Chapter 1: Introduction

Classifying object shapes and identifying the category of a known or previously unseen shape most similar to a prototype are fundamental tasks in computer vision. Successful solution of these tasks is a prerequisite for higher levels of object and scene analysis in still images, and for object tracking in video sequences. Research into the problem is increasingly motivated by new applications in content-based retrieval for image, trademark, and video databases, in addition to traditional industrial inspection, biomedical, and target identification applications. In multimedia applications, it is often particularly important to conform to the human judgments of similarity, and to have a representation which can be matched with efficient computations.

While such practical interests are one driving force, the questions of similarity and object identification are of intrinsic interest from a psychological and neurophysiological standpoint as well; distinct perspectives on these problems have evolved within these disciplines, and will be given considerable space here. Computational and signal processing concepts have often informed these fields, framing the nature of experiments and interpretations of results.

In this thesis, the related problems of similarity and stimulus equivalence are addressed through a method involving the construction of a metric space by partial synchronization of states, under the twin influence of the intrinsic dynamics of the oscillators and of the shapes presented as initial conditions. The current form of the system is a hybrid of algorithmic processing for learning and recognition, with the construction of the space and the resulting recognition algorithm using a regular lattice of chaotic oscillating units with a local mean field coupling.

MARR'S PROGRAM IN VISUAL NEUROSCIENCE

Because the visual system has been studied most intensely, the mutual influence of theory and experiment in that realm has dominated thinking about neural mechanisms and sensory problem solving. Marr described three levels [1] of abstraction for the study of vision (Marr 1982) . The lowest level is the **mechanism** level, concerned with details of what the physical elements of the brain (or computer are doing). Historically, neurons have been taken as the elementary units of interest. The next level is **algorithmic**: a description of the process controlling the hardware. Finally, the highest level is the **computational** level, in which the problem should be understood in terms of information processing: what is being computed, why, and what are the appropriate models for these operations?

Marr produced such a computational theory for object recognition, in which information is reorganized by flowing through a series of parallel and serial modules or processing stages. The raw, two dimensional array of image intensity first goes through a

[1] The outline of Marr's work here follows closely the synopsis in (Yuille and Ullman 1990).

process of edge-detection to produce an intermediate representation known as the *primal sketch*. This representation undergoes further processing to produce a viewer centered representation (the $2\frac{1}{2}-\mathrm{D}$ sketch), Finally, an objected centered 3-D representation is computed, describing the objects in terms of volumetric primitives. Low level vision is concerned solely with the first two levels; interpreting and recognizing an object from a sketch is generally considered as high level vision. The notion of object centered representations has proven controversial; I will examine the controversy in some detail. In this thesis, I have adopted the *viewer-centered* approach.

Because the Marr program is especially influential and successful in replicating the types of function thought to be performed in *low level* vision, many researchers have tried to extend it to high level vision, but chiefly using the assumptions concerning *mechanisms* adopted from low level vision. These mechanisms are chiefly a "front end" of filters employing rate coding, with "back ends" consisting of specific arrangements of excitatory and inhibitory neurons, doing the work of the algorithms within a computational module. While the general framework remains valid, the role of detailed structural arrangements of individual neurons will be questioned in this thesis.

I will now present an overview of the problem - the computational analysis - and a brief treatment of the mechanism or implementation level.

SITUATING SIMILARITY AND OBJECT RECOGNITION

The over-arching problem chosen for study here is similarity, with a particular focus on similarity of curves and silhouettes. By silhouettes, I mean two dimensional projections of three dimensional objects, with all shading data removed. These binary images are essentially closed curves in the plane, with the enclosed region filled in. Considering curves, the similarity problem can be simply but generally formulated as:

Given planar, non-self-intersecting closed curves C1 and C2:

Define a distance function d(C1,C2) which returns a scalar value indicating their distance in some high dimensional space.

