SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE DTIC FILE COPY

Form Approved OMB No. 0704-0188

DTIC ELECTE FEB 15 1991 B D

AD-A231 835

Contract F49620-88-C-0025	1b. RESTRICTIVE MARKINGS
	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 91 0009
6a. NAME OF PERFORMING ORGANIZATION Yale University	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION AFOSR/NE	
6c. ADDRESS (City, State, and ZIP Code) 333 Cedar Street New Haven, CT 06510	7b. ADDRESS (City, State, and ZIP Code) Bldg 410 Bolling AFB DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NE
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-88-C-0025	
8c. ADDRESS (City, State, and ZIP Code) Bolling AFB Washington, DC 20332	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2305; TASK NO. B1; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

A NEURAL NETWORK APPROACH TO MODEL-BASED RECOGNITION

12. PERSONAL AUTHOR(S)

Gene R. Gindi

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 12/1/87 TO 3/1/90	12/26/90	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The problem of designing neural networks for visual object recognition was pursued on several fronts. The main thrust of the work involved theory and simulation of networks that store some notion of a relational model of an object and perform recognition via a version of graph matching. This approach is governed by the use of objective functions to both specify the network and the problem to be solved. The dynamics of the net thus carry out an optimization procedure. Key here is the incorporation into the objective function of a compositional and specialization hierarchy of models, and provision to perform dynamic grouping (perceptual organization) of the input data. Good performance was achieved for problem versions where raw data items were pregrouped into subsets consisting of parts of one object, but poorer performance was attained on more difficult problems where the network itself had to perform the perceptual grouping problem. Other experiments on optimal segmentation were carried out that relied on much the same methodology. *CONTINUED...*

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Craig	(202) 767-4931	NE

DD Form 1473, JUN 86   *Previous editions are obsolete.*   SECURITY CLASSIFICATION OF THIS PAGE

91 2 12 032

Applications to real images were carried out and a system for industrial part recognition was demonstrated. Progress was also made in applying learning to these networks. Prematched pairs of parts were presented to the network and feedback circuits were designed to alter connection strengths so as to modify local minima appropriately in the network. A related effort explored aspects of traditional associative memories that may be of use in more complex networks. Questions of performance, storage and robustness were addressed. A new hybrid fast learning algorithm was proposed and tested for a CMAC network. Work in optical implmentation of some of these networks constituted a third front. The main problem here is to use optics to form a fixed interconnection network between layers of 2-D nodes (neurons). A design and implementation of multifaceted holograms was accomplished. Computational procedures for calculating a binary hologram pattern were explored, and several photolithographic versions fabricated. Tests of these in an optical implementation of associative memory showed performance near the theoretical limit. Theoretical work in support of the optical research resulted in slightly improved versions of Hopfield memories.

Final Progress Report
AFOSR Contract F49620-88-C-0025
A Neural Network Approach to Model-Based Recognition

Gene Gindi
Yale University

28 December 1990

1 Summary

In this work, the investigators explored neural nets for object recognition and also investigated, in a subcontractual arrangement with the University of Arizona, hardware issues in the optical implementation of neural nets.

The approach underlying the design of networks was that of optimization theory to both specify the problem and the network for its solution. These networks differ considerably from simple pattern matchers (such as the Hopfield content-addresssable memory) where an iconic version of the pattern itself is stored. Instead, the network implements a form of model matching in which the model base is organized as structured graphs, the input data is organized , again by the optimization procedure, into similarly structured graphs by the neural net, and recognition is accomplished by a form of graph matching. The model base is organized hierarchically in that both compositional (whole-part) and specialization (class membership) notions are captured by sparse matrices which serve as pointer structures in the objective functions. This notion of design allows for a uniform style of addressing both low-level visual problems and more traditional high-level recognition problems.

Our main experiments were conducted in visual domains involving the recognition of simple figures. The following questions were pursued via theory and simulation: (1) Can these nets recognize simple articulated objects when given the important clue of class membership? , i.e. this subset of parts belongs to this object? Answer: Yes, the so-called labelling of parts when object identity is known works quite well. (2) How well do they do when this clue is missing? Answer: Not so well; the simultaneous grouping and recognition is a complex task not well handled by the approach adopted here.(3) What can be done to improve the optimization algorithm? Answer: We tried a Langrange multiplier approach to implement hard constraints as wel as mean-field annealing to escape local minima, Both resulted in improved performance relative to simple descent methods on penalty function objective functions. (4) Can these nets learn from examples? Answer: Some initial progress was made in presenting supervised examples of *pairs* of labelled parts to the net and updating connections accordingly. A new learning algorithm was invented for this purpose. The much more difficult question of learning the database was not pursued at all.

