this one.

e) **Competition**: A robotic competition should be organized that encourages graduate students to explore research in combining machine learning, perception and robotics. This could be held in conjunction with some major conference in either robotics or artificial intelligence, such as IJCAI, AAAI or IEEE Robotics and Automation.

f) **Goals**: Selected projects should be established as long-term goals. These projects should focus on important potential applications. Applying machine learning and avoiding (re)programming as much as possible is critical for future practical applications of machine vision.

ACKNOWLEDGMENTS

This report is based on the discussions at the NSF/ARPA Workshop on Machine Learning and Vision and on a preliminary report written by the participants. The editors express special thanks Dr. Oscar Firschein and Dr. Howard Moraff for providing support for the Workshop, for their belief in the importance of research in machine learning and vision, and for their continuing interest in and support of this research.

Thanks go also to the members of the GMU Machine Learning and Vision group, Jerzy Bala, Ali Hadjarian, Mark Maloof and Peter Pachowicz, for their comments on an earlier draft of this report.

REFERENCES

Aloimonos, Y. and Shulman, D., *Integration of Visual Modules: An Extension of the Marr Paradigm*, Academic Press, Boston, MA, 1989.

Ballard, D.H. and Brown, C.M., *Computer Vision*, Prentice Hall, Englewood Cliffs, NJ, 1982.

Bloedorn, E. and Michalski, R.S., "Data driven constructive induction in AQ17-DCI: A method and experiments," *Proceedings of the Third International Conference on Tools for AI*, San Jose, CA, November 9–14, 1991.

Bloedorn, E., Wnek, J. and Michalski, R.S., "Multistrategy constructive induction," *Proceedings of the Second International Workshop on Multistrategy Learning* (MSL93), Harpers Ferry, WV, May 27–29, 1993.

Cohen, W.W. and Hirsh, H. (eds.), *Machine Learning*, *Proceedings of the Eleventh International Conference*, Rutgers University, New Brunswick, NJ, July 10–13, 1994.

Carbonell, J. (ed.), *Machine Learning: Paradigms and Methods*, MIT Press, Cambridge, MA, 1990.

Fawcett, T. (ed.), *Working Papers from the Workshop on Constructive Induction and Change of Representation,* organized in connection with the Eleventh International Conference on Machine Learning, Rutgers University, New Brunswick, NJ, July 10–13, 1994.

Kodratoff, Y. and Michalski, R.S., *Machine Learning: An Artificial Intelligence Approach,* Vol. III, Morgan Kaufmann, San Mateo, CA, 1990.

Landelius, T., "Behavior representation by growing a learning tree," Thesis No. 397, Dept. of Electrical Engineering, Linköping University, Sweden, 1993.

Marr, D., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, CA, 1982.

Michalski, R.S., "A theory and methodology of inductive learning," *Artificial Intelligence,* Vol. 23, pp. 111–161, 1983.

Michalski, R.S., "Inferential theory of learning: Developing foundations for multistrategy learning," in *Machine Learning: A Multistrategy Approach*, Vol. IV, Michalski, R.S. and Tecuci, G. (eds.), Morgan Kaufmann, San Mateo, CA, 1994.

Michalski, R.S., Carbonell, J. and Mitchell, T. (eds.), *Machine Learning: An Artificial Intelligence Approach,* Morgan Kaufmann, Los Altos, CA, 1983.

Michalski, R.S., Carbonell, J. and Mitchell, T. (eds.), *Machine Learning: An Artificial Intelligence Approach*, Vol. II, Morgan Kaufmann, Los Altos, CA, 1986.

Michalski, R.S., Rosenfeld, A., Pachowicz, P., and Aloimonos, Y., "Machine learning and vision: Research issues and promising directions," Preliminary Report on the NSF/ARPA Workshop on Machine Learning and Vision, Technical Report No. MLI-93-1, George Mason University, Fairfax, VA, 1993.

Michalski, R.S. and Tecuci, G. (eds.), *Machine Learning: A Multistrategy Approach,* Vol. IV, Morgan Kaufmann, San Mateo, CA, 1994.

Pomerleau, D., "ALVINN: An autonomous land vehicle in a neural network," Technical Report No. CMU-CS-89-107, Computer Science Dept., Carnegie-Mellon University, Pittsburgh, PA, 1989.

Rosenfeld, A. and Kak, A.C., *Digital Picture Processing*, Academic Press, New York, second edition, 1982.

Warmuth, M. (ed.), *Proceedings of the Seventh Annual ACM Conference on Computational Learning Theory*, Rutgers University, New Brunswick, NJ, July 12–15, 1994.

Wittgenstein, L., *Philosophical Investigations*, Blackwell, London, 1958.

Wnek, J. and Michalski, R.S., "Comparing symbolic and subsymbolic learning: A case study," in *Machine Learning: A Multistrategy Approach,* Vol. IV, Michalski, R.S. and Tecuci, G. (eds.), Morgan Kaufmann, San Mateo, CA, 1994a.

Wnek, J. and Michalski, R.S., "Hypothesis-driven constructive induction in AQ17-HCI: A method and experiments," *Machine Learning,* Vol. 14, pp. 139–168, 1994b.

SUPPLEMENTAL BIBLIOGRAPHY

Agin, G.J. and Duda, R.O., "SRI vision research for advanced automation", *Proc. of the Second USA-Japan Computer Conf.*, Tokyo, Japan, pp. 113–117, 1975.

Albus, J.S., "A new approach to manipulator control: The cerebellar model articulation controller", *Trans. of the ASME: Journal of Dynamic Systems, Measurement and Control*, Vol. 97, pp. 220–227, 1972.

Albus, J.S., "Outline for a theory of intelligence", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 21, pp. 473–509, 1991.

Aloimonos, J. and Shulman, D., "Learning early-vision computations", *J. of the Optical Society of America*, Vol. A6, pp. 908–919, 1987.

Atick, J.J., "Could information theory provide an ecological theory of sensory processing?", *Network*, Vol. 3, pp. 213–251, 1992.

Atick, J.J. and Redlich, A.N., "Quantitative tests of a theory of retinal processing: Contrast sensitivity curves", Report 90/51, Institute for Advanced Study, 1990.

Atkeson, C.G., "Using locally weighted regression for robot learning", *Proc. of the IEEE Conf. on Robotics and Automation*, Sacramento, CA, pp. 958–963, 1991.

Ayache, N. and Faugeras, O.D., "Hyper: A new approach for the recognition and positioning of two-dimensional objects", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 44–54, 1986.

Bajcsy, R., Paul, R., Yun, X. and Kumar, V., "A multiagent system for intelligent material handling", *Proc. of the Intl. Conf. on Advanced Robotics*, Pisa, Italy, pp. 18–23, 1991.

Bala, J., "Combining structural and statistical features in a machine learning technique for texture classification", *Proc. of the Intl. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Charleston, SC, July 1990.

Bala, J. and DeJong, K., "Generation of feature detectors for texture discrimination by genetic search", *Proc. of the Intl. Conf. on Tools for Artificial Intelligence*, Washington, DC, 1990.

Bala, J., DeJong, K. and Pachowicz, P.W., "Using genetic algorithms to improve the performance of classification rules produced by symbolic inductive method", *Proc. of the Intl. Symp. on Methodologies for Intelligent Systems*, Charlotte, NC, pp. 286–295, 1991.

