Recognition Cones: A Neuronal Architecture for Perception and Learning

by

Vasant Honavar
and
Leonard Uhr

Computer Sciences Technical Report #717
September 1987

RECOGNITION CONES: A NEURONAL ARCHITECTURE FOR PERCEPTION AND LEARNING

Vasant Honavar   Leonard Uhr

Computer Sciences Department
University of Wisconsin-Madison
Madison, WI 53706. U.S.A.

ABSTRACT

There is currently a great deal of interest and activity in developing connectionist, neuronal, brain-like models, in both Artificial Intelligence and Cognitive Science. This paper specifies the main features of such systems, and examines "recognition cone" models of perception from this perspective, as examples of structures of neuron-like units combined into successively larger brain-like modules. Issues addressed include architecture, information flow, and the parallel-distributed nature of processing and control in recognition cones; and their use in perception and learning.

1. Introduction

It is now widely recognized that massively parallel hardware/software structures are needed to perceive and understand the constantly changing environment. The recognition cones model of perception (Uhr1972, Uhr1987) examined in this paper is suggested by the brain, which serves as a living existence-proof of achievable perceptual and cognitive capabilities, and as a source of potentially useful mechanisms for achieving its capabilities by a computer system. The underlying hypothesis is that the larger structures of the brain - how its billions of neurons are connected in complex networks - e.g., areas of layers containing hypercolumns of columns of layered ensembles of neurons - is central to its function.

This is not to deny that neural nets are general-purpose computing structures, in exactly the same sense that Turing machines are general-purpose, since they are equivalent to finite state automata and McCulloch-Pitts networks. The problem is to develop the particular neural structure that can actually do what is needed, exactly as the problem is to develop the particular Turing machine program for a given function (e.g., chess, computer vision).

Both an analysis of what is needed and observations of the way the brain structures ensembles of neurons strongly suggest local and successively more global brain-like architectures that organize the components which, in most connectionist systems, are linked quasi-randomly. The particular structure of the brain has evolved to handle perceptual, cognitive, and motor functions needed to interact effectively with the larger environment while it is still capable of, perhaps less efficiently, other forms of computation.

This approach to machine perception differs significantly in philosophy from that in which "intelligent behavior" is realized in computer vision programs with little reference to underlying neural mechanisms. Perhaps less obviously, it also differs from brain modeling, wherein a neural net model of a specific portion of the brain is built and simulated, as faithfully as possible, based on the current neurophysiological

This research was partially supported by the NSF and the University of Wisconsin-Madison.

data, and explanations for and predictions of neural phenomena that might take place in that part of the brain are advanced. As do most systems that are today called "neuronal" or "connectionist," it uses basic units reminiscent of simplified neurons to build the larger system. But it also employs successively larger structures that are suggested by the brain's larger structures, in order to give the system the power needed for real-world perception.

Human beings recognize complex objects in substantially less than a second. This speed is achieved despite the fact that the brain's basic processing unit, the neuron, needs at least a millisecond to fire (compute a function of its inputs and transmit its outputs). Thus a neuron is 10^5 to 10^9 times slower in sending its electrochemical messages to other neurons to which it is connected than a typical computer switch. Yet, the brain is at least 10^3 to 10^6 times faster than today's computers (This is an estimate, since no existing computer vision program comes close to perceiving as well as the brain does). The only way to achieve adequately fast perception with such slow basic processors is by having very large numbers of them working in parallel, as does the brain. The brain contains on the order of 10^{11} neurons each of which may be connected to as many as 10^6 others (typically $3x10^4$ in the visual cortex). These facts have been a prime motivation for the current interest in massively parallel architectures for machine perception.

Over the past few years, a number of investigators in several different fields have begun to employ massively parallel models in their work. The general idea has been advocated in animal models (Arbib1979) and cognitive models (Anderson1977, Ratcliff1978). In machine perception research, massively parallel, cooperative computational theories have become a dominant paradigm (Marr1982, Davis1981, Ballard1984). Highly parallel, hierarchical models as well as computer architectures for perception have been advanced (Uhr1972, Hanson1978, Fukushima1975, Tanimoto1980, Schaefer1985, Uhr1987). Computational architectures for complex cognitive information processing with neuronal implementations have been suggested (Barnden1986, Fahlman1983, Grossberg1980). Computational properties of neural networks have been studied (Hopfield1982, Smolensky1984, Amari1977). Associative memories and parallel distributed knowledge representations have been advanced (Kohonen1977, Kohonen1984) and analyzed (Hanson1987). This list simply indicates, but by no means exhausts, the activity in this area.

2. Neuronal Models and Brain-like Models

Several authors have attempted a general characterization of the class of brain-like, neuronal, connectionist, parallel distributed processing models (Feldman1982, Rumelhart1986). There appears to be more or less general agreement as to what constitutes their salient aspects, although the various models discussed in the literature clearly differ from each other in detail to various degrees. We now simply state what we mean by neuronal and brain-like models, omiting the details of how the models can differ in terms of the stated criteria, but pointing out as appropriate how our criteria differ from those suggested by others.

2.1. Neuronal Systems Defined and Characterized

A neuronal system is a directed graph whose nodes compute functions on information passed to them via their input links, and send results via their output links.

Each node has, associated with it, an activation level or a state variable.

Each node computes one or more relatively simple neuron-like functions: its inputs and activation are integrated or in some other way combined, these results might then be evaluated (e.g., against a threshold); it then outputs accordingly and updates its activation level.

Each link has, associated with it, a transfer function.

Links transmit signals (e.g., packets of bits, symbols, numbers, etc.) between nodes.

Learning rules modify any of the following: processing functions of the nodes, transfer functions of the links, topology of the graph, and learning rules themselves.

The topology of the total graph, along with the functions that the individual nodes compute and the information input to them, determine the system's over-all behavior.

The total graph may be (successively) decomposable into relatively regular sub-graphs (e.g., layers, windows, columns, trees). From this, the system's behavior (including output, coordination, control, adaptation, learning) follow and emerge.

This contrasts with most "connectionist" systems (Rumelhart1986, Feldman1982) as follows:

In connectionist systems, links transmit only weights (typically, real numbers);

Nodes usually output a simple sum or threshold function over the inputs and activation;

Learning rules usually can modify only the weights associated with the links;

And usually organization into higher-level structures assumes a completely or randomly connected graph, or is left unspecified.