This formulation allows for geometric approaches relying on knowledge of explicit coordinates, as well as approaches which operate on the instantiation of the curves in an image. These approaches will vary according to various criteria for effectiveness, for sensitivity to noise and distortions, and with regard to plausibility as biological models.

In both machine vision and the psychological study of similarity, it has been common to construct a *metric space* and to embed objects as points in the space. The distance function defined above can then be computed as standard Euclidean distance, or by a variety of more sophisticated weighted distance functions. Non-metric methods of assessing similarity exist as well, and will be briefly mentioned; these emerged chiefly in psychology, in response to evidence that many aspects of similarity in cognitive and perceptual phenomena may not be metric.

Similarity is seen by cognitive science and psychology of vision as tightly linked to segregation and grouping processes in visual scene analysis, to identifying objects seen from different viewpoints, and to categorization. Thus, it underlies nearly all higher level cognitive phenomena regarding our processing of the visual world. In everyday life we spontaneously identify objects, and create categories from the diverse retinal images of an object seen from different viewpoints, or from the diverse individual members of a species. However, this fundamental problem of stimulus identity has only recently begun to be addressed satisfactorily in computer vision. While many geometric methods have been developed for handling translation and rotation in the plane ((Simoncelli, Freeman et al. 1992); ((Wolfson and Yehezkel 1992), rotation *in depth* of diverse objects has been addressed most recently and successfully through statistical approaches with a rich feature space (Mel 1997), and by an ensemble of radial basis function neural networks implementing a view normalization and interpolation strategy (Edelman 1999).

RESTRICTIONS ON THE PROBLEM SPACE

For exploring, the power of a class of dynamical networks on problems of similarity and stimulus invariance, I have chosen data limited to isolated outline objects (silhouettes), i.e. binary images in the plane. I do not address more general problems of segmentation and scene analysis; clearly extensions to this method and extensive preprocessing would be required to account for those aspects of visual processing.

The approaches just mentioned developed by Edelman and Mel for object recognition also use isolated objects, but include shape, shading and (in Mel's SEEMORE system) color. These additional cues may increase the performance of those algorithms, which have not been tested on silhouettes. Humans clearly can recognize common objects from outline information only (Hayward 1998), and the focus in this thesis has been on that domain.

Nevertheless, including shading data would allow discrimination of certain objects from views which occlude (hide) parts of their own structure. With silhouettes only, it would be impossible to discriminate a round ball from a clarinet seen looking directly down the bell, for example. Shading information would allow the convex and concave nature of the two objects to be discriminated. I do attempt to handle objects which present radically different views; the following figure illustrates an extreme view of one of the test objects in a family of recognition experiments.

Fig. 1. Two views of the same paperclip object, illustrating the extreme nature of the distortion due to rotation in depth. The left view is the 0°, the right view is +90°. Base images provided by Michael Tarr, Psychology Dept., Brown University.

In many object recognition studies, the performance of images degraded with noise and with distortions such as scrambling or occlusion is examined.
Examining performance in the presence of image noise is particularly important for a complete scene analysis system attempting noise sensitive processes such as segmentation.

At the time of writing, the approach here has not been tested with noise degradation, and I will assume segmentation is handled by pre-processing which must be noise tolerant and ideally has some noise suppression ability.

An Overview of the Methods

The use of a metric representation on a space of extracted features, or on higher-order features discovered by a learning process, is one of the oldest methods in machine vision, and in pattern recognition in general. The method developed here is novel in that the dimensions of the space are derived from phase space partition cells of a dynamical system[2]. Stated differently, the representation or encoding is based on the *statistics* of the *total network state* in the constructed space, without reference to the individual nodes in the network. This is in contrast to most connectionist representation spaces, where the dimensions of the space capturing the statistics of the modeled world are bound to individual nodes, whether these are internal or output nodes whose values constitute a code. The linkage of representation dimensions to nodes is referred to as *localist* or *place coding* and remains the most widespread assumption in neuroscience and neural network theory.