Bala, J., DeJong, K. and Pachowicz, P.W., "Integration of inductive learning and genetic algorithms to learn optimal descriptions from engineering data", *Intl. Workshop on Multistrategy Learning*, Harpers Ferry, WV, 1991.

Bala, J., DeJong, K. and Pachowicz, P.W., "Multistrategy learning from engineering data by integrating inductive generalization and genetic algorithms", *Machine Learning: A Multistrategy Approach IV*, R.S. Michalski and G. Tecuci (eds.), Morgan Kaufmann, San Mateo, CA, 1993.

Bala, J. and Michalski, R.S., "Recognition of textural concepts through multilevel symbolic transformations", *Proc. of the Intl. Conf. on Tools for Artificial Intelligence*, San Jose, CA, 1991.

Bala, J., Michalski, R.S. and Wnek, J., "The principal axes method for constructive induction", *Proc. of the Intl. Conf. on Machine Learning*, Aberdeen, Scotland, 1992.

Bala, J. and Pachowicz, P.W., "Application of symbolic machine learning to the recognition of texture concepts", *Proc. of the IEEE Conf. on Artificial Intelligence Applications*, Miami Beach, FL, pp. 224–230, 1991.

Bala, J. and Pachowicz, P.W., "Recognizing noisy patterns via iterative optimization and matching of their rule descriptions", *Intl. J. of Pattern Recognition and Artificial Intelligence*, Vol. 6, pp. 513–538, 1992.

Barlow, H.B., "Possible principles underlying the transformation of sensory messages", *Sensory Communication*, W.A. Rosenblith (ed.), MIT Press, Cambridge, MA, 1961.

Barlow, H.B., "Unsupervised learning", *Neural Computation* , Vol. 1, pp. 295–311, 1989.

Barth, M., Das, S. and Bhanu, B. "Learning-based control of perception for mobility", *Proc. of the IEEE Intl. Symp. on Intelligent Control*, pp. 329–334, 1992.

Barto, A.G., "Connectionist learning for control: An overview", *Neural Networks for Control*, T. Miller , R.S. Sutton and P.J. Werbos (eds.), MIT Press, Cambridge, MA, pp. 5–58, 1990.

Barto, A.G., "Some learning tasks from a control perspective", *1990 Lectures in Complex Systems*, L. Nadel and D.L. Stein (eds.), Addison-Wesley, Redwood City, CA, pp. 195–223, 1991.

Barto, A.G., "Reinforcement learning and adaptive critic methods", *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D.A. White and D.A. Sofge (eds.), Van Nostrand Reinhold, New York, pp. 469–491, 1992.

Basri, R. and Ullman, R., "Linear operator for object recognition", in *Advances in Neural Information Processing Systems,* Vol. 4, Morgan Kaufmann, San Mateo, CA, pp. 452–459, 1991.

Basye, K., Dean, T. and Vitter, J., "Coping with uncertainty in map learning", *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, Detroit, MI, pp. 663–668, 1989.

Becker, S. and Hinton, G.E., "Self-organizing neural network that discovers surfaces in random dot stereograms", *Nature*, Vol. 355, pp. 161–163, 1992.

Bergadano, F., Matwin, S., Michalski R. S. and Zhang, J., "Learning two-tiered descriptions of flexible concepts: The POSEIDON system", *Machine Learning*, Vol. 8, pp. 5–43, 1992.

Bhanu, B., "Automatic target recognition: State of the art survey", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 22, 1986.

Bhanu, B., "Machine learning in computer vision", Technical Report, Honeywell Systems and Research Center, Minneapolis, MN, 1988.

Bhanu, B. and Ming, J., "TRIPLE: A multi-strategy machine learning approach to target recognition", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, pp. 537–547, 1988.

Bhanu, B., Lee, S. and Ming, J., "Adaptive image segmentation using a genetic algorithm", *Proc.*

of the DARPA Image Understanding Workshop, Palo Alto, CA, pp. 1043–1055, 1989.

Bhanu, B., Lee, S. and Ming, J., "Self-optimizing control system for adaptive image segmentation", *Proc. of the DARPA Image Understanding Workshop,* Pittsburgh, PA, pp. 583–596, 1990.

Bhanu, B., Lee, S. and Ming, J., "Self-optimizing image segmentation system using genetic algorithm", *Proc. of the Intl. Conf. on Genetic Algorithms*, pp. 362–369, 1991.

Bhanu, B., Ming, J. and Lee, S., "Closed-loop adaptive image segmentation", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition,* Maui, HI, pp. 734–735, 1991.

Bhanu, B. and Das, S., "Computational learning for adaptive computer vision", Technical Report, University of California, Riverside, CA, 1992.

Bhanu, B. and Lee, S., *Adaptive Image Segmentation Using a Genetic Algorithm,* Kluwer Academic, Boston, MA, 1994 (in press).

Bialek, W., Rieke, F., de Ruyter-van-Steveninck, R.R. and Warland, D., "Reading a neural code", *Science*, Vol. 252, pp. 1854–1857, 1991.

Bolles, R.C. and Cain, R.A., "Recognizing and locating partially visible objects: The local feature focus method", *Intl. J. of Robotics Research,* Vol. 1, pp. 57–82, 1982.

Bolles, R.C., Horaud, P. and Hannah, M.J., "3DPO: A three-dimensional part orientation system", *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, pp. 1116–1120, 1983.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., *Classification and Regression Trees*, Wadsworth International, Belmont, CA, 1984.

Breuel, T.M., "Adaptive model base indexing", A.I. Memo 1008, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1989.

Brooks, R.A., "Symbolic reasoning among 3D models and 2D images", *Artificial Intelligence*, Vol. 17, pp. 285–348, 1981.

Brunelli, R. and Poggio, T., "HyperBF networks for real object recognition", *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, 1991.

Brunelli, R. and Poggio, T. "Face recognition: Features versus templates", Technical Report 9110-04, IRST, Trento, Italy, 1991.

Brunelli, R. and Poggio, T., "Caricatural effects in automated face perception", *Biological Cybernetics*, Vol. 69, pp. 235–241, 1993.

Brunelli, R. and Poggio, T., "Face recognition: Features versus templates", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, pp. 1042–1052, 1993.

Buelthoff, H.H. and Edelman, S., "Psychophysical support for a 2-D view interpolation theory of object recognition", *Proc. of the Natl. Academy of Science,* Vol. 89, pp. 60–64, 1992.

Camps, O.I., Shapiro, L.G. and Haralick, R.M., "PREMIO: An overview", *Proc. of the IEEE*

Workshop on Directions in Automated CAD-Based Vision, Maui, HI, pp. 11–21, 1991.

Carpenter, G.A., "Neural network models for pattern recognition and associative memory", *Neural Networks*, Vol. 2, pp. 243–257, 1989.

Carpenter, G.A. and Grossberg, S., "A massively parallel architecture for a self-organizing neural pattern recognition machine", *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp. 54–115, 1987.

Carpenter, G.A., Grossberg, S. and Reynolds, J., "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network", *IEEE Expert*, Vol. 6, 1991.

Carpenter, G.A., Grossberg, S. and Reynolds, J., "A neural network architecture for fast on-line supervised learning and pattern recognition", *Neural Networks for Perception*, Vol. 1, H. Wechsler (ed.), Academic Press, Boston, MA, pp. 248–264, 1992.