For the optics research, we investigated two means of using spatial multiplexing to effect a 4-D interconnect between two 2-D node planes: multifaceted planar holograms and multichannel incoherent imaging systems. Nonlinearities are implemented electrically after detection of the light intensity. Much progress was made with the hologram approach. Computational procedures using random search and error diffusion were implemented in order to calculate the 2-D binary transparency pattern, and photolithography was used in order to shrink this pattern to a hologram. As a benchmark, connections for a simple associatice memory were encoded and the optical neural net performed quite similarly to an equivalent computer simulation.

2 Introduction

This project grew out of earlier work in which an approach to visual object recognition as an associative memory problem was pursued. The big problem with this earlier approach is that objects must be recognized regardless of changes in position, rotation, scaling, and a host of other deformations. One may store iconic patterns directly in a memory and design such invariances directly into the connection pattern, but the circuits quickly become complicated. We follow the approach taken in traditional computer vision systems and store relational models instead of iconic models into the circuit, and demand that the input data itself be organized into such a relational structure and optimally matched to the "nearest" model. Relational models are designed to automatically capture the desired invariances; the perceptual organization of input data into relational structure proceeds simultaneously with the matching process. Like some associative memory approaches, this one also uses objective functions for specification of the problem and design of the circuit.

Early work on optical implmentations was also continued in this grant. The original work focused on the use of optics for fast recognition of 2-D geometrical patterns independent of scale, rotation, translation. By stages, the focus shifted toward implementation of neural nets to accomplish the same goal, and, by extension, toward the general problem of optical imterconnects.

3 Summary of Completed Research

Below is a succint summary of accomplishments of the two yeargrant period. Each item is followed by a reference to an appropriate paper or technical report that gives a more detailed account.

- We implemented two networks for simple object recognition and performed analysis and simulation experiments. Each of these successfully matched simple stick figures to a database of models. It was able to find multiple objects and specializations of objects. One version [6] used an unconstrained optimization technique for net dynamics; the other incorporated "Lagrange multiplier neurons" to implement hard constraints [17]. In each case, the difficult task was made tractable by hand coding the input data into relational structures suitable for matching. The resulting objective functions were quadratic and the net worked well.

- A more difficult version of the above task was attempted. Again, the idea is recognize simple stick figures, but now the network, as part of the optimization process, had to group input sticks into potentially meaningful relational structures. This resulted in 5th-order objective functions. While success of the network was limited, much was learned. Results are reported in [1], [10], [12], and [8].

- A version involving recognition of 3D objects was completed. This [18] network recognizes 3D-stick figures from a 2D projection.

- Photolithographic multifacetted holograms for optical neural net interconnects were fabricated for an associative memory problem. The optical results were compared to results from a simulation program written to model sources of error in the optical scheme. Results showed [11] , [13] close agreememt with theory.

- In support of the optical effort, we completed a study on performance of outer-product associative memories. One result [2] showed that versions with self interconnects perform better than those without; another [4] showed that a version with positive only interconnects can be made to perform well. Both of these results have ramifications for optical schemes.
- An additional study on optimal architectures for outer product associative memories was completed [16]. A universal architecture that makes optimally efficient use of hardware is proposed in this study.
- Though learning hasn't been a major theme of the work, some progress was made in initial studies for learning distance metrics in the graph matching networks mentioned above. In an unrelated study, a fast, general purpose, supervised learning algorithm based on CMAC models was devised [14] [15]. On a popular test case, it greatly outperforms backprop.
- The object recognition system was applied to a real doain of industrial parts. [3]
- We tried mean-field annealing optimization on these networks with some success. [3]
- A network for model-based segmentation of signals was devised. The idea was to address the grouping problem in a simplified setting. (Work described later in this report.)
- The beginnings of learning schemes were applied to the recognition nets. (Workdescribed later in this report.)

4 Discussion

The work splits into three categories: networks for model matching, optical implementations, and analyses of associative memory. The optical work was carried out as a subcontract at the University of Arizona and is discussed in a self-contained section.