---

[2] This is translated roughly as non-overlapping intervals in the set of possible real valued states, with each node in a network having a state in only one such interval.

The present system and network model assumes an alternative spatio-temporal population code, and achieves some functional computing properties which help to overcome classical dilemmas in pattern recognition. Issues of combining local features (binding), capturing structural relationships at slightly larger scales, and handling view invariance can be addressed simultaneously by the spatio-temporal interactions that occur during learning and recognition in this system.

In addition, the work raises principled issues of computation in biological systems. While rooted in theory of recurrent networks, chaos, and complexity (high dimensional coupled chaotic systems), it makes some substantive departures from previous work in these fields. The most important is that it introduces two stages with a sharp change in the parameters, rather than stationary or smoothly changing dynamics. By some definitions, this precludes its consideration as a dynamical system altogether and certainly makes any analysis based on continuous mathematics more difficult.

The present work also breaks with typical practice in recurrent networks by concentrating on the *transient* regime of dynamics rather than on equilibrium states (i.e. attractors) of the network. In other words, the network is measured and finishes its work *prior* to reaching any stable asymptotic state. This staged processing, with desynchronization and partial synchronization of transient trajectories is designated as a *Synchronization Opponent Cooperative Activity* (**Soca**) network.

To date the study of high dimensional, spatially coupled nonlinear systems is chiefly experimental, with the experiments conducted by numerical simulation. Given this fact, there is little existing theory to build rigorous proofs of the system's capability, bounds on performance, or expectations on memory capacity and scaling. The most relevant recent theoretical developments are presented briefly, though none are directly applicable in their present form. The present study, like the bulk of spatially-extended network studies, is exclusively computational in nature.

OPERATIONS AND ARCHITECTURES FOR COOPERATIVE NEURAL COMPUTATION: AN OVERVIEW

The application of the Turing scheme to describe neural computation in a real brain is not completely obvious … Where is the program in the brain? And what is a memory? If a program exists, its mere definition will, in my view, be a revolutionary step toward the understanding of brain function. The mere demonstration of the existence of a program is beyond what seems imaginable. And all we have are noisy neurons and unreliable synapses. (Amit 1995)

For the purposes of this introduction, let us note two observable properties of a *chaotic* system. It is a deterministic system charactized by some scalar or vector state *S* which

1. exhibits an aperiodic time evolution (called its orbit or trajectory)
2. exhibits exponentially rapid divergence over time for nearby initial conditions.

We can model such systems with differential equations or difference equations. Using difference equations, chaotic behavior can be generated with a one dimensional system – a system with a single state variable. Such a system, with a single time varying state variable, can be considered as a simple model of an *oscillator*.

If we want to compute similarity of inputs based on the state values of a system observed at some later time *t*, chaos seems to be the very opposite of what we want. Only early in the evolution of a chaotic system, in the *transient* stage, is there reliable correlation between the input and the state of the evolving system.

We can construct *higher dimensional* systems by connecting two or more such systems as nodes in a network. At every time step, an averaging function results in mutual influence of the state values, counteracting the divergent tendencies of chaotic dynamics. This coupled high dimensional system behaves quite differently; depending on the "strength" of chaos and coupling, the connected oscillators may *synchronize* in an aperiodic or periodic mode.

In this case, there are rather intricate dependencies in the approach to synchronization on the specific spatial form of the input to the system. The *basic intuition* underlying this thesis was the idea that by exploiting these two opposing tendencies, of divergence and synchronization, a high dimensional system may effectively compute a representation suitable for use as a distance function.

Since an important goal of this thesis is to extend and solidify the connections between high dimensional nonlinear dynamics, computational approaches to pattern recognition, and biological networks, it is necessary to understand developments in several fields which have motivated the approach taken. The relationship between arrays of chaotic network elements, programs and neurons would seem even less obvious than the relationship of programs and neurons.