Chan, T.Y.T. and Goldfarb, L., "Primitive pattern learning", *Pattern Recognition*, Vol. 25, pp. 883–889, 1992.

Channic, T., "Texpert: An application of machine learning to texture recognition", Report MLI 89-17, George Mason University, Fairfax, VA, 1989.

Chapman, D. and Kaelbling, L., "Input generalization in delayed reinforcement learning: An algorithm and performance comparisons", *Proc. of the Intl. Joint Conf. on Artificial Intelligence*, Sydney, Australia, 1991.

Chen, C.H. and Mulgoankar, P.G., "Automatic vision programming", *CVGIP: Image Understanding*, Vol. 55, pp. 170–193, 1992.

Cheng, S., et al., "Development of renal osteodystrophy", *Proc. of the SPIE,* Vol. 1450, pp. 90–98, 1991.

Cho, K., "Learning shape classes", Ph.D. thesis, Rutgers University, New Brunswick, NJ, 1992.

Christiansen, A.D., Mason, M.T. and Mitchell, T.M., "Learning reliable manipulation strategies without initial physical models", *Robotics and Autonomous Systems*, 1991.

Connell, J., "A colony architecture for an artificial creature", Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

Connell, J.H. and Brady, M., "Learning shape descriptions", *Proc. of the Intl. Joint Conf. on Artificial Intelligence,* Los Angeles, CA, pp. 922–925, 1985.

Cooperstock, J. and Milios, E., "A neural network operated vision-guided mobile robot arm for docking and reaching", Technical Report RBCV-TR-92-39, University of Toronto, 1992.

Cromwell, R.L. and Kak, A.C., "Automatic generation of object class descriptions using symbolic learning techniques", *Proc. of the Natl. Conf. on Artificial Intelligence*, Anaheim, CA, pp. 710–717, 1991.

Cutkosky, M.R., "On grasp choice, grasp models and the design of hands for manufacturing tasks", *IEEE J. of Robotics and Automation*, Vol. 5, pp. 269–279, 1989.

Das, S. and Bhanu, B., "Computational vision: A learning perspective", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, submitted.

Daugman, J.G., "An information-theoretic view of analog representation in striate cortex", *Computational Neuroscience,* E.L. Schwartz (ed.), MIT Press, Cambridge, MA, pp. 403–423, 1990.

Davis, E., *Representing and Acquiring Geographic Knowledge*, Morgan Kaufmann, Los Altos, CA, 1986.

Davis, E., "Inferring ignorance from the locality of visual perception", *Proc. of the Natl. Conf. on Artificial Intelligence,* St. Paul, MN, pp. 786–790, 1988.

Dean, T., Angluin, D., Basye, K., Engelson, S., Kaelbling, L., Kokkevis, E. and Maron, O., "Inferring finite automata with stochastic output functions and an application to map learning", *Proc. of the Natl. Conf. on Artificial Intelligence*, San Jose, CA, pp. 208–214, 1992.

Dean, T., Basye, K., Chekaluk, R., Hyun, S., Lejter, M. and Randazza, M., "Coping with uncertainty in a control system for navigation and exploration", *Proc. of the Natl. Conf. on Artificial Intelligence*, Boston, MA, pp. 1010–1015, 1990.

Dean, T., Basye, K. and Kaelbling, L., "Uncertainty in graph-based map learning", *Robot Learning*, Connell, Jonathon and Mahadevan (eds.), Kluwer Academic, Boston, MA, 1992.

Dickmanns, E.D. and Graefe, V., "Dynamic monocular machine vision", *Machine Vision and Applications*, Vol. 1, pp. 223–240, 1988.

Donald, B. and Jennings, J., "Sensor interpretation and task-directed planning using perceptual equivalence classes", Technical Report, Cornell University, Ithaca, NY, 1991.

Draper, B.A., "Generalizing recognition strategies", *Proc. of the Applied Imagery Pattern Recognition Workshop*, Washington, DC, pp. 40–51, 1992.

Draper, B.A., "Learning object recognition strategies", Technical Report 93-50, University of Massachusetts, Amherst, MA, 1993.

Draper, B.A., "Statistical properties of learning recognition strategies", *Proc. of the ARPA Image Understanding Workshop*, Washington, DC, pp. 557–565, 1993.

Draper, B.A. and Hanson, A., "An example of learning in knowledge-directed vision", *Scandinavian Conf. on Image Analysis*, Aalborg, Denmark, August 1991. Reprinted in *Theory and Applications of Image Analysis,* P. Johansen and S. Olsen (eds.), World Scientific, Singapore, pp. 237–252, 1992.

Draper, B.A., Hanson, A. and Riseman, E., "Learning knowledge-directed visual strategies", *Proc. of the DARPA Image Understanding Workshop*, San Diego, CA., pp. 933–940, 1992.

Draper, B.A., Hanson, A.R. and Riseman, E.M., "Learning blackboard-based scheduling algorithms for computer vision", *Intl. J. of Pattern Recognition and Artificial Intelligence*, Vol. 7, pp. 309–328, 1993.

Draper, B.A. and Riseman, E.M., "Learning 3D object recognition strategies", *Proc. of the Intl. Conf. on Computer Vision*, Osaka, Japan, pp. 320–324, 1990.

Du Buf, J.M.H., Kardan, M. and Spann, M., "Texture feature performance for image segmentation", *Pattern Recognition*, Vol. 23, pp. 291–309, 1990.

Dunn, G.B. and Segen, J., "Automatic discovery of robotic grasp configuration", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 396–401, 1988.

Dunn, S. and Cho, K., "Shape-based object recognition by inductive learning", *Proc. of the NATO Advanced Research Workshop on Shape in Picture*, Driebergen, The Netherlands, 1992.

Edelman, S. and Buelthoff, H.H., "Viewpoint-specific representations in 3D object recognition", A.I. Memo 1239, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Edelman, S. and Buelthoff, H.H. "Orientation dependence in the recognition of familiar and novel views of 3D objects", *Vision Research*, 1994 (in press).

Edelman, S. and Poggio, T., "Bringing the grandmother back into the picture: A memory-based view of object recognition", Artificial Intelligence Memo 1181/Center for Biological Information Processing Paper 52, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Fayyad, U.M., "On the induction of decision trees for multiple concept learning", Ph.D. thesis, University of Michigan, Ann Arbor, MI, 1991.

Fayyad, U.M., "Applying machine learning classification techniques to automate sky object cataloguing", *Proc. of the Intl. Space Year Conf. on Earth and Space Science Information Systems*, Pasadena, CA, 1992.

Fayyad, U.M., Doyle, R.J., Weir, N. and Djorgovski, S.G., "Automating sky object classification in astronomical survey images", *Proc. of the Machine Discovery Workshop, Intl. Conf. on Machine Learning*, Aberdeen, Scotland, 1992.

Fayyad, U.M. and Irani, K.B., "The attribute selection problem in decision tree generation", *Proc. of the Natl. Conf. on Artificial Intelligence*, San Jose, CA, pp. 104–110, 1992.

Field, D.J., "Relations between the statistics of natural images and the response properties of cortical cells", *J. of the Optical Society of America*, Vol. A4, pp. 2379–2394, 1987.