4.1 Networks for model matching

We describe here qualitatively our model match networks. Quantitative details can be found in [1] or in the attached reprints.

Neural net tasks for visual recognition is often thought of as a variant of some simple pattern matcher, such as the the Hopfield associative memory or a simple perceptron. These schemes are limited in two ways: "objects" are represented iconically instead of in the more efficient manner of relational structures, and there is no provision for efficiency in search by using notions of hierarchy. Of course, both of these ideas are common in traditional computer vision, but here, we propose a way of incorporating these crucial notions into a neural-net paradigm.

We introduce an optimization approach for solving problems in computer vision that involve multiple levels of abstraction. Specifically, our objective functions can include compositional hierarchies involving object-part relationships and specialization hierarchies involving object-class relationships. The advantage of hierarchical organization is that it makes the search process involved in image interpretation easier to express and more efficient. The large class of vision problems that can be subsumed by this method includes traditional model matching, perceptual grouping, dense field computation (regularization), and even early feature detection which is often formulated as a simple filtering operation. This raises the possibility of solving within a single vision system both low-level and high-level problems in a uniform manner.

Our approach involves casting a variety of vision problems as inexact graph matching problems, formulating graph matching in terms of constrained optimization, and using analog neural networks to perform the constrained optimization.

Our extension of graph-matching to model-based object recognition involves regarding one of the graphs as a "model" graph, which is supposed to represent the knowledge of shapes within the system, and the other graph as a "data" graph which is obtained from the current input data to the system. The model-side nodes are simply called "models" and data-side nodes are called "frames" (denoted F_i), which are collections of analog neurons representing parameters of an object (and are denoted F_{is}, where s indexes the parameters of a single frame). The instantiation of a model in the image is expressed by "turning on" a match neuron $M_{\alpha i}$ between a model α and its matched Frame F_i. We will refer to such a network of frames and models as "Frameville".

The incorporation of a compositional hierarchy and a specialization hierarchy on the model side is achieved via graph-arcs called *INA* links and *ISA* links respectively. The objective function includes terms representing a simultaneous match of a model to an object (on the data side) and the parts of the model to parts of the object in a consistent fashion. An objective function may be inherited through the *ISA* links from a model to its specializations and there may be an incremental objective function for each of the specializations. Numerical parameters are represented by using analog neurons and the verification of metrical relationships involving these parameters is achieved by corresponding consistency terms in the objective function.

In order to perform perceptual organization, the data-side compositional hierarchies must be dynamic. To achieve this we introduce dynamic *ina* links on the data side. The *ina* links connect

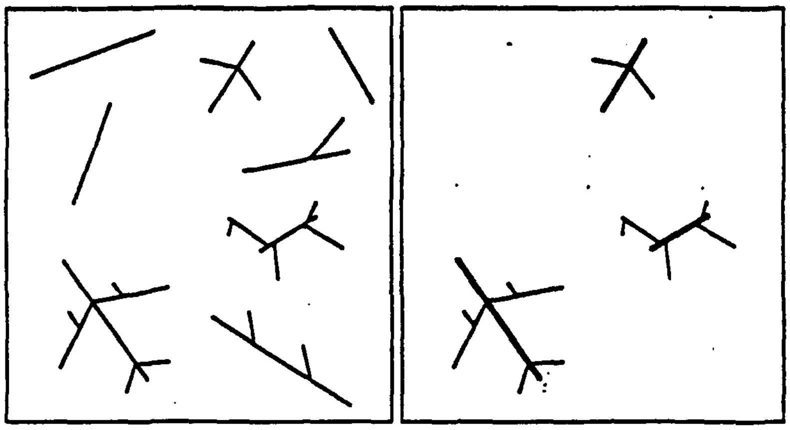


Figure 1: **The task in Stickville** Image at left shows objects among noise sticks. The goal is to find all instances of objects in the model base in a manner independent of of translation rotation scale and some distortion. Two instances of *plane* and one of *mammal* are found as shown at right. Objects are abstracted by parameters of a main part (heavy line).

more and less abstract frames, and their evolution corresponds to a search for a perceptual organization consistent with the model graph. Specialization is implicitly achieved by the simultaneous match of a frame to a model and to one of its *ISA* specializations.

4.2 Experimental results

The Frameville approach described above is quite general; implementations for specific domains is a design problem requiring some care. Here, we describe some specific implementations.