I will argue that a more diverse, and ultimately clearer, picture of neural computation is emerging from theoretical areas such as synchronization dynamics of ensembles of coupled chaotic elements, and the field of *symbolic dynamics* which bridges classical computing concepts (symbols and formal languages) with dynamical systems theory. I will also argue that a wealth of experimental evidence in neurophysiology supports this view, even though (with a few exceptions) this has not been the interpretation framework used by experimentalists working on synchronization dynamics [3].

This style of computation differs from the classical connectionist models, including attractor neural networks, in at least one significant way. *The structure supporting the computation is implicit in the graph corresponding to the flow of states, when a particular partitioning (coarse graining) of the state space of network elements is chosen*. This state flow graph must adapt to both the statistics of the sampled world of patterns, and must also support a cooperative transformation of the input, such as the normalization function developed here in the context of solving the stimulus equivalence

[3] There is some evidence that this is changing. W. Singer, a well known investigator of synchronized oscillations, stresses that *aperiodic* oscillations may be synchronized in a recent review. The work of Hampel and Sompelinsky, reviewed later, is an outstanding exception.

version of the similarity problem. Some similarity of this computational structure to Markov chains and decision processes is apparent, but again the present emphasis on *changes* in the dynamical parameters precludes an obvious mapping of techniques from that field, which chiefly models random processes with stationary probability transitions between states.

It could be argued that the statement highlighted above is too strong; recurrent network architectures *in general* do their work through complex state flows controlled by weights. However, these still depend on fine structure of weights and specific connection topologies, while the networks studied here do their work with homogenous local connections between units in a single layer.

With this iterative computation style and corresponding implicit state flow graphs, networks with few (i.e. 6) parameters yield effective computations for a perceptual task. These parameters are spatially homogenous across a spatially regular, locally connected network, rather than requiring a specific topology of connections and weights fed by feature detectors. These parameters correspond to *average parameters* of micro-circuit component systems tolerating a great deal of randomness in structure and noise in their local, neuron level operation, because computation is carried out by hierarchies of oscillating sub-populations.

This is not to say that classical ideas of receptor fields, and of neural coding and computation mediated by single cell rates are unimportant, as they are relatively simple and predict the responses of superficial neurons to simple stimuli. However, in the emerging view, these classical concepts interface with more dynamically complex computational systems, functioning in part by rapid coordination between cortical regions and between cortical and sub-cortical areas. In sensory pathways, these simplest computational stages may act as *perturbations* to higher level computational systems supporting segmentation, object recognition, and attended search functions. In motor systems they serve as final actuators to the muscular systems, driven by complex dynamical pattern processes in pre-motor systems.

Still, even the classical concepts are being questioned and revised, based on a variety of evidence which I will review in some detail. Omitting citations for the moment, the evidence pointing to a revision of classical ideas on neural signaling and computation includes:

1. Modulations of response profiles from natural stimuli, including areas outside classical receptive field.
2. Multi-channel spike rate and field potential studies indicating a role for temporal synchrony and modulations in synchrony.
3. A changing role for dendritic action from passive conductance to an active role in processing, related to complex spike arrival time processing and synchrony.
4. Interactions between local fields (electrotonic coupling) and dendritic processing, again related to spike arrival time processing and synchrony.
5. Indications that the same neurons perform different functional roles in stages as the time course of perception unfolds.

One way to view the historical interactions between experimental neuroscience and neural network theory is to consider that ideas about the elementary operations computed by neurons condition ideas about network architectures, which in turn condition experimental approaches. For example, in the history of neural modeling, neurons were first considered as binary threshold units, combined into networks functioning as boolean logic systems. Experimentally observed excitatory and inhibitory synapses, along with the connectionist concepts of weights, advanced the neuron level computation model to support a variety of architectural approaches, shifting the architectural dialog towards signal processing, and to the mapping of learned environmental regularities to representations or behaviors. Competitive principles came to play a large role in network architectures.