Field, D.J., "What the statistics of natural images tell us about visual coding", *Proc. of the SPIE*, Vol. 1077, pp. 269-276, 1989.

Fox, J. and Walker, N., "Knowledge-based interpretation of medical images", in *Mathematics and Computer Science in Medical Images*, M.A. Viergever and A. Todd-Pokropek (eds.), Springer-Verlag, Berlin, 1988.

Girosi, F. and Poggio, T., "Networks and the best approximation property", *Biological Cybernetics*, Vol. 63, pp. 169–176, 1990.

Girosi, F., Poggio, T. and Caprile, B., "Extensions of a theory of networks for approximation and learning: outliers and negative examples", Artificial Intelligence Laboratory Memo 1220/Center for Biological Information Processing Paper 46, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Goldfarb, L., "On the foundations of intelligent processes—I. An evolving model for pattern learning", *Pattern Recognition,* Vol. 23, pp. 595–616, 1990.

Gong, L., Kulikowski, C. and Mezrich, R., "Automatic generation of plans for biomedical image interpretation", *Proc. of the SCAMC,* pp. 465–469, 1991.

Gong, L., Kulikowski, C. and Mezrich, R., "Knowledge-based experimental design for planning biomedical image interpretation", *Proc. of MEDINFO-92*, pp. 628–634, 1992.

Gool, L.V., Dewaele, P. and Oosterlink, A., "Texture analysis anno 1983", *Computer Vision, Graphics, and Image Processing,* Vol. 29, pp. 336–357, 1985.

Grimson, W.E.L. and Lozano-Perez, T., "Localizing overlapping parts by searching the interpretation tree", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 9, pp. 469–482, 1987.

Haralick, R.M., "Statistical and structural approaches to texture", *Proc. of the IEEE,* Vol. 67, pp. 786–804, 1979.

Haralick, R.M., Shanmugam, K. and Dinstein, I., "Texture features for image classification", *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. 3, pp. 610–621, 1973.

Harlee, W., et al., "MRI tissue characterization and segmentation of human brain tissues using a Prolog-based expert system", in *Tissue Characterization in MRI Imaging,* H. Higer and G. Bielke (eds.), Springer Verlag, Berlin, pp. 313–319, 1990.

Hebb, D.O., *The Organization of Behavior,* Wiley, New York, 1949.

Hinton, G.E. and Sejnowski, T.J., "Optimal perceptual inference", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 448–453, 1983.

Hinton, G.E., Williams, C.K.I., and Revow, M.D., "Adaptive elastic models for hand-printed character recognition", in *Advances in Neural Information Processing Systems,* Vol. 4, J.E. Moody, S.J. Hanson and R.P. Lippmann (eds.), Morgan Kaufmann, San Mateo, CA, pp. 512–519, 1992.

Hurlbert, A. and Poggio, T., "Synthesizing a color algorithm from examples", *Science,* Vol. 239, pp. 482–485, 1988.

Huttenlocher, D.P. and Ullman, S., "Object recognition using alignment", *Proc. of the Intl. Conf. on Computer Vision,* London, England, pp. 102–111, 1987.

Ikeuchi, K., "Generating an interpretation tree from a CAD model for 3D object recognition in bin-picking tasks", *Intl. J. of Computer Vision,* Vol. 1, pp. 145–165, 1987.

Ikeuchi, K. and Hong, K.S., "Determining linear shape change: Toward automatic generation of object recognition programs", *CVGIP: Image Understanding,* Vol. 53, pp. 154–170, 1991.

Ikeuchi, K. and Suehiro, T., "Towards an assembly plan from observation: Task recognition with polyhedral objects", Technical Report CMU-CS-91-167, Carnegie Mellon University, Pittsburgh, PA, 1991.

Intrator, N., Gold, J.I., Bulthoff, H.H. and Edelman, S., "3D object recognition using unsupervised feature extraction", in *Advances in Neural Information Processing Systems*, Vol. 4, pp. 461–467, Morgan Kaufmann, San Mateo, CA, 1991.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J. and Hinton, G.E., "Adaptive mixtures of local experts", *Neural Computation*, Vol. 3, 1991.

Jordan, M.I. and Rumelhart, D.E., "Forward models: Supervised learning with a distal teacher", *Cognitive Science*, Vol. 16, No. 3, pp. 307-354, 1992.

Julesz, B., "Visual pattern discrimination", *IRE Trans. on Information Theory*, Vol. 8, pp. 84–92, 1962.

Julesz, B., "Experiments in visual perception of texture", *Scientific American*, Vol. 232, pp. 34–43, 1975.

Kaelbling, L.P., "Learning in embedded systems", Report STAN-CS-90-1326, Stanford University, Stanford, CA, 1990.

Karpinski, J. and Michalski, R.S., "A digital system recognizing handwritten alphanumeric characters: General mathematical concepts and experimental data from an operational computer simulator", *Reports of the Institute of Automatic Control*, No. 35, Polish Academy of Sciences, Warsaw, Poland, p. 89, 1966 (in Polish).

Keeler, J.D., Rumelhart, D.E., and Leow, W.K., "Integrated segmentation and recognition of hand-printed numerals", in *Advances in Neural Information Processing Systems*, Vol. 3, R.P. Lippmann, J.E. Moody and D.S. Touretzky (eds.), pp. 557–563, Morgan Kaufmann, San Mateo, CA, 1991.

Kodratoff, Y. and Lemerle-Loisel, R., "Learning complex structural descriptions from examples", *Computer Vision, Graphics, and Image Processing*, Vol. 27, pp. 266–290, 1984.

Kuipers, B., "Modeling spatial knowledge", *Cognitive Science*, Vol. 2, pp. 129–153, 1978.

Kuipers, B. and Levitt, T.S., "Navigation and mapping in large-scale space", *AI Magazine*, Vol. 9, pp. 25–43, 1988.

Kuipers, B. and Byun, Y.-T., "A robust, qualitative method for robot spatial reasoning", *Proc. of the Natl. Conf. on Artificial Intelligence*, St. Paul, MN, pp. 774–779, 1988.

Kuniyoshi, T., Inaba, M. and Inoue, H., "Teaching by showing: Generating robot programs by visual observation of human performance", *Proc. of the Intl. Symp. on Industrial Robotics*, pp. 119–126, 1989.

Kuperstein, M., "Neural model of adaptive hand-eye coordination for single postures", *Science*, Vol. 239, pp. 1308-1311, 1988.

Le Cun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L., "Backpropagation applied to handwritten zip code recognition", *Neural Computation*, Vol. 1, pp. 541–551, 1989.

Le Cun, Y., Bozer, B., Denker, J., Howard, D., Hubbard, R. and Jackel, L., "Handwritten digit recognition with a back-propagation network", in *Advances in Neural Information Processing Systems*, Vol. 2, pp. 396–404, Morgan Kaufmann, San Mateo, CA, 1990.

Levitt, T.S., Lawton, D.T., Chelberg, D.M. and Nelson, P.C., "Qualitative landmark-based path planning and following", *Proc. of the Natl. Conf. on Artificial Intelligence*, Seattle, WA, pp. 689–694, 1987.

Levitt, T.S. and Lawton, D.T., "Qualitative navigation for mobile robots", *Artificial Intelligence*, Vol. 44, pp. 305–360, 1990.