The earliest effort was in a domain, Stickville, in which the problem was to recognize simple stick figures in a 2-D image. Accounts are found in [6] and [17].

Figure 1 illustrates the problem; we are given a clutter of sticks and we must find objects among them. An object is an assemblage of connected sticks (parts) whose internal geometry satisfies model demands. We must find all instances of a given object and there may also be present more than one kind of object. An object must be found in a manner invariant to translation, rotation, scaling; a certain amount of distortion or missing parts is to be tolerated too. Through incorporation of an *ISA* hierarchy, Stickville finds general classes of objects (e.g. it locates a *plane*) and their specific subclasses (specifically a *jet*).

A low-level frame is associated with a single stick and stores the end points of the stick. In Stickville, we use a mainpart abstraction; the parameters of a single designated stick summarizes the entire assemblage of sticks constituting an object. At the fixpoint, high level frame is matched to a stick identified as a mainpart. The parameters of this mainpart need not be computed so the slots in these frames need not be represented by analog neurons as in full-blown Frameville. A second restriction applies here. The *ina* matrix is assumed constant through the use of a connectedness hueristic: If two sticks i and j are attached,then $ina_{ij} = 1$ else it is zero. The only remaining dynamic variables are the match neurons M.

In Stickville, a match metric compares relative quantities of a data stick pair i, j to ideal values

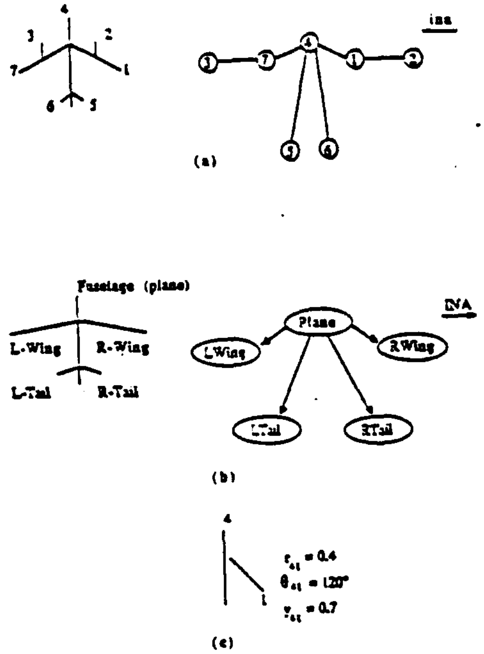


Figure 2: **Definitions in Stickville** (a) A stick figure of a plane, and its representation as a graph with *ina* links. (b) A plane model and its representation as a graph with asymmetric *INA* links. (c) Parameters characterizing the relationship between two data sticks.

associated with model pair α, β. The relative quantities are $(r_{ij}, \theta_{ij}, q_{ij})$ defined as relative size, relative angle, and position of attach point for stick pair i, j. Since the slots for any frame pair F_i, F_j are known *a priori*, we may precompute $(r_{ij}, \theta_{ij}, q_{ij})$ and hence the entire metric $H^{\alpha,\beta}(F_i, F_j)$. In Stickville, all of the metrics are thus precomputed. Figure 2 shows an assemblage of sticks and its *ina* representation, the model to which it should match and its *INA* representation, and a stick pair with its relative coordinates. In figure 2b , the mainpart of *plane* is *fuselage*; the corresponding data stick in Fig 2a is stick 4.

A successful recognition of two *jets* is shown in Fig. 3 in terms of the match matrix. The result here used the unconstrained optimizatin method, though [17] shows results using the constrained method. Though the objective here is quadratic with only the M's variable), the constraints were of order 3 when expressed as penalty terms.

It is interesting to note that the experiments were quite successful and that Stickville could almost always correctly find objects for the size of problems (about 10 models) considered. In [17], the results of experiments are analyzed in some detail. On the other hand, the problem was made "easy" by restricting the search so that no dynamic grouping of data was needed (ie. *ina* was constant) and no dynamic computation of analog parameters was done. Stickville showed that the formidable combinatorial matching problem could be solved.