The new operations conceived here derive from a research stream rooted in experimental and theoretical biology of neuron populations[4]. Small microcircuit models (Chapeau-Blondeau and Chauvet 1992), small living neural circuits maintained on a silicon electrode array (Kowalski, Albert et al. 1992) and living sub-networks in awake behaving animals (Freeman 2000) have all been demonstrated to show complex, time-varying population firing activity. Observation in living systems and detailed modeling of increasing scales of populations can result first in synchrony, then eventually chaos (by coupling synchronous excitatory and inhibitory *sub-populations* with incommensurable frequencies).

Perhaps finally (but awaiting explicit experimental confirmation) sequences of desynchronization and partial synchronization emerge as "populations of populations" are coupled, mediated by slower waves or impulses of activity *from spatially separate regions* which serve to modify dynamical parameters. The idea that slow wave rhythms in the brain act as a kind of clock was proposed long ago by Wiener (Wiener 1985), but a clock was conceived of as a kind of "gating" operator on signals as in digital logic. An alternative view expressed here is that such rhythms act as a clock controlling synchronization operators by modulating bifurcations in a nonlinear system[5] Synchronization and clustering operations can implement competitive interactions, but their hallmark is cooperative effects, resulting from the interaction of spatially organized input with the geometry implicit in the dynamics. Binding of disparate sensory features into a unified code which preserves compositionality – the ability to recognize sub-parts – is an important possibility for these new primitive operators. The binding problem has so far eluded a satisfying solution by localist, rate coded feature combination hierarchies (i.e. gnostic or grandmother cells), and by feed-forward and attractor networks, due to the compositionality issue. This issue will be taken up in a later chapter in some detail.

These new synchronization operators and the corresponding strategies for coding and computation operate in 16 iterations or less, an upper bound chosen to correspond to plausible biological recognition times (assuming certain spatial and temporal scales of neural computation). Simulations described in this thesis, and by other investigators, indicate that rapid synchronization can occur with coupled discrete oscillators. Emerging

---

[4] While there is a literature on chaos in single neuron responses, I will not discuss it here.

[5] Baird earlier argued that discrete time clocking dynamics in cortical assemblies establish (fast) entrainment and (slower) bifurcation frame rates, with Hebbian learning occurring during the latter.

theories on ensemble density evolution in chaotic systems and on parameter lower bounds guaranteeing synchronization may provide a deeper mathematical explanation for the time course of transients and synchronization, and will be introduced briefly.

In addition to the present work demonstrating that rapid synchronization of transient responses governed by chaotic dynamics has applications in shape recognition, many investigators have been exploring similar systems for the related vision tasks of segmentation and grouping. The generality of the computational strategy with chaotically evolving fields of coupled state variables, suggests that such dynamics and conceptual approaches, while originally suggested for large scale "mass action " models, may also apply to smaller networks. The state variables in large scale models may be *ensemble average firing probabilities or spike rates*, while in microcircuit models (where each node represents a single cell) the state variable can be mapped to *spike firing phase* relative to some reference slow cycle, for example, resulting in local phase distributions which can rapidly perform computations. In the rest of the thesis, I will adopt the convention of discussing firing rates, but the flexibility of these principles is a rather important point.

Historically, pioneering researchers in psychology and neuroscience have emphasized the *differences* between dynamical approaches and computational or symbolic approaches. I will be more conciliatory, following the lead of a research community sometimes termed "physics of information", which examines dynamical systems as information processing systems, often using the tools of symbolic dynamics to bridge these domains.

ORGANIZATION OF THE THESIS

The organization of the thesis reflects several goals beyond the description of a pattern recognition algorithm, chiefly the grounding in psychology, neurobiology and the justification of the coupled map approach as an appropriate technique in computational neuroscience.

In this introductory chapter and continuing in the next, I describe the selected problem of similarity and pattern recognition, emphasizing well-known dilemmas which are negotiated in the present study with a combination of novel approaches and existing methods.