Lew, M.S., Huang, T.S. and Wong, K., "Learning and feature selection in stereo matching", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994, in press.

Li, R.Y., "Model-based target recognition using laser radar imagery", *Optical Engineering*, Vol. 31, 1992.

Lin, L.-J., "Programming robots using reinforcement learning and teaching", *Proc. of the Natl. Conf. on Artificial Intelligence*, Anaheim, CA, 1991.

Linsker, R., "From basic network principles to neural architecture", *Proc. of the Natl. Academy of Science*, Vol. 83, pp. 7508–7512, pp. 8390–8394, pp. 8779–8783, 1986.

Linsker, R., "Towards an organizing principle for a layered perceptual network", *Proc. of the Conf. on Neural Information Processing Systems*, Denver, CO, pp. 485–494, 1987.

Linsker, R., "Self-organization in a perceptual network", *IEEE Computer*, Vol. 21, pp. 105–117, 1988.

Linsker, R., "An application of the principle of maximum information preservation to linear systems", in *Advances in Neural Information Processing Systems*, Vol. 1, D.S. Touretzky (ed.), Morgan Kaufmann, San Mateo, CA, pp. 86–94, 1989.

Linsker, R., "How to generate ordered maps by maximizing the mutual information between input and output signals", *Neural Computation*, Vol. 1, pp. 402–411, 1989.

Linsker, R., "Perceptual neural organization: Some approaches based on network models and information theory", *Annual Review of Neuroscience*, Vol. 13, pp. 257–281, 1990.

Linsker, R., "Emergence of organization in cortex", *Neuroscience Facts*, Vol. 3, p. 60, 1992.

Linsker, R., "Local synaptic learning rules suffice to maximize mutual information in a linear network", *Neural Computation*, Vol. 4, pp. 672–683, 1992.

Linsker, R., "Deriving receptive fields using an optimal encoding criterion", in *Advances in Neural Information Processing Systems*, Vol. 5, S.J. Hanson et al. (eds.), Morgan Kaufmann, San Mateo, CA, pp. 953–960, 1993.

Linsker, R., "Sensory processing and information theory", in *From Statistical Physics to Statistical Inference and Back*, P. Grassberger and J.-P. Nadal (eds.), Kluwer, Dordrecht, The

Netherlands, pp. 237–247, 1994.

Liu, H., Iberall, T. and Bekey, G.A., "The multi-dimensional quality of tasks requirements of task requirements for dexterous hand control", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 452–457, 1989.

Logothetis, N.K., Pauls, J. and Poggio, T., "Viewer-centered object recognition in monkeys", Artificial Intelligence Memo 1473/Center for Biological and Computational Learning Paper 95, Massachusetts Institute of Technology, Cambridge, MA, April 1994.

Lowe, D.G., *Perceptual Organization and Visual Recognition*, Kluwer Academic, Boston, MA, 1986.

Lozano-Perez, T., "A simple motion-planning algorithm for general robot manipulators", *IEEE J. of Robotics and Automation*, Vol. 3, pp. 224–238, 1987.

Lu, S.Y. and Fu, K.S., "A syntactic approach to texture analysis", *Computer Graphics and Image Processing,* Vol. 7, pp. 303–330, 1978.

Mahadevan, S. and Connell, J., "Automatic programming of behavior-based robots using reinforcement learning", *Proc. of the Natl. Conf. on Artificial Intelligence*, Anaheim, CA, 1991.

Mataric, M.J., "A distributed model for mobile robot environment learning and navigation", Technical Report 1128, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Mataric, M.M., "Environment learning using a distributed representation", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 402–406, 1990.

Matsuyama, T., "Expert systems for image processing: Knowledge-based composition of image analysis processes", *Computer Vision, Graphics, and Image Processing*, Vol. 48, pp. 22–49, 1989.

McDermott, D.V. and Davis, E., "Planning routes through uncertain territory", *Artificial Intelligence*, Vol. 22, pp. 107–156, 1982.

Mel, B., *Connectionist Robot Motion Planning: A Neurally Inspired Approach to Visually Guided Reaching*, Academic Press, San Diego, CA, 1991.

Michalski, R.S., "Problems of computer simulation of pattern recognition systems and a description of system recognizing handwritten alphanumeric characters", *Proc. of the Learning Automata and Biological Processes of Perception Conference*, pp. 152–182, Polish Academy of Sciences, Warsaw, Poland, 1966 (in Polish).

Michalski, R.S., "A variable-valued logic system as applied to picture description and recognition", in *Graphic Languages*, F. Nake and A. Rosenfeld (eds.), North-Holland, Amsterdam, pp. 20–47, 1972.

Michalski R.S., "AQVAL/1: Computer implementation of the variable valued logic system VL1 and examples of its application to pattern recognition", *Proc. of the Intl. Joint Conf. on Pattern Recognition,* Washington, DC, 1973.

Michalski, R.S., "Variable-valued logic and its applications to pattern recognition and machine learning", *Computer Science and Multiple-Valued Logic Theory and Applications*, D.C. Rine (ed.), North-Holland, Amsterdam, pp. 506–534, 1975.

Michalski, R.S., "A theory and methodology of inductive learning", *Artificial Intelligence*, Vol. 23, pp. 111–161, 1983.

Michalski, R.S., Rosenfeld, A., Pachowicz, P. and Aloimonos, Y., Machine learning and vision: Research issues and promising directions, a preliminary report on the NSF/ARPA Workshop on Machine Learning and Vision, Report MLI-93-1, George Mason University, Fairfax, VA, 1993.

Millan, J.D.R. and Torras, C., "Learning to avoid obstacles through reinforcement", *Proc. of the Intl. Workshop on Machine Learning*, pp. 298–302, 1991.

Miller, W.T., "Sensor-based control of robotic manipulators using a general learning algorithm", *IEEE J. of Robotics and Automation*, Vol. 3, pp. 157–165, 1987.

Miller, W.T., Sutton, R.S. and Werbos, P.J., *Neural Networks for Control*, MIT Press, Cambridge, MA, 1990.

Ming, J. and Bhanu, B., "A multistrategy learning approach for target model recognition, acquisition, and refinement", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA, pp. 742–756, 1990.

Ming, J. and Bhanu, B., "TRIPLE: A multistrategy machine learning approach to target recognition", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, pp. 537–547, 1988.

Moore, A.W., "Acquisition of dynamic control knowledge for a robotic manipulator", *Proc. of the Intl. Conf. on Machine Learning*, 1990.

Moravec, H.P. and Elfes, A., "High resolution maps from wide-angle sonar", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, St. Louis, MO, pp. 138–145, 1985.

Nazif, A. and Levine, M.D., "Low level image segmentation: An expert system", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 555–577, 1984.

Negahdaripour, S. and Jain, A.K. (eds.), "Challenges in computer vision research: Future directions of research", report of an NSF Workshop, Maui, HI, June 1991.

Ourston, D., "Changing the rules: A comprehensive approach to theory refinement", *Proc. of the Natl. Conf. on Artificial Intelligence*, Boston, MA, pp. 815–820, 1990.

Ozveren, C.M., "Analysis and control of discrete event dynamic systems: A state space approach", Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Pachowicz, P.W., "Low-level numerical characteristics and inductive learning methodology in texture recognition", *Proc. of the Intl. Workshop on Tools for Artificial Intelligence*, Washington, DC, pp. 91–98, 1989.