Given that success, the next escalation was to try a domain that included dynamic grouping and computation of abstractions, i.e. a full-blown Frameville. A Description in detail is found in [10]

Figure 4 shows the problem. Input data consists of vertical or horizontal unit-length sticks. The net must find instances of T-juctions (e.g. sticks 1-2-3) or L-junctions (e.g. sticks 5-6-7) amid extraneous sticks and label the parts as well as compute parameters summarizing each object. The simplicity of the problem is misleading; why not use a matched filter for example? The problem is

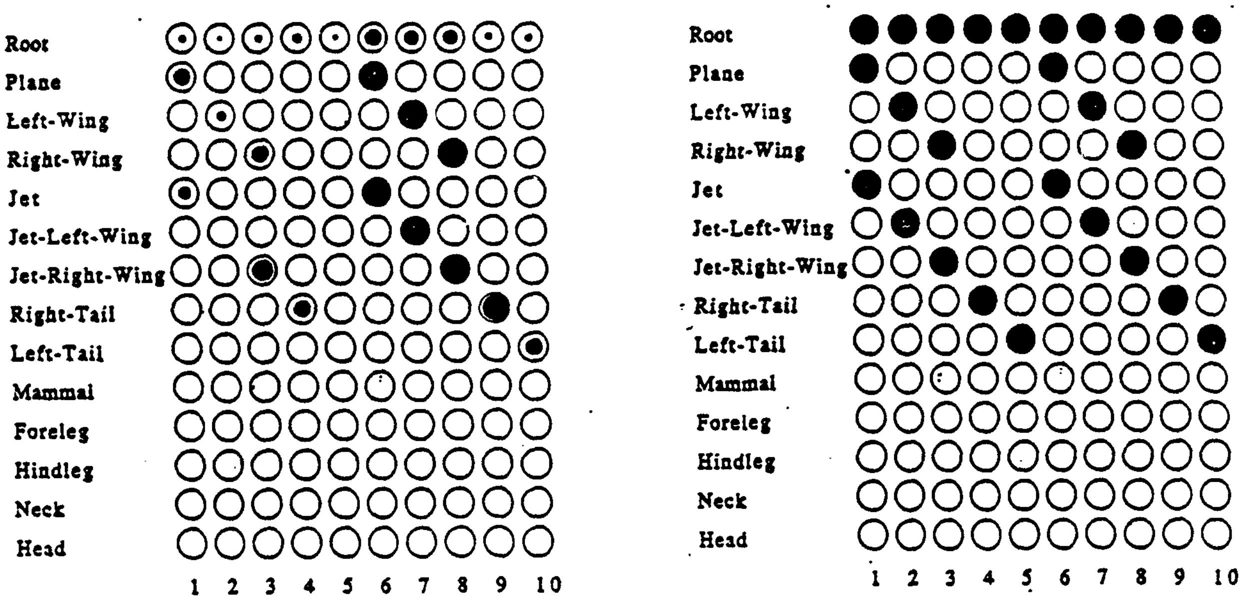


Figure 3: **Stickville results** are displayed in terms of the match matrices (circles) whose value is encoded as the radius of the shaded portion. Each match neuron is indexed by the model (row) and stick (column). (a) Shows match matrix at 28 time steps. (b) Shows it at 70 tme steps. It has correctly found two instances of a *plane* and its specialization *jet*. Note that a stick matches to a model and all of its generalizations (more than one on in a column) and that both *jet*s have been found (more than one on in a row).

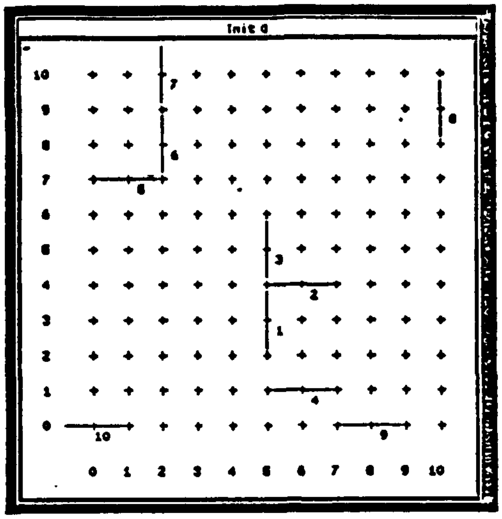


Figure 4: **Input data for TLville**

kept simple for diagnostic purposes; the same machinery used here could solve recognition problems for which simple template matching would be impractical.