2.2. The Neuronal System and its External Environment

It is often convenient to treat the neuronal system as forming a closed system along with the "external environment". The external environment provides some of the inputs to, and accepts some of the outputs from, interface nodes in the neuronal system. This closed system may be partitioned into two sub-graphs: an "internal" subgraph (that is, the neuronal system with its input and output links with the external environment removed), and an "external subgraph" (that is, the external environment with its input and output links with the neuronal system removed). The input and output links between the internal and external sub-graphs are functionally similar to input and output links between the nodes of the internal sub-graph. The environment may be part of the real world, linked with the neuronal system via transducers like TV cameras or robot effectors, or a version of some aspects of the real world, simulated either by a computer program or by human beings interacting via keyboards and monitors.

2.3. Specification of a Neuronal System

In order to completely specify a particular neuronal system, one needs to define the topology of the graph, the processing functions of the nodes, the transfer functions associated with the links, the learning rules (which can, potentially, modify any of the above), and the external environment that, together with the neuronal system, forms a closed system. The behavior of the system results from the dynamics of the system, that is, from the interaction between the large number of units acting in parallel at each moment, over a period of time.

2.4. Requirements for a Brain-like System

How can we characterize brain-like systems? The criteria listed above for neuronal systems state nothing about how plausible, if at all, the model should be in terms of the neurophysiology and neuroanatomy of the brain. If the goal is to categorize brain-like systems, we should probably add the following additional requirements:

The nodes and links should (at least, to a first approximation and without gross violations) model neurons or functional units realizable with neuron-like units and connections between them.

The topology of the graph, processing functions of the nodes, transfer functions, and learning rules should be plausible in terms of the structure and the function of the brain.

If the total system is decomposable into higher level structures (sub-graphs, columns, areas, etc.), such structures must be reasonably brain-like (to a first approximation and without gross violations, or at least not altogether implausible in terms of the physiology and anatomy of the brain).

This description of neuronal and brain-like systems is broad enough to include a fairly large class of brain like models of perception, ones that can differ significantly from each other in detail.

3. The Human Brain and Visual System - An Overview

This section presents a brief overview of the human brain and the visual system. (More extensive discussions can be found in Barlow1982, Kuffler1984, and Peters1985; A more detailed overview than the one presented here is given in Uhr1986.)

3.1. The Retina and the Geniculate

The eye's lens focuses images of the scene on the 2-dimensional retinal array of rods (that sense small changes in position and intensity) and cones (that sense color). Each of these light-activated sensors sends excitatory signals straight back to some neurons, and inhibitory signals to surrounding neurons in the local neighborhood in the adjacent layer. There are several such layers consisting of bipolar and ganglion cells, with horizontal and amacrine cells providing rich lateral linkings to greater distances. The basic excitatory and inhibitory processes appear to enhance differences and emphasize gradients, as though using a Laplacian or DOG (difference of gaussians) operator (Marr1982). The larger structures appear to serve as detectors of different types and sizes of local spots. About a million ganglion cells carry signals from the retina via one layer of synapses in the lateral geniculate to the primary visual area of the cortex, retaining the retinotopic nature of the map, much like the original image.

3.2. The Cortex

The cerebral cortex is a thin (1-2mm in thickness), crumpled, sheet containing a complex structure of six layers of densely linked neurons, under which lies a mass of cortico-cortical axons that carry signals from one part of the cortex to another. Sets of "columns" rise through all six layers of cortex (Hubel1982). Neurons sensitive to simple features such as oriented edges, colors, and textures have been identified (for e.g., simple cells - (Hubel1962)). It was originally thought that simple feature detector neurons are followed by ones that respond to more complex features in a more or less strict hierarchy (Hubel1962, Hubel1977). But recent evidence suggests more complex parallel-serial structures, rather than a simple hierarchy (Stone1979). Information flows, possibly cycling, through the layers of neurons in the cortical columns and between columns. Studies of firing patterns of neurons in columns using multi-electrode recordings (Krueger1985) have shown different but stable patterns for different stimuli; many different stimuli appear to fire each neuron, but the pattern of firings over a set of neurons differs from stimulus to stimulus.

At least 20 visual areas, and many nonvisual areas involved in perception, have been identified in the brain. Some of these handle different intrinsic scene characteristics such as motion, color and shape. Two major pathways, one processing color and shape, leading to recognition of objects, and the other handling spatial relations between objects and temporal changes (due to motion), are suggested by a large body of anatomical and physiological evidence (Mishkin1983). Similar evidence has been used to show how the 20 visual areas found to date (in macaque monkey) are wired together by about 40 major and 40 minor pathways (Van Essen1985). Note that this is far from the "complete connectivity" that is sometimes assumed to exist, which would link each of the 20 areas to all 19 others; Nor is it a simple hierarchical tree of the sort implemented in many computer vision systems.

Very few studies to date have attempted to identify neurons that respond to successively more complex structures of features found in successively larger regions of the visual field; so at present the existence of such neurons is at best a reasonable conjecture. However, high-level neurons that respond in a very specific manner to extremely complex stimuli (such as a hand, face, or even a specific face) have been found (Gross1972, Gross1984, Bruce1981, Perret1982, Rolls1984, Perret1982, Perret1984). Such neurons are not "grandmother cells" each of which somehow mysteriously recognizes some complex object, but are simply members at high levels of complex ensembles of neurons that have accumulated evidence for the patterns recognized through successive transformations of the input stimuli.

3.3. The Over-all Design of the Visual System

The extremely complex human visual system is massively parallel, shallowly serial and roughly hierarchical, with functional organizations into larger structures of neurons interconnected by pathways that help to integrate diverse sources of information. Neurons usually (but by no means always) interact with near neighbors, and organize into successively larger structures (e.g., columns, hypercolumns, areas). The brain functions effectively in extremely noisy, distorted, rapidly changing environments. It can tolerate loss of neurons due to damage and aging, and change in thresholds and levels of firing due to drugs and deprivation, indicating large amounts of built-in redundancy and self-regulating mechanisms. It is able to gain knowledge of the environment through a life-long process of learning, suggesting considerable plasticity in its structure. Our objective is to capture at least some of this complexity in an artificial system, to help realize substantial perceptual and cognitive abilities.