Pachowicz, P.W., "Learning-based architecture for the robust recognition of variable texture to navigate in natural terrain", *Proc. of the Intl. Workshop on Intelligent Robot Systems*, Japan, pp. 135–142, 1990.

Pachowicz, P.W., "Integrating low-level feature computation with inductive learning techniques for texture recognition", *Intl. J. of Pattern Recognition and Artificial Intelligence*, Vol. 4, pp. 147–165, 1990.

Pachowicz, P.W., "Learning invariant texture characteristics in dynamic environments: A model evolution approach", Report MLI-2-91, George Mason University, Fairfax, VA, 1991.

Pachowicz, P.W., "Application of symbolic inductive learning to the acquisition and recognition of noisy texture concepts", *Applications of Learning and Planning Methods*, World Scientific, Singapore, pp. 99–127, 1991.

Pachowicz, P.W., "Recognizing and evolving texture concepts in dynamic environments: An incremental model generalization approach", Report MLI-91-10, George Mason University, Fairfax, VA, 1991.

Pachowicz, P.W., "A learning-based evolution of concept descriptions for an adaptive object recognition", *Proc. of the IEEE Conf. on Tools with Artificial Intelligence*, Arlington, VA, pp. 316–323, 1992.

Pachowicz, P.W., "Invariant object recognition: A model evolution approach", *Proc. of the DARPA Image Understanding Workshop*, Washington, DC, 1993.

Pachowicz, P.W., "Semi-autonomous evolution of object models for adaptive object recognition", *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 24, 1994.

Pachowicz, P.W. and Bala, J., "Improving recognition effectiveness of noisy texture concepts through optimization of their descriptions", *Proc of the Intl. Workshop on Machine Learning*, Evanston, IL, pp. 625–629, 1991.

Pachowicz, P.W., Bala, J. and Zhang, J., "Methodology for iterative noise-tolerant learning and its application to object recognition in computer vision", *Proc. of the Intl. Conf. on Systems Research, Informatics and Cybernetics*, Baden-Baden, Germany, 1992.

Pachowicz, P.W., Bala, J. and Zhang, J., "Iterative rule simplification for noise tolerant inductive learning", *Proc. of the IEEE Conf. on Tools with Artificial Intelligence*, Arlington, VA, pp. 452–453, 1992.

Pachowicz, P.W., Hieb, M. and Mohta, P., "A learning-based incremental model evolution for invariant object recognition", *Proc. of the Intl. Conf. on Systems Research, Informatics and Cyberetics*, Baden-Baden, Germany, 1992.

Park, I.P., "Qualitative navigation using isolated landmarks", Technical Report CUCS-015-92, Columbia University, New York, 1992.

Pazzani, M. and Kibler D., "The utility of knowledge in inductive learning", *Machine Learning*, Vol. 9, pp. 57–94, 1992.

Pearlmutter, B.A. and Hinton, G.E., "G-maximization: An unsupervised learning procedure for discovering regularities", *Neural Networks for Computing*, J.S. Denker (ed.), pp. 333–338, American Institute of Physics, New York, 1986.

Pednault, E.P.D., "Some experiments in applying inductive inference principles to surface reconstruction", *Proc. of the Intl. Joint Conf. on Artificial Intelligence,* pp. 1603–1609, 1989.

Pipitone, F., Tripods, Report No. 6780, Naval Research Laboratory, Washington, DC, 1991.

Pipitone, F., "Tripod operators for recognizing objects in range images: Rapid rejection of library objects", *Proc. of the IEEE Conf. on Robotics and Automation*, Nice, France, pp. 1596–1601, 1992.

Pipitone, F. and Adams, W., "Rapid recognition of freeform objects in noisy range images using tripod operators", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, New York, pp. 715–716, 1993.

Piraino, D., et al., "Segmentation of magnetic resonance images using an artificial neural network", *Proc. of the SCAMC*, pp. 470–472, 1991.

Poggio, T., Gamble, E. and Little J., "Parallel integration of vision modules", *Science,* Vol. 242, pp. 436–440, 1988.

Poggio, T., "A theory of how the brain might work", Artificial Intelligence Memo 1253/Center for Biological Information Processing Paper 50, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Poggio, T., "3D object recognition: On a result of Basri and Ullman", Technical Report 9005-03, IRST, Trento, Italy, 1990.

Poggio, T., Little, J., Gamble, E., Gillett, W., Geiger, D., Weinshall, D., Villalba, M., Larson, N., Cass, T., Bulthoff, H., Drumheller, M., Oppenheimer, P., Yang, W. and Hurlbert, A., "The MIT Vision Machine", *Proc. of the DARPA Image Understanding Workshop*, McLean, VA, pp. 177–198, 1988.

Poggio T. and Girosi, F., "A theory of networks for approximation and learning", Artificial Intelligence Laboratory Memo 1140/Center for Biological Information Processing Paper 31, Massachusetts Institute of Technology, Cambridge, MA, 1989.

Poggio, T., "A theory of how the brain might work", *Proc. of the Cold Spring Harbor Symposium on Quantitative Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 899–910, 1990.

Poggio, T. and Brunelli, R., "A novel approach to graphics", Artificial Intelligence Laboratory Memo 1354/Center for Biological Information Processing Paper 71, Massachusetts Institute of Technology, Cambridge, MA, 1992.

Poggio, T. and Edelman, S., "A network that learns to recognize three-dimensional objects", *Nature*, Vol. 343, pp. 263–266, 1990.

Poggio, T., Edelman, S. and Fahle, M., "Learning of visual modules from examples: A framework for understanding adaptive visual performance", *CVGIP: Image Understanding*, Vol. 56, pp. 22–30, 1992.

Poggio, T. and Girosi, F., "Networks for approximation and learning", *Proc. of the IEEE,* Vol. 78, pp. 1481–1497, 1990.

Poggio, T. and Girosi, F., "Regularization algorithms for learning that are equivalent to multilayer networks", *Science*, Vol. 247, pp. 978–982, 1990.

Poggio, T. and Girosi, F., "Extensions of a theory of networks for approximation and learning: Dimensionality reduction and clustering", Artificial Intelligence Laboratory Memo 1167/Center for Biological Information Processing Paper 44, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Poggio, T. and Hurlbert, A., "Observations on cortical mechanisms for object recognition and learning", Artificial Intelligence Memo 1404/Center for Biological and Computational Learning Paper 77, Massachusetts Institute of Technology, Cambridge, MA, 1993.

Poggio, T. and Vetter, T., "Recognition and structure from one 2-D model view: Some observations on prototypes, object classes and symmetries", *Artificial Intelligence Memo 1347/Center for Biological Information Processing Paper 69*, Massachusetts Institute of Technology, Cambridge, MA, 1992.

Poggio, T. and Vetter, T., "Recognition and structure from one 2-D model view: Some observations on prototypes, object classes and symmetries", *Optics 1992 Conference: From Galileo's "Occhialino" to Optoelectronics*, Padova, Italy, 1992.

Pomerleau, D., "ALVINN: An autonomous land vehicle in a neural network", Technical Report CMU-CS-89-107, Carnegie-Mellon University, Pittsburgh, PA, 1989.