Figure 5 shows the model base and the frames. Models (ovals) occur at two levels, line segments and junctions. The *INA* links for a T-junction are shown. We resrict the form that an object can have by allowing at most seven positional roles for parts, arranged in the form of the familiar seven-segment LED display as illustrated iconically in Fig. 5. The triangles are frames, each with three slots. The low-level frames Q_j matched to segments contain the parameters $Q_{j,s} = (x, y, \theta), s = 1, 2, 3$ denoting position and orientation of the segment. High-level frames P_i matched to objects have the same parameters, but these apply to a mainpart. Unlike the case in Stickville, they are unknown ahead of time and must be computed dynamically as part of the optimization. Match neurons neurons ina_{ij} that group low-level frames to high-level frames are shown (circles). Unlike the case in Stickville, the grouping is computed dynamically, thus performing a "perceptual organization".

Typical results are shown in Fig. 6 Given a random start and the input data of Fig 4, the network often gets trapped in an unfavorable local minimum as depicted in Fig 6. With only a single model in the database, however, our results are much improved. The network, if given a "hint" in the form of the *ina* neurons initialized to correct values, does indeed converge to the correct solution as shown in Fig. 6b.

The methodology we use can be applied to an entire vision hierarchy from low-level feature extraction to high-level structural matching. At a high enough level, data is assumed grouped in some meaningful way (e.g. pixels to sticks) and these grouped data entities compete for matches to parts. At the lowest level, we are usually facing a problem such as edge-detection or shape-from-x that often amounts to local retinotopic processing. At intermediate levels, the data must in many applications be partitioned into contiguous regions guided only weakly by the specific expectations of high-level models. It is at this intermediate level of segmentation that we now focus.

Vision systems eventually face the problem of segmenting a scene into *parts* for purposes of object recognition or other goals. Neural network formulations of vision problems have usually been applied to early vision where the style of processing is uniform. Here, we formulate a network for partitioning an image into disjoint parts, and simultaneously delivering appropriate values of

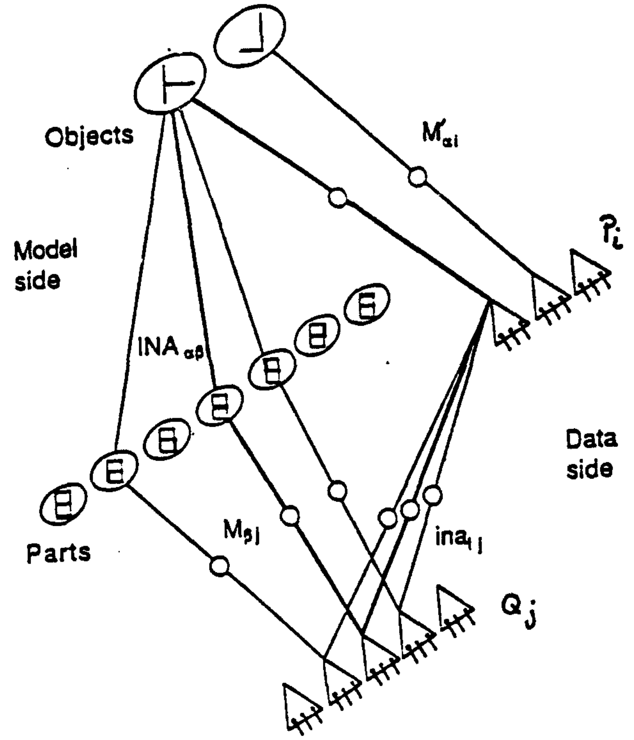


Figure 8: **Structure for TLville** Models occur at two levels, line segments and junctions. The *INA* links for a T-junction are shown. Each frame has three slots: two position coordinates and one orientation coordinate. The bold lines highlight a possible consistent rectangle. The M and *ina* neurons are displayed as circles. This does *not* show the connections between neurons, just the entities that the neurons match up.