4. Recognition Cones - Parallel-Serial, Distributed, Brain-like Models

This section describes highly parallel, hierarchical, recognition cone perceptual structures, and examines this model in the context of the general characteristics of neuronal, brain-like systems outlined above. The emphasis is on visual perception, although some of the underlying principles are of relevance to other sensory modalities.

4.1. The Architecture of Recognition Cones

The basic building block of recognition cones is an adaptive neuron-like unit, which accepts inputs from other similar units via its input links, does some simple processing of the inputs, and sends out signals over its output links to the units into which it fires. A large number of such units are organized, roughly into a layered, hierarchically converging structures (hence the name recognition cones).

Within each layer, each unit is linked predominantly to nearby units in a relatively small surrounding neighborhood. This reflects the property of the real world that nearby points in the scene are likely to influence each other more than those that are farther apart. This observation has been exploited, with considerable success, in algorithms for computation of edges (Zucker1977), shape (Ikeuchi1981), motion (Horn1981), and stereo disparity (Marr1976). For simplicity, some regularity could be imposed on the size and shape (and to make implementations on today's computers feasible it typically is; for example, each unit can link to its 8 nearest neighbors in a square grid) of the neighborhoods. Or the connectivity patterns could model, albeit in a simplified manner, the structure of the retina, the lateral geniculate and the visual cortex (for a discussion of the primate, including the human, visual system, see Uhr1986).

The connectivity between layers is predominantly retinotopic (as in the human visual system). It is also roughly converging. Each unit is linked via input and output links to a small set of units directly above it, the number of such units usually being smaller than the neighborhood size in its own layer. It is also linked to a small set of units directly below it, the number of such units usually being somewhat larger than the neighborhood size in its own layer. This gives a logarithmically converging fan-in moving upward and diverging fan-out moving downward.

The input to the system is the image of the scene sensed by transducers (e.g., TV cameras) at the base layer of the recognition cone. This may well be a dynamic input, a new image every 30 msec or so, of the changing scene of moving objects. The total system is made of one, or possibly several, cone-like structures emerging from the retinal layer. There are several outputs from the system, typically, but not necessarily, from the higher levels. Further, there may be a rich set of additional links, e.g., for feedback loops, between the different layers (including the outputs).

4.1.1. Adaptive Neuron-like Units - A Functional Description

It was mentioned earlier that the basic building block of recognition cones is an adaptive neuron-like element. It can be treated in several alternative ways, some of which are described below.

One way to look at the adaptive neuron-like unit is to treat it as a simple finite-state automaton. It accepts inputs via its input links, computes some simple function over its inputs and its current state, updates its state, and sends outputs via output links.

The adaptive neuron-like unit can be viewed as an abstract process that computes one or more probabilistic or fuzzy transforms over its inputs. The fuzzy nature of the transforms is a consequence of the weights associated with the inputs and thresholds associated with the outputs. The output of the transforms at successive layers results in a logarithmically converging hierarchy of successively more complex abstractions of the input.

A unit typically has a small set of inputs which gather potentially relevant information (usually over a small compact window, e.g., a region large enough to extract a local feature like an edge, angle, or, at a higher level where abstracted image arrays form the inputs to the arrays of processing units - contours, enclosures, and other higher-level characteristics). The units may be designed to mimic the operation of the appropriate kinds of neurons in the retina, the lateral geniculate, and the visual cortex. For example, assume that a layer of the recognition cone has units that respond to short edges. The units in the next layer might look for longer vertical lines, among other things. Such units might look at a small neighborhood of units in the layer below, and imply a vertical line with a certain weight if some of the units in the neighborhood looked at (say three out of five) gave supporting evidence (in the form of short vertical edges). The adaptive nature of the units allows information to cycle between the layers until a stable state is reached. For example, the vertical line detectors may provide feedback that, in turn, raises the outputs of short vertical edge detectors, and so on.

Still another way to look at the operation of the adaptive neuron-like units is to treat it as the firing of "IF (conditions) THEN (implieds)" type rules (these differ from typical production rules in their use of weights and thresholds, and of spatial interrelations got from the arrays containing images and their successive abstractions).

4.1.2. Higher Level Structures

Each layer of the recognition cone may have several different kinds of units, interspersed among each other, designed to execute the appropriate fuzzy transforms to compute different intrinsic properties of the scene, such as surface shape, motion, color, etc. Each unit may have more distant links to predominantly its own kind of units (as found between hypercolumns in the primary visual cortex (Ts´o1986)), and occasionally to units of other kinds (for combining different kinds of information, say motion and shape).

Alternatively, relatively independent processing functions may be handled by separate cones (roughly analogous to the different areas of visual cortex), with sparse links between cones to combine different kinds of information. The brain seems to use both these techniques - different slope detectors are interspersed among each other, but to a large extent, motion and shape are handled in different areas.

4.1.3. Implementational Issues

The basic structure of a recognition cone is reasonably simple: Hierarchies of each-simple processes are executed on large structures (usually 2-dimensional arrays) of information, transforming information as it flows through its converging and diverging structures. The model becomes rather complex when detailed sub-structures (e.g., different types of processors and additional links among them) are specified, although it still captures only some of the complexity of the brain. Several simplifications need to be made in implementing the model in an actual program for a serial computer or, better, in silicon for a hardware-software system, using current VLSI technology.

When transforms to be performed are simple, they can be implemented by individual units. For example, a vertical edge detector is implemented by a unit that has excitatory and inhibitory inputs via links from a small neighborhood of units in the layer below, which sums its inputs and fires when a threshold is exceeded. But often, a transform may be too complex to be implemented by a single neuronal unit. It may however, be implemented by a small network of neuronal units. In such cases, the implementations may be made more efficient by increasing the complexity of individual units and links to perform such transforms.

It is useful to impose some regularity over the size and shape of neighborhoods, connectivity between layers of units, the organization into higher level structures and links between such structures, etc., to make VLSI implementations feasible. Space does not permit a detailed discussion of the various issues involved here. Several of the possible alternative implementations (both current as well as potential) have been examined in detail in (Uhr1987).