Pomerleau, D.A., "Efficient training of artificial neural networks for autonomous navigation", *Neural Computation*, Vol. 3, pp. 88–97, 1991.

Pomerleau, D.A., Gowdy, J., and Thorpe, C.E., "Combining artificial neural networks and symbolic processing for autonomous robot guidance", *Engineering Applications of Artificial Intelligence*, Vol. 4, pp. 279–285, 1991.

Prescott, T.J. and Mayhew, J.E.W., "Obstacle avoidance through reinforcement learning", *Advances in Neural Information Processing Systems*, Vol. 4, pp. 523–530, Morgan Kaufmann, San Mateo, CA, 1991.

Quinlan, J.R., "Learning efficient classification procedures and their application to chess end games", pp. 463–481, Morgan Kaufmann, Los Altos, Ca., 1986.

Ramadge, P.J. and Wonham, W.M., "Supervisory control of a class of discrete event processes", *SIAM J. of Control and Optimization*, 1987.

Ramirez, M. and Sunanda, M., "Three-dimensional target recognition from fusion of dense range and intensity images", *Proc. of the SPIE*, Vol. 1567, 1991.

Rao, A.R. and Jain, R.C., "Computerized flow field analysis: oriented texture fields", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, pp. 693–709, 1992.

Raya, S.P., "Low-level segmentation of 3-D magnetic resonance brain images—A rule-based system", *IEEE Trans. on Medical Imaging*, Vol. 9, pp. 327–337, 1990.

Reed T.R., "Region growing using neural networks", *Neural Networks for Perception*, Vol. 1, H. Wechsler (ed.), pp. 386–397, Academic Press, Boston, MA, 1992.

Rimey, R. D. and Brown, C. M., "Controlling eye movements with hidden Markov models", *Intl. J. of Computer Vision*, Vol. 7, pp. 47–66, 1991.

Ritter, H., Martinetz, T. and Schulten, K., "Topology conserving map for learning visuomotor coordination", *Neural Networks*, Vol. 2, pp. 159–168, 1989.

Ritter, H., Martinetz, T. and Schulten, K., *Neural Computation and Self-Organizing Maps: An Introduction*, Addison-Wesley, Reading, MA, 1991.

Rosenfeld, A. and Troy, E., "Visual texture analysis", *Proc. of the IEEE Conf. on Feature Extraction and Selection in Pattern Recognition*, pp. 115–124, 1970.

Salganicoff, M., "Learning and forgetting for perception-action: A projection-pursuit and density-adaptive approach", Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, 1992.

Salganicoff, M., "Density-adaptive learning and forgetting", *Proc. of the Intl. Workshop on Machine Learning*, Amherst, MA, 1993.

Segen, J., "Learning structural descriptions of shape", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 96–98, 1985.

Segen, J., "Learning structural descriptions of shape", *Machine Vision: Algorithms, Architectures and Systems*, H. Freeman (ed.), Academic Press, Boston, MA, 1988.

Segen, J., "Learning graph models of shape", *Proc. of the Intl. Conf. on Machine Learning*, Ann Arbor, MI, pp. 29–35, 1988.

Segen, J., "Model learning and recognition of nonrigid objects", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, pp. 597–602, 1989.

Segen, J., "From features to symbols: Learning relational models of shape", *From Pixels to Features*, J.C. Simon (ed.), North Holland, Amsterdam, 1989.

Segen, J., "Graph matching, interpretation and clustering using minimal representation criterion", *AAAI Spring Symp. Series: Theory and Applications of Minimal-Length Encoding*, Stanford, CA, 1990.

Segen, J., "Graph clustering and model learning by data compression", *Proc. of the Intl. Conf. on Machine Learning*, Austin, TX, pp. 93–101, 1990.

Segen, J., "GEST: An integrated approach to learning in computer vision", *Intl. Workshop on Multistrategy Learning*, Harpers Ferry, WV, 1991.

Segen, J., "GEST: A learning computer vision system that recognizes hand gestures", in *Machine Learning IV*, R.S. Michalski and G. Tecuci (eds.), Morgan Kaufmann, San Mateo, CA, 1992.

Segen, J., "Inference of stochastic graph models for 2-D and 3-D shape", *Proc. of the NATO Advanced Research Workshop on Shape in Picture*, Driebergen, The Netherlands, 1992.

Segen, J. and Dana, K., "Parallel symbolic recognition of deformable shapes", *From Pixels to Features* II, H. Burkhardt, Y. Nuevo and J.C. Simon (eds.), North-Holland, Amsterdam, 1991.

Seibert, M. and Waxman, A.M., "Learning aspect graph representations from view sequences", *Advances in Neural Information Processing*, Vol. 2, D.S. Touretzky (ed.), Morgan Kaufmann, San Mateo, CA, pp. 258–265, 1990.

Seibert, M. and Waxman, A.M., "Learning and recognizing 3D objects from multiple views in a neural system", *Neural Networks for Perception*, Vol. 1, H. Wechsler (ed.), Academic Press, Boston, MA, pp. 426–443, 1992.

Shannon, C.E., in *The Mathematical Theory of Communication*, C.E. Shannon and W. Weaver (eds.), University of Illinois Press, Urbana, IL, 1949.

Shvaytser, H., "Learnable and nonlearnable visual concepts", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 12, pp. 459–466, 1990.

Simmons, R. and Krotkov, E., "An integrated walking system for the Ambler planetary rover", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 2086–2091, 1991.

Singh, S., "Transfer of learning across compositions of sequential tasks", *Proc. of the Intl. Workshop on Machine Learning*, pp. 348–352, 1991.

Sobh, T.M., "A framework for visual observation", Technical Report MS-CIS-91-36, GRASP LAB 261, University of Pennsylvania, Philadelphia, PA, 1991.

Stansfield, S.A., "ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtraction angiograms", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 8, pp. 188–199, 1986.

Stansfield, S.A., "Knowledge-based robotic grasping", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 1270–1275, 1990.

Suganuma, Y., "Learning structures of visual patterns from single instances", *Artificial Intelligence,* Vol. 50, pp. 1–36, 1991.

Sullins, J.R., "Boolean learning in neural networks", Technical Report CAR-TR-359, University of Maryland, College Park, MD, 1988.

Sullins, J.R., "Distributed learning: Motion in constraint space", Technical Report CAR-TR-412, University of Maryland, College Park, MD, 1988.

Sullins, J.R., "Distributed learning of texture classification", Technical Report CAR-TR-444, University of Maryland, College Park, MD, 1989.

Sullins, J.R., "Distributed learning under invariance", Technical Report CAR-TR-479, University of Maryland, College Park, MD, 1989.

Sutton, R.S., "Learning to predict by the methods of temporal differencing", *Machine Learning*, Vol. 3, pp. 9–44, 1988.

Sutton, R.S., "Integrated architectures for learning planning and reacting based on approximating dynamic programming", *Proc. of the Intl. Conf. on Machine Learning*, Austin, TX, 1990.

Sutton, R.S., Barto, A.G. and Williams, R.J., "Reinforcement learning is direct adaptive optimal control", *Proc. of the American Control Conf.*, Boston, MA, pp. 2143–2146, 1991.

Swain, M.J. and Stricker, M. (eds.), "Promising directions in active vision, a report written by the attendees of the NSF Active Vision Workshop", Report CS 91-27, University of Chicago, Chicago, IL, 1991.