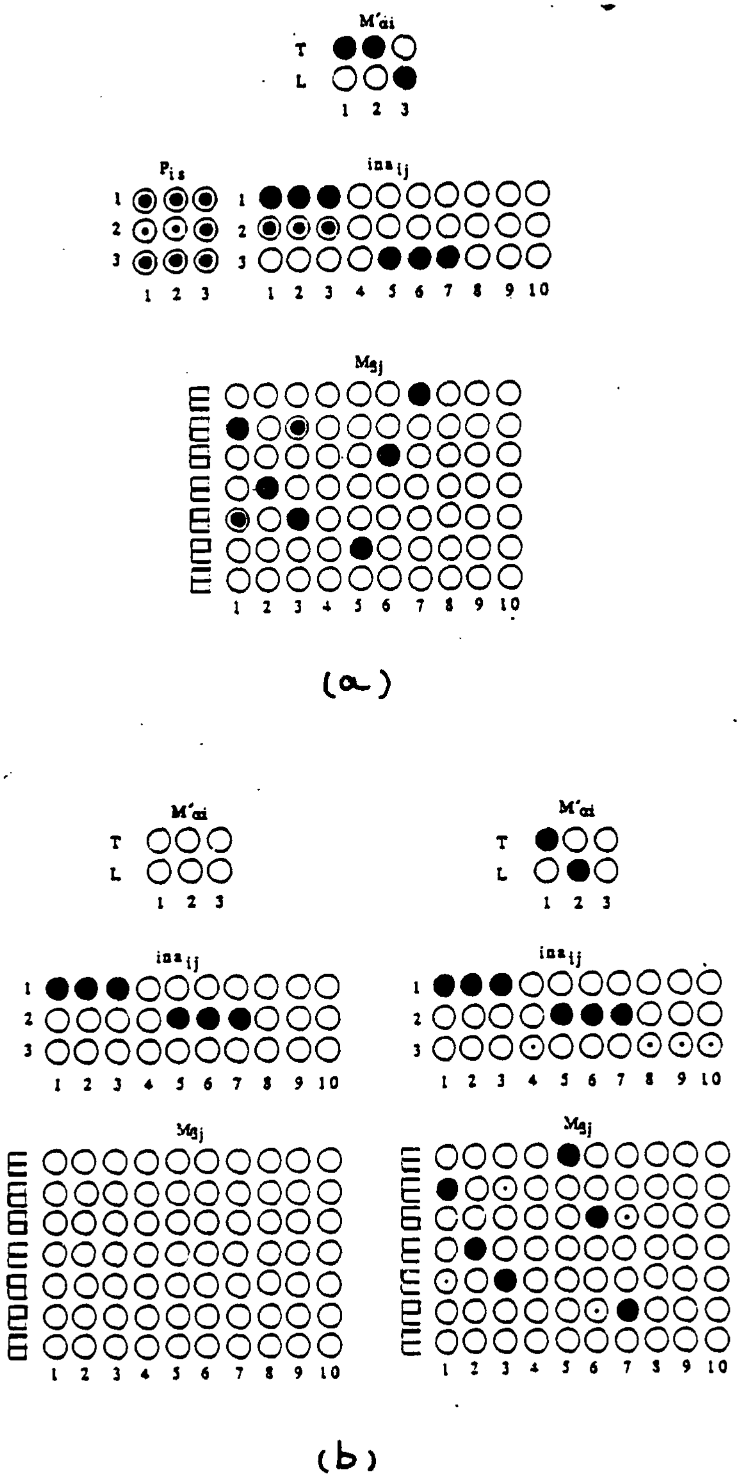


Figure 9: **Experimental results in TLville** The matrices show the state of the various neurons depicted in Fig 8 (a) Shows a typical failure. (b) Shows a correct result after network has been initialized with the correct *ina* neurons on.

parameters that describe the parts. A part is a parameterized function that closely approximates the pixel values comprising the part.

The goal of the segmentation calculation is to partition an image into relatively few contiguous regions so that each region is closely approximated by a part. We can formulate a version of the Frameville rectangle rule for segmentation. In the language of Frameville, a low-level frame Q_j is associated with pixel j of the image; the value of the single parameter in Q_j is simply the gray-level value, which we just denote by Q_j itself. A high-level frame P_i holds the parameters of the part. We do not address the question of choosing a "good" part parameterization, e.g. a superquadric; the vocabulary of parts models is assumed supplied by a user. Parts models are indexed by α. Since there is only a single pixel model, all Q frames are matched to it and $M_{\beta j}$ is unity. Since a pixel model is part of all high level parts models, $INA_{\alpha\beta} = 1$ for all model pairs. The rectangle rule becomes simply

$$\sum_{\alpha} \sum_{ij} M_{\alpha i} ina_{ij} H^{\alpha}(P_i, Q_j) \tag{1}$$

The difficulty of the segmentation problem lies, of course, in choosing a partition of the data into nonoverlapping contiguous subsets so as to minimize the cost of the parameter fit. The segmentation of the data is expressed by the values of the neurons ina_{ij}.