4.2. Information Flow In Recognition Cones

This section describes the flow of information in recognition cones as a consequence of their topology - a roughly retinotopic, converging-diverging network of adaptive neuron-like units.

Information is input into a sheet of photoreceptors, which serves a function analogous to that of the retina. This layer samples the continuous input from the environment in space (because of limited, although fairly high, resolution of the retina) and time (just as the photoreceptors respond to light impinging on the retina in the human visual system - once a receptor fires, it cannot fire again for a duration known as the refractory period). A whole layer of adaptive neuron-like units are excited and respond by firing in parallel (this is a simplification over the human visual system in which the firing of photoreceptors may not be completely synchronized, although at any given instant a large number of them may fire in parallel). In the process, they compute fuzzy transforms over their inputs of the kind described earlier. The firing of the units in one layer leads to the subsequent firings of units in adjacent layers, which, because of the layered hierarchical converging and diverging structure, naturally results in the computation of successively more complex, increasingly global, transformations (for, e.g., edges, long line segments, corners, contours, and so on).

The rich feedback loops that provide feedback between adjacent layers of neuronal units and, less often, between more distant layers, allow the whole process to take on the nature of a multi-level (involving several levels of abstraction) and multi-modal (involving several different kinds of information, such as shape, color, texture, motion) transformations and "relaxations" (for an example of the relaxation technique, see Zucker1977). The outputs of the units at any layer are merged with the outputs of other units computing the same or different transforms at that layer, and converged into the next layer (just as the information the retinal cones and rods sense converges into layers of primary visual cortex after several retinal transformations). Similar processing takes place continually in parallel at every layer. However, it must not be forgotten that a certain serial depth of processing (due to the layered, hierarchical structure of the system) is essential: it is this serial depth that enables the computation of successively more complex, more global, transforms of the input.

4.3. Perception in Recognition Cones

This section outlines how recognition cones can be used for visual perception - the recognition of objects in the environment - given the overall structures and processes described earlier.

Recognition cones can be given a specific structure of transforms as follows: The image is input into the retinal layer. It is processed there with smoothing (noise suppression) transforms and then by local gradient detectors, say for example, a high-resolution difference of Gaussian operator. The next layer then looks for a family of edges at several different orientations, as well as color and textural features. The next layer combines oriented edges into corners, longer lines, curves, etc; colored regions into contrast-corrected larger regions; and so on. This process of successive transformation and merging of information to detect more and more complex features can continue, possibly all the way to the top, until enough information is gathered so that specific objects are sufficiently highly implied by the features detected. In addition, continuing feedback from higher to lower layers activates processes at those layers.

Programs simulating this kind of processing, with some simplifications, have been coded and evaluated (Uhr1979, Li1986, Li1987, Uhr1987), with reasonably good and extremely fast (Li1985) performance on complex real-world objects like windows, shutters, and houses. That is, although they are highly parallel, and also neuronal and largely connectionist - albeit with additional more global brain-like structures - they have been shown able to handle complex vision problems at least as well as do computer vision systems that rely on explicit serial model-matching, and are, as a consequence, much slower.

4.4. Learning in Recognition Cones

Learning is an extremely complex phenomenon. The learning process includes the acquisition of new knowledge, development of motor, perceptual, and cognitive skills through instruction or experience, organization and integration of acquired knowledge into effective representations, and the discovery of new facts, theories, or ideas through observation, experimentation, and thought. (For discussions of various forms of machine learning, see Michalski1983, Michalski1986, Uhr1973, Fischler1987.) This section outlines how recognition cones can handle various aspects of learning.

4.4.1. Learning as Constrained Induction

A major aspect of learning involves acquiring knowledge and an understanding of the environment in which the learner-perceiver (whether natural or artificial) operates. This can be thought of as the building of usable models of the significant aspects of the environment and the important objects and other things that it contains. If the environment is sufficiently rich, that is, if it captures at least a significant portion of the complexity and the variety present in the real world, the number of possible inputs and the number of possible structures relating them is extremely large. Only a small fraction of these associations is meaningful in modeling the environment (including gaining knowledge of meaningfully related features of the environment). This suggests that the perceptual system should be designed so that it is best suited for detecting those associations that are most "useful".

In neuronal systems like recognition cones, the "structure" is determined loosely by connections between adaptive neuron-like units, within layers, from layer to layer, between cones, and so on, leaving enough room for modification by experience or training. Given a certain structure of the perceptual system (whether it is the living brain or an artificial system), knowledge of the environment is gained by a process of induction (constrained by the structure of the system) applied to the information provided by the senses. Induction is the process by which a system develops an understanding of principles or theories that are useful in dealing with the environment by generalization and specialization from specific examples or instances presented to it (Michalski1983, Holland1986).

4.4.2. Basic Neuronal Mechanisms for Learning

How can a neuronal system begin to acquire knowledge of the environment? Each unit receives excitatory or inhibitory connections from many others in its vicinity. Some authors (Feldman1982) have proposed an additional system of mutual inhibition. If this is strong enough, as soon as one unit becomes active, it will inhibit all others (alternatively, just the neighboring ones) in its layer, so that only one unit will be active in a small region of a given layer. Similar results can be realized by having the next layer simply choose the maximum over a neighborhood. It may even be preferable to postpone the choice as much as possible and make the decision a few layers above (Uhr, 1972).

The co-occurrences that occur regularly and frequently between inputs are likely to be indicative of some useful relation between them. Suppose a neuron that responds to a particular co-occurrence between its inputs is thereby sensitized to those inputs after a number of repetitions; then it will signal the event whenever it occurs. This is the basic idea behind the Hebbian learning rule (Hebb1949) which, stated simply, is: if neuron x receives input from another neuron y; then if both have sufficiently high activation, the input link from y to x must be strengthened. This is the basic neuronal mechanism of learning used in most neuronal systems.

4.4.3. The Role of Recognition Cones' Structures in Learning

The basic neuronal mechanism for learning outlined above, which changes the strengths of links with the repetition of stimulus patterns, possibly associated with feedback (either reinforcement or an error signal indicating the would-have-been correct behavior) can be used in recognition cones in any of the various ways that they are used in other connectionist or neuronal systems. (For some examples, see (Rumelhart1986a, Rumelhart1986b). Simple repetition of stimulus patterns and the resulting modification of the links gives a realization of unsupervised learning. Feedback (which could simply be an indication of whether the result of perception was right or wrong, or the correct result that should have been output) following each presentation of a pattern gives the mechanism for supervised learning. But the more global structures in the architecture of recognition cones (e.g., retinotopic mapping, layered, converging-diverging flow of information) make possible several interesting variations and additions to the basic learning mechanisms.