Tan, M., "A cost-sensitive learning system for sensing and grasping objects", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 1990.

Tham, C.K. and Prager, R.W., "Reinforcement learning for multi-linked manipulator control", Technical Report CUED/F-INFENG 104, Cambridge University, Cambridge, UK, 1991.

Theunissen, F.E. and Miller, J., "Representation of sensory information in the cricket cercal sensory system", *Neurophysiology* , Vol. 66, pp. 1680–1689, pp. 1690–1703.

Thompson, D.W. and Mundy, J.L., "Three-dimensional model matching", *Proc. of the IEEE Conf. on Robotics and Automation*, Raleigh, NC, pp. 208–220, 1987.

Thrun, S. and Moller, K., "Active exploration in dynamic environments", *Advances in Neural Information Processing Systems*, Vol. 4, pp. 531–538, Morgan Kaufmann, San Mateo, CA, 1991.

Thrun, S.B., Bala, J., Bloedorn, E., Bratko, I., Cestnik, B., Cheng, J., De Jong, K.A., Dzeroski, S., Fahlman, S.E., Hamann, R., Kaufman, K., Keller, S., Kononenko, I., Kreuziger, J., Michalski, R.S., Mitchell, T., Pachowicz, P., Vafaie, H., Van de Velde, W., Wenzel, W., Wnek, J. and Zhang, J., "The MONK's problems: A performance comparison of different learning algorithms", Report CMU-CS-91-197, Carnegie Mellon University, Pittsburgh, PA, 1991.

Tomovic, R., Bekey, G. and Karplus, W., "A strategy for grasp synthesis with multifingered robot hands", *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pp. 83–89, 1986.

Towell, G.G. and Shavlik, J.W., "Hybrid symbolic-neural methods for improved recognition using high-level visual features", *Neural Networks for Perception*, Vol. 1, H. Wechsler (ed.), Academic Press, Boston, MA, pp. 445–461, 1992.

Towell, G.G., Shavlik, J.W. and Noordewier, M.O., "Refinement of approximately correct domain theories by knowledge-based neural networks", *Proc. of the Natl. Conf. on Artificial Intelligence*, Boston, MA, pp. 861–866, 1990.

Tririon, E. and Quan, L., "Geometrical learning from multiple stereo views through monocular based feature grouping", *Proc. of the Intl. Conf. on Computer Vision,* Osaka, Japan, pp. 481–538, 1990.

Tucker, L.W., Feynman, C.R. and Fritzsche, D.M., "Object recognition using the Connection Machine", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition,* Ann Arbor, MI, pp. 871–878, 1988.

Ullman, S. and Basri, R., "Recognition by linear combination of models", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 992–1006, 1991.

Van Essen, D.C., Anderson, C.H. and Felleman, D.J., "Information processing in the primate visual system: An integrated systems perspective", *Science*, Vol. 255, pp. 419–423, 1992.

Vistnes, R., "Texture models and image measures for texture discrimination", *Intl. J. of Computer Vision*, Vol. 3, pp. 313–336, 1989.

Watkins, C.J.C.H., "Learning from delayed rewards", Ph.D. thesis, King's College, Cambridge, UK, 1989.

Wechsler, H., *Computational Vision*, Academic Press, Boston, MA, 1990.

Weinshall, D., "Model based invariants, and their use for representation, constant-time indexing, and linear structure from motion", IBM Report RC 17505, 1992.

Weir, N., Djorgovski, S.G., Fayyad, U.M., Roden, J. and Rouquette, N., "SKICAT: A system for the scientific analysis of the Palomar—STScI digital sky survey", in *Astronomy from Large Data Bases II*, A. Heck and F. Murtagh (eds.), Haguenau, France, 1992.

Weir, N., Djorgovski, S.G., Fayyad, U.M., Doyle, R.J. and Roden, J., "An analysis of the Palomar observatory-STScI digital sky survey: Catalog construction and classification results", American Astronomical Society, Boston, MA, 1992.

Weng, J.J., Ahuja, N. and Huang, T.S., "Learning recognition and segmentation of 3D objects from 2D images", *Proc. of the Intl. Conf. on Computer Vision*, Berlin, Germany, 1993.

Weszka, J.S., Dyer, C.R. and Rosenfeld, A., "A comparative study of texture measures for terrain classification", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 6, pp. 269–285, 1976.

White, D.A. and Sofge, D.A., *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, Van Nostrand Reinhold, New York, 1992.

Whitehead, S.D. and Ballard, D.H., "Active perception and reinforcement learning", *Neural Computation*, Vol. 2, pp. 409–419, 1990.

Winston, P.H., "Learning structural descriptions from examples", in *The Psychology of Computer Vision*, P.H. Winston (ed.), McGraw Hill, New York, 1975.

Wixson, L. and Ballard, D., "Learning efficient sensing sequences for object search", *Proc. of the AAAI Fall Symp. Series: Sensory Aspects of Robotic Intelligence*, pp. 166–173, 1992.

Wnek, J. and Michalski, R.S., "Hypothesis-driven constructive induction in AQ17: A method and experiments", *Proc. of the IJCAI-91 Workshop on Evaluating and Changing Representation in Machine Learning*, Sydney, Australia, 1991.

Wnek, J. and Michalski, R.S., "An experimental comparison of symbolic and subsymbolic learning paradigms: Learning logic-style concepts", *Proc. of the Intl. Workshop on Multistrategy Learning*, Harpers Ferry, WV, 1991.

Wnek, J. and Michalski, R.S., "Experimental comparison of symbolic and subsymbolic learning", *HEURISTICS, The Journal of Knowledge Engineering*, Vol. 5, 1992.

Wnek, J. and Michalski, R.S., "Comparing symbolic and subsymbolic learning: A case study", in *Machine Learning: A Multistrategy Approach*, R.S. Michalski and G. Tecuci (eds.), Morgan Kaufmann, San Mateo, CA, 1993.

Wong, A.K.C. and You, M., "Entropy and distance of random graphs with application to structural pattern recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 7, pp. 599–609, 1985.

Yang, G. and Huang, T.S., "Human face detection in a scene", *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 453–458, 1993.

Zhou, Y.T. and Chellappa, R., "Computation of optical flow using a neural network", *Proc. of the Intl. Conf. on Neural Networks*, San Diego, CA, pp. 71–78, 1988.

Zhou, Y.T. and Chellappa, R., "A network for motion perception", *Proc. of the Intl. Joint Conf. on Neural Networks*, San Diego, CA, pp. 875–884, 1990.

Zhou, Y.T. and Chellappa, R., "A neural network for motion processing", *Neural Networks for Perception*, Vol. 1, H. Wechsler (ed.), Academic Press, Boston, MA, pp. 492–516, 1992.

Zucker, S.W., Rosenfeld, A. and Davis, L.S., "Picture segmentation by texture discrimination", *IEEE Trans. on Computers*, Vol. 24, pp. 1228–1233, 1975.

Zytkow, J.M. and Pachowicz, P.W., "Fusion of vision and touch for spatio-temporal reasoning in learning manipulation tasks", *Proc. of the SPIE Symp. on Advances in Intelligent Robotics Systems*, Philadelphia, PA, pp. 404–415, 1989.