Chapter 2 reviews psychological and computer vision approaches to the issues of similarity, as well as reviewing a selected set of algorithmic approaches to shape recognition. These algorithms were chosen for review from the vast literature on the basis of recency and some family resemblance to the present method. Finally, a section on pure computer science theory relating dynamical systems to pattern classification is raised in this context.

Chapter 3 reviews some neurophysiology and systems neuroscience which is especially relevant to the systems approach followed here. The emphasis is on recent findings and controversies over the role of single neurons versus larger assemblies in coding and computational tasks. Further, I explore the possible nature of computations in such assemblies and at various scales of organization.

Chapter 4 will more systematically introduce the required dynamical concepts. Since this dynamical framework at first appears remote from conventional neural network modeling, I discuss dynamics in the context of larger scale networks and the concepts of macrostate or ensemble variables. Some work of other investigators in dynamical networks is reviewed, emphasizing the visual task of segmentation where most previous work in oscillations has been focused.

Chapter 5 discusses representation and learning. I begin with some discussion of how I approach the problems of similarity and stimulus equivalence with this dynamical framework, then proceed to computation and framing the learning problem.

The dynamical complexity of recurrent, spatially extended systems poses challenges for analytical determination of network parameters, though some recent progress on related issues was described in Chapter 4. As a result, in order to construct network dynamics, the *evolutionary computing*[6] paradigm is used to construct networks. The emphasis on this study is more on dynamics than on learning per se, so the treatment of this aspect of the work is brief.

Chapter 6 describes the computational experiments undertaken. First I describe a variety of simple but novel studies of transients and convergence in single maps, pairs of coupled maps, and lattices with random or simple algebraic initial conditions. Most of these studies examine system responses over multiple parameter and stimulus dimensions, with 3-D animations of the system state versus a parameter plane produced as a result. These animations are provided as an internet based supplement.

Next, I describe a family of experiments involving the construction of a quasi-metric space which serves to order a family of parametric curves. The motivation for using this toy problem is that these visual forms can be unambiguously ordered in a way that maps naturally to their perceptual appearance. This is rarely the case for natural objects. The resulting family of images for a curve, considered as distortions of a canonical image, also resemble distortions due to scaling and rotation in depth of a single object.

Finally, an extensive family of experiments on recognizing the equivalence of different views of objects rotated in depth is described. This problem is formulated in two ways to follow experimental paradigms used by other researchers; one is a database search problem, with a second set following a psychological paradigm of determining whether two successive presentations are the same object or different. The *paperclip* images and variants used here have been used in both psychophysics and neural recordings, proving challenging to monkey and human subjects, with error rates ranging from 30% to chance depending on training circumstances, indicating that the problem is non-trivial. For the stimulus equivalence problem, the objective function used during evolutionary computation involves separate terms to achieve a balance between clustering different views of an object and avoiding the mapping of views with similar local statistics to the same region in the representation space. The concepts of

---

[6] While the term genetic algorithm is more widely used, some authors (e.g. J. Pollack) restrict that term to bitwise parameter encoding and mutation which is blind to “gene” boundaries, with evolutionary computing subsuming both blind mutations and gene-boundary-aware mutations.

regularization or normalization and cross-entropy are introduced. The results of the hybrid (network and algorithmic) system are compared with two other state of the art computational approaches.

Chapter 7 is devoted to discussion, future work and conclusions. I highlight some of the limitations of the study and system, suggesting possible improvements from the standpoint of both engineering and biological plausibility. I revisit competing theories of biological coding briefly, and provide a new interpretation and computational role for synchronous observations and modulations of synchrony (rate correlation) seen in multi-channel neuron experiments, and suggest neurophysiological and computational work. Finally, I summarize the main contributions and conclusions of the thesis.

Appendices are included introducing several technical topics: formal language theory, signal processing, and Markov chains.