4.4.3.1. The Role of Retinotopic Mapping and Spatial Contiguity in Learning

Typically, neighboring points in the visual field are correlated, exhibiting dependencies on one another (spatio-temporal contiguity - a classic idea), since they often belong to parts of a single object. In general, each physical object is enclosed in a surrounding skin, and separate objects interact more intensely the closer they are. Perception involves recognition of physical objects from images that reflect this fact. To learn about an object, it is useful to discover associations between its sub-parts; and these sub-parts, because of spatial contiguity, are likely to be imaged into neighboring parts of the visual field.

Topographical mapping of the visual field is a prominent feature of recognition cones - just as it is of the visual cortex. The effect of this mapping is to cause neighboring parts of the visual field to project to neighboring parts of recognition cone layers. This topographical mapping of the visual field preserves the spatial relations between different parts of the scene, and the connectivity patterns (each unit is connected predominantly to others in a small neighborhood directly below it) favor discovery of useful associations between neighboring parts of a scene. So recognition cones are predisposed, by their very structure, towards discovering associations between neighboring parts of the visual field, and hence towards learning associations between sub-parts of an object. For example, if the object imaged is a chair, its sub-parts (e.g.; legs, seat, back-rest) are projected onto the neighboring parts of the visual field with spatial relations between them intact. Since units receive inputs from other units in their vicinity, and since learning is effected by strengthening the links that fan into active units, this facilitates strengthening of links that signal the spatial relations between sub-parts of the chair. It will be shown later that this does not preclude the learning of more global relations.

4.4.3.2. The Role of Hierarchical, Converging-Diverging Structure in Learning

In recognition cones, at each level, the learning of associations can be effected by the adaptive neuron-like units, working in parallel. And since each layer feeds the next, the learning of structure is a hierarchical, repeated (in the layers as well as over several presentations of the patterns) operation. (The same appears to be true for the brain, especially the cortex). For example, the lower layers may learn the associations between several vertical edges more or less aligned with each other and thus discover (learn) the concept of a long vertical line. At higher levels, a vertical line and a horizontal line that intersect may facilitate the learning of the more complex concept of a corner, and so on. The effects of this are two-fold: Simpler concepts are learned before more complex concepts. Successively more global relations are learned at successively higher layers.

For perception, this model supports the capacity for discovering the defining transforms over a set of patterns essentially by a process of induction that involves generalization whenever possible and specialization whenever needed. One type of generalization is the sharing of several transforms by the internal representations of different objects that the system learns to recognize. For example, a chair and a stool both have legs and seat and hence they can share several transforms that enable the system to recognize these parts. Specialization is initiated by the addition of new transforms when the existing set proves inadequate for the perceptual tasks that the system must perform. For example, additional transforms that allow the system to recognize the back-rest are needed if it is to differentiate between chairs and stools. This is explained in more detail in the next section.

4.4.4. The Learning of Useful Transforms

In systems like recognition cones, which rely on the application of a large number of transforms to the input, successively abstracting, learning by induction can be viewed as the development and retention of a set of transforms that are adequate for the perceptual tasks demanded of the system. The system can learn, or be initialized with (as though by evolution) a set of low-level transforms such as edge detectors, color detectors, etc., to which new transforms are added by combining several of the existing transforms. The layered, converging- diverging structure provides the framework for this process, by determining the space of possible transforms that can be generated (since the links that are allowed, either a-priori or through learning, determine the various ways in which simpler transforms can be combined into more complex transforms). This is examined in more detail in the next section.

Transforms can be modified by changing weights of implieds (output links of the adaptive neuron-like units) and conditionals (input links of the units); and by changing thresholds of firing (output functions of the units). New transforms can be added by growing new links between units - implieds, conditionals, with appropriate weights, which are themselves learned; or by recruiting new units from the available pool of units. These mechanisms will be examined in some detail in what follows.

4.4.4.1. Modification of Existing Transforms

When available, feedback can be used to strengthen the links of the transforms that implied the "right" thing (or to weaken those that implied the "wrong" thing), by propagating the information from the unit or units that made the choice, usually moving from the apex of the cones backward through the network, possibly down to the lowest layer. One way to achieve this is by release of a "transform reinforcement signal" by units that received positive feedback, which would strengthen the input links feeding them that have positive weights associated with them and trigger the release of the reinforcement signals in those units from which those input links originated. For example, if the system recognizes a square correctly, and if the square was implied by a set of units whose outputs implied the four corners, then the links from those units that feed into the units that recognized a square will be strengthened. Similarly, those units that led to the wrong output and therefore received negative feedback would have links weakened.

4.4.4.2. Creation of New Transforms

Feedback can also be used for the development of new transforms. Suppose the feedback indicates that the system implied the "wrong" thing. This can be used to cause the release of a "transform creation signal" by units that received negative feedback, which would be transmitted to every unit in the lower layer which contributed inputs to the units in question. This can continue, possibly all the way down to the lowest layer of transforms. At one or more, possibly several layers, some random subset of the units receiving the transform creation signal would "recruit" one or more units from the next higher layer by growing a link to that unit. Growth of new links such as these would take place without violating the topological constraints of predominantly near neighbor connectivity, retinotopy, convergence and organization into higher level structures mentioned earlier. The effect of this is to add new transforms to the existing set of transforms at those layers. The transforms so added would participate in the learning process according to the same principles as those described above.

4.4.4.3. Regulatory Mechanisms to Control the Development of Transforms

Some sort of balance is needed between the modification of existing transforms and the creation of new ones. If the former predominates, the system can get bogged down in a situation where no amount of reweightings will enable it to perform adequately. If the latter predominates, the system may accumulate a large number of transforms most of which have not proven their usefulness. One way to maintain the required balance is to regulate the frequency at which a particular unit receiving a transform creation signal can respond to it. The dual mechanisms of reinforcement of "good" transforms and the creation of new transforms explained above, over a period of time, result in the development and retention of a set of fuzzy transforms that are useful for recognizing the objects in the environment. On the other hand, transforms found not useful fade away, since they are negatively reinforced (or just left undisturbed) by the learning mechanism.

4.4.5. A Summarizing Look at Learning in Recognition Cones

As pointed out earler, the structure of a neuronal system determines the possible transforms that can be computed efficiently. Since the structure of recognition cones is suggested by the spatio-temporal contiguity of objects in the environment (as well as the structure of the brain, which functions so effectively in such environments), it appears that the structure of recognition cones favorably constrains the development and retention of transforms, and hence the acquisition of perceptual skills necessary to recognize the objects in the environment. This model puts into the learning process relatively more useful structure than is typically found in many connectionist systems (which often reduces learning to an essentially random search through the space of all possible correlations among all possible stimuli and responses).

4.5. Distributed Processing and Control in Recognition Cones

Recognition cones obviously process large amounts of information in parallel. In this section, we show that they embody distributed processing of information, and examine some properties that emerge as a result. An information processing system has three aspects, namely processing, memory, and control, each of which may be "distributed" to different degrees, in several different ways, over the system. We will not get into the details of the alternative ways in which this can be done here, but simply point out how processing, memory, and control are distributed in recognition cones, and the desirable properties that result from it.

4.5.1. Distributed Processing in Recognition Cones

First, each of the many micro-modular units computes only a tiny part of the large and complex function that perceives - so processing is clearly distributed. The activation of each unit is a function of the inputs from several other units, and each unit contributes to the activation of several units. At a more abstract level, this can be viewed as recognition of a pattern being the result of utilizing knowledge distributed over several successively more complex feature detectors. Each 'feature' is a pattern of activation, whose 'meaning' becomes evident to processes (often but not exclusively at the layers above) that immediately or eventually receive their outputs. Note that the activation of a single unit at any layer does not convey much information; it is the collective contribution of a pattern of activity that is meaningful. Further, each thing (feature, sub-object, etc.) is implied by several different transforms and each transform implies several different things. Thus, processing functions are distributed over a large number of units and fuzzy transforms.

4.5.2. Distributed Memory in Recognition Cones

Memory is distributed over the large number of units (and each unit's output functions, thresholds, activation levels, etc.) and the links (the weights and learning rules associated with them).

4.5.3. Distributed Control in Recognition Cones

The control or supervisory role is shared by all the units involved. Each unit, by virtue of its connectivity, 'knows' where its inputs come from and where its outputs should go. In other words, whether or not a unit fires is determined by its inputs, which are themselves the outputs of several other units, and so on. In such a system, the firing of a unit or a set of units (performing a transform of the kind described earlier) is the analog of execution of an instruction in a conventional computing system. Control - that is, the decision to execute one instruction (or apply a transform) or another - in a neuronal system is locally but collectively exerted by individual units providing the inputs for other units. The control exerted by a unit is immediately felt in a small neighborhood around it; but the control over the behavior of the entire system emerges from all the units exerting control over units in their neighborhoods.

Different transforms are executed depending on the units that get fired as a function of the input to the retinal layer. Spatial relations between parts of a scene are "recognized" by virtue of the links and the location of sub-parts of a pattern as determined by the patterns of activation. (Contrast this with the situation in which a global controller - a single processor or the logic in the program being executed on a conventional computer instructs the system to look for specific sub-parts in the image and, depending on the success of failure of the search, to execute different sections of the program).

Some control is exerted by the layered converging-diverging structure of the recognition cones. Since the direction of information flow determines what transforms get applied, the control exerted by the structure of the system itself is significant (note, however, that there are no global controllers). Thus, control is distributed over the system. There is no need for a global supervisor or executive, since the function of the executive is collectively assumed by the adaptive neuron-like units and the overall structure of the system.

4.5.4. Noise Immunity and Graceful Degradation Under Damage

As pointed out earlier, recognition cones work essentially by means of a large number of fuzzy transforms that operate over fields of information (typically relatively small windows of units surrounding each unit). To function effectively in the real world (as the human visual system clearly does), they must incorporate sufficient noise immunity as well as redundancy. Thus the transforms must be executed by a set of neurons in a relatively imprecise manner. In recognition cones, this is accomplished by the fuzzy nature of the transforms. Also, larger windows and samplings can be used, with thresholds that accept a variety of different patterns of firings. Further, one can weight the certainty of the implied feature by means of weights on the links to the units that are fed by the output of the unit detecting the feature. In addition, there can be several units that are looking for the same feature (say long vertical line) in approximately the same neighborhood, using several different strategies. Each transform is a relatively simple, hence weak, process; therefore, many transforms are needed. For example, to detect vertical edges, one might well use several different edge detectors, and combine all their results; to detect angles one might use several different detectors, one concentrating on properly interrelated edges, another on the vertex, and others on various aspects of the interior and exterior regions. Thus the different but closely related processes that are serving overlapping purposes provide the redundancy that is needed to ensure robustness (graceful degradation under damage - failure of a few random units or loss of a few random links, analogous to the death of a few neurons) as well as noise immunity.

4.5.5. Automatic Generalization in Recognition Cones

Automatic generalization is an emergent property of the representation - similar things (objects, concepts, features, or whatever) have overlapping patterns of activation. This is again a consequence of the distributed nature of the model. That, combined with its hierarchical, roughly converging organization, leads to efficient representations of large numbers of patterns. In other words, information is abstracted into successively more compact representations by discarding and generalizing over unessential details - by the inductive learning process outlined earlier.

4.5.6. Top-down Processing Power in a Data-driven Architecture

The recognition cone model clearly incorporates data-driven or bottom up processing of information. It may not be equally obvious that it simultaneously embodies top down (model based, goal directed) processing also. When a thing (feature, object, or whatever) is implied, it (or the collection of units representing it) can back-fire so that the lower levels can apply other transforms (that imply the same thing), thus accumulating more evidence for the implied thing. The activation of a unit can activate units at lower levels that are connected to it, and this effect may percolate all the way down to the lowest layer. In other words, when there is partial evidence for a certain feature, this can cause the system to gather more evidence for that feature. All this is achieved without the need for explicit models or explicit control, as a consequence of the distributed nature of processing and control.

5. Summary and Conclusions

Real world Perception and Cognition are extremely complex phenomena that require adequately complex systems. The brain is a living existence-proof of such capabilities, and hence a rich source of suggestions for the required structures and mechanisms. Neuronal, brain-like artificial systems attempt to capture some of the complexity of the brain, aiming to achieve some of the perceptual and cognitive abilities of the brain in the process. The recognition cones described in this paper incorporate some of the larger mechanisms and structures of the brain, as well as neuron-like basic units. Many of the capabilities of recognition cones emerge from the use of brain-like structures.

The use of adaptive neuron-like units structured into an over-all organization suggested by the brain, albeit with several simplifications, endows recognition cones with fairly powerful capacity for perception. Perception takes place by means of a multi-level, multi-modal, relaxation process of detecting successively more abstract and more global characteristics, utilizing a large number of relatively simple, and hence weak, fuzzy transforms that extract, abstract, combine, and integrate information from diverse sources. The structure of recognition cones constrains the learning process in such a way that the development of useful new associations and structures is facilitated.

Recognition cones are highly parallel, shallowly serial, roughly hierarchical structures that embody distributed processing of large amounts of information. Out of the distributed nature of the system arise the desirable properties of noise immunity, robustness, and flexibility in handling variations. Thus, recognition cones appear to usefully incorporate high-level brain-like structures in neuronal models.

References

Amari1977.
Amari, S. A., ''A mathematical approach to neural systems,'' in *Systems Neuroscience*, ed. Metzler, J., pp. 67-117, Academic Press, New York, 1977.

Anderson1977.
Anderson, J. A., Silverstein, J. W., Ritz, S. A., and Jones, R. S., ''Distinctive features, categorical perception, and probability learning,'' *Psychological Review*, vol. 84(5), pp. 413-451, September 1977.

Arbib1979.
Arbib, M. A., ''Perceptual structures and distributed motor control,'' *COINS Technical Report 79-11*, University of Machusetts, Computer and Information Science, and Center for Systems Neuroscience, Amherst, Massachusetts, June 1979.

Ballard1984.
Ballard, D. H., ''Parameter networks - Towards a theory of low-level vision,'' *Artificial Intelligence*, vol. 22, pp. 235-267, 1984.

Barlow1982.
Barlow, H. B. and Mollon, J. D. (eds.), *The Senses*, Cambridge University Press, New York, 1982.

Barnden1986.
Barnden, J., ''Complex cognitive information processing: a computational architecture with a connectionist implementation,'' *Technical Report 211*, Department of Computer Science, Indiana University, Bloomington, Indiana, December, 1986.

Bruce1981.
Bruce, C. J., Desimone, R., and Gross, C. G., ''Visual properties of neurons in a polysensory area in

the superior temporal sulcus of the macaque,'' *Journal of Neurophysiology*, vol. 46, pp. 369-384, 1981.

Davis1981.
Davis, L. S. and Rosenfeld, A. R., ''Cooperating processes for low-level vision: a survey,'' *Artificial Intelligence*, vol. 17, pp. 245-263, 1981.

Fahlman1983.
Fahlman, S. E., Hinton, G. E., and Sejnowski, T. J., ''Massively parallel architectures for AI: NETL, Thistle, and Boltzmann machines,'' *Proceedings of the National Conference on Artificial Intelligence AAAI-83*, 1983.

Feldman1982.
Feldman, J. A. and Ballard, D. H., ''Connectionist models and their properties,'' *Cognitive Science*, vol. 6, pp. 205-254, 1982.

Fischler1987.
Fischler, M. A. and Firschein, O., *Intelligence - The Eye, the Brain, and the Computer*, Addison-Wesley, Reading, Massachusetts, 1987.

Fukushima1975.
Fukushima, K., ''Cognitron: A self-organizing multi-layered neural network,'' *Biological Cybernetics*, vol. 20, pp. 121-136, 1975.

Gross1972.
Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B., ''Visual receptive fields of neurons in the inferotemporal cortex of the macaque,'' *Journal of Neurophysiology*, vol. 35, pp. 96-111, 1972.

Gross1984.
Gross, C. G., Desimone, R., Albright, T. D., and Schwartz, E. L., ''Inferior temporal cortex as a visual integration area,'' in *Cortical Integration*, ed. C. Ajmone-Marsan, Raven, New York, 1984.

Grossberg1980.
Grossberg, S., ''How does the brain build a cognitive code?,'' *Psychological Review*, vol. 87, pp. 1-51, 1980.

Hanson1978.
Hanson, A. R. and Riseman, E. M. (eds.), *Computer Vision Systems*, Academic Press, New York, 1978.

Hanson1987.
Hanson, S. J. and Burr, D. J., *Knowledge representation in connectionist networks*, Bell Communications Research, Morristown, New Jersey, February 1987.

Hebb1949.
Hebb, D. O., *The Organization of Behavior*, Wiley, New York, 1949.

Holland1986.
Holland, J. H., Holyoak, K. J., Nisbett, R. E., and Thagard, P. R., *Induction: Processes of Inference, Learning, and Discovery*, The MIT Press, Cambridge, Massachusetts, 1986.

Hopfield1982.
Hopfield, J. J., ''Neural networks and physical systems with emergent collective computational abilities,'' *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 79, pp. 3088-3092, 1982.

Horn1981.
Horn, B. K. P. and Schunck, B. G., ''Determining optical flow,'' *Artificial Intelligence*, vol. 17, pp. 185-203, August 1981.

Hubel1962.
Hubel, D. H. and Wiesel, T. N., ''Receptive fields, binocular interaction and functional architecture in the cat's visual cortex,'' *Journal of Physiology*, vol. 160, pp. 106-154, 1962.

Hubel1977.
Hubel, D. H. and Wiesel, T. N., ''Ferrier lecture: Functional architecture of the macaque monkey visual cortex,'' *Philosophical Transactions of the Royal Society, London (Biology)*, vol. 198, pp. 1-

59, 1977.

Hubel1982.
Hubel, D. H., "Explorations of the primary visual cortex, 1955-1978," *Nature*, vol. 299, pp. 515-524, 1982.

Ikeuchi1981.
Ikeuchi, K. and Horn, B. K. P., "Numerical shape from shading and occluding boundaries," *Artificial Intelligence*, vol. 17, pp. 141-184, August 1981.

Kohonen1977.
Kohonen, T., *Associative memory: A system theoretical approach*, Springer, New York, 1977.

Kohonen1984.
Kohonen, T., *Self-organization and associative memory*, Springer-Verlag, Berlin, 1984.

Krueger1985.
Krueger, J., "Investigation of a small volume of neocortex with multiple microelectrodes: Evidence for principles of self-organization," in *Complex Systems - Operational Approaches*, ed. H. Haken , pp. 71-80, Springer-Verlag, Berlin, 1985.

Kuffler1984.
Kuffler, S. W., Nicholls, J. G., and Martin, A. R., *From Neuron To Brain*, Sinauer Associates Inc., Sunderland, Massachusetts, 1984.

Li1985.
Li, Z. N. and Uhr, L., "Comparative timings for a neuron recognition program on serial and pyramid computers," *Proceedings of Conference on Computer Architectures for Pattern Analysis and Image Data Base Management*, pp. 99-106, IEEE Computer Society Press, 1985.

Li1986.
Li, Z. N. and Uhr, L., "A pyramidal approach for the recognition of neurons using key features," *Pattern Recognition*, vol. 19, pp. 55-62, 1986.

Li1987.
Li, Z. N. and Uhr, L., "Pyramidal algorithms for analysis of house images," *Systems, Man and Cybernetics*, 1987.

Marr1982.
Marr, D., *Vision*, W. H. Freeman and Company, New York, 1982.

Marr1976.
Marr, D. C. and Poggio, T., "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283-287, 1976.

Michalski1983.
Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.), *Machine Learning - An Artificial Intelligence Approach, Vol. 1*, Tioga , Palo Alto, California, 1983.

Michalski1986.
Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.), *Machine Learning - An Artificial Intelligence Approach, Vol. 1*, Tioga, Palo Alto, California, 1986.

Mishkin1983.
Mishkin, M., Ungerleider, L. G., and Macko, K. A., "Object vision and spatial vision: Two cortical pathways," *Trends in Neuroscience*, vol. 6, pp. 414-417, 1983.

Perret1982.
Perret, D. L., Rolls, E. T., and Caan, W., "Visual neurones responsive to faces in the monkey temporal cortex," *Experimental Brain Research*, vol. 47, pp. 329-342, 1982.

Perret1984.
Perret, D. L., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M. A., "Neurones responsive to faces in temporal cortex: studies of functional organization, sensitivity to identity and relation to perception," *Human Neurobiology*, vol. 3, pp. 197-208, 1984.

Peters1985.
Peters, A. and Jones, E. G. (eds.), *Cerebral Cortex: Vol. 3. Visual Cortex*, Plenum, New York, 1985.

Ratcliff1978.
Ratcliff, R. A., "A theory of memory retrieval," *Psychological Review*, vol. 85(2), pp. 59-108, March 1978.

Rolls1984.
Rolls, E. T., "Neurons in the cortex of the temporal lobe and in the amygdala of monkey with responses selective for faces," *Human Neurobiology*, vol. 3, pp. 209-222, 1984.

Rumelhart1986.
Rumelhart, D. E., Hinton, G. E., and McClelland, J. L., "A general framework for parallel distributed processing," in *Parallel Distributed Processing vol. 1: Explorations into the Microstructure of Cognition*, The MIT Press, Cambridge, Massachusetts, 1986.

Rumelhart1986a.
Rumelhart, D. E. and Zipser, D., "Feature discovery by competitive learning," in *Parallel Distributed Processing vol. 1: Explorations into the Microstructure of Cognition*, The MIT Press, Cambridge, Massachusetts, 1986.

Rumelhart1986b.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning internal representations by error propagation," in *Parallel Distributed Processing vol. 1: Explorations into the Microstructure of Cognition*, The MIT Press, Cambridge, Massachusetts, 1986.

Schaefer1985.
Schaefer, D. H., "A pyramid of MPP processing elements - experiences and plans," *Proceedings of the 18th International Conference on System Sciences*, Honolulu, 1985.

Smolensky1984.
Smolensky, P., "The mathematical role of self-consistency in parallel computation," *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, 1984.

Stone1979.
Stone, J., Dreher, B., and Leventhal, A., "Hierarchical and parallel mechanisms in the organization of visual cortex," *Brain Research Reviews*, vol. 180, pp. 345-394, 1979.

Tanimoto1980.
Tanimoto, S. L. and Klinger, A. (eds.), *Structured Computer Vision: Machine Perception Through Hierarchical Computation Structures*, Academic Press, New York, 1980.

Ts´o1986.
Ts´o, D. Y., Gilbert, C. D., and Wiesel, T. N., "Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis," *Journal of Neuroscience*, vol. 6, pp. 1160-1170, 1986.

Uhr1972.
Uhr, L., "Layered recognition cone networks that preprocess, classify, and describe," *IEEE Transactions on Computers*, vol. 21, pp. 758-768, 1972.

Uhr1973.
Uhr, L., in *Pattern Recognition, Learning and Thought*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

Uhr1979.
Uhr, L. and Douglass, R., "A parallel-serial recognition cone system for perception," *Pattern Recognition*, vol. 11, pp. 29-40, 1979.

Uhr1986.
Uhr, L., "Toward a computational information processing model of object perception," *Computer Sciences Technical Report #651*, Computer Sciences Department, University of Wisconsin-Madison, Madison, Wisconsin, July 1986.

Uhr1987.

Uhr, L., ''Highly parallel, hierarchical, recognition cone perceptual structures,'' in *Parallel Computer Vision*, ed. L. Uhr , Academic Press, New York, 1987.

Van Essen1985.

Van Essen, D. C., ''Functional organization of primate visual cortex,'' in *Cerebral Cortex: Vol. 3. Visual Cortex*, ed. A. Peters and E. G. Jones , pp. 259-329, Plenum, New York, 1985.

Zucker1977.

Zucker, S. W., Hummel, R. A., and Rosenfeld, A. R., ''An application of relaxation labeling to line and curve enhancement,'' *IEEE Transactions on Computers*, vol. C-26, pp. 394-403, April 1977.