The ina must obey constraints that must obtain at the fixpoint of the optimization. The four used in the network are:

- A given pixel is or isn't part of a given segment. ($V \Rightarrow 1\, or\, 0$)
- A given pixel may belong to only one part.
- The pixels belonging to a part must form a spatially contiguous set.
- A part i may have one or none contiguous pixel sets assigned to it.

The constraints appear as penalty terms and are similar to those used for TLville except for the contiguity constraint. The penalty term for this counts 0-1 and 1-0 transitions in a row of the ina matrix and forces one or none transitions of each type.

It is still possible to satisfy the constraints and achieve a good fit by allocating a huge number of parts, but the abstraction of pixels to parts is useless unless the segmented image contains much less information than the original. To prevent this part proliferation, we include a parsimony term in the objective function:

- use as few parts as possible

Or. current parsimony term just counts the number of parts used. The weight on this additive term determines the relative cost savings of fewer parts vs. better fits, but we have no good criterion yet for choosing this weight.

We use the objective function to specify a high-order Hopfield network of neurons ina, M and P to carry out the optimization. As the optimization proceeds, groups of pixels clump together to form parts, and the neurons representing parameter vectors settle to appropriate values.

Figure 7 shows a result of the segmentation net partitioning 1-D data into constant regions. Here, P_i is a single number equal to the value of the constant approximating region i. The appropriate metric is simply

$$H(P_i, Q_j) = \sum_{ij} ina_{ij}(P_i - Q_j)^2. \tag{2}$$

where H needs no superscript since only one model is present. The partition is stable; we get the same answer with different random starts and with small changes in the data.

Other more recent implementations are described in following pages. One page shows the recognition of real industrial parts while another shows the use of these nets for recognition of three-dimensional objects.

The networks discussed here were hand designed in that both the model base and the definition of the match metric $H^{\alpha\beta}$ were chosen in an *ad hoc* manner. It may be possible to improve the performance of a Frameville network by supervised learning. Learning the database is quite difficult, but improving $H^{\alpha\beta}$ for greater discrimination may be possible. A possible strategy would be to present the network with examples of fully matched models (all match variables set). The match metric $H^{\alpha\beta}$ is then changed. For example, it may be desirable to make its value zero or negative for positive training examples and have it assume a large positive value for negative examples. The measure governing the change of H might be the separability of the clusters of energy values reported by the analog computation term for positive and negative examples. The match metric itself could be modified by representing H in parameterized form and descending in the parameter space. The next page shows results of initial work summarizing this aspect.

4.3 Work on associative memory and learning

The model matching networks described above differ from the more familiar associatve memories well known to neural netters, but since associative memory serves as a test problem for optical implementaions, several problems in analysis presented themselves. In addition, we continued some previous work in new types of associative memory.

It turns out that a Hopfield style ACAM (Associative Content Addressable Memory) may be implemented optically as projection of a binary input state onto a set of stored memories to obtain a set of inner products, followed by summation of stored memories weighted by the innner products, followed by thresholding. With this kind of optical architecture, it turns out to be convenient to use connections that are positive only instead of bipolar, and to not restrict onesself to eliminating self connections from the network. These two resrictions (bipolar nodes, non self connected nodes) are present in the Hopfield model, but in two papers [2] [4] we show how to remove them. In particular, we show through statistical arguments that the model with self-connected nodes actually works better than the one with without these, and that an all-positive network is possible if the threshold point is selected judiciously.

While the outer-product memories have been well studied and implemented optically, there is reason to prefer alternative models that use simple template matching in conjunction with a layer of internal decision units which compete to perform a winner-take-all (WTA) function. We refer to this a a unary model. There are 3 reasons for our interest: WTA networks are modules of the model-match networks described earlier, they constitute an implementation challenge for the optical effort (see next section), and unary models are interesting in their own right.

With this latter reason as motivation, we completed a study of unary models [16] and showed the following: We present a universal architecture for standartd auto-associative memory models which makes optimally efficient use of hardware. This architecture is described by a bilinear energy functiom. Bot outer-product and unary models can be viewed as special cases of this unversal architecture. The universal architecture uses only the minimal number of binary connections required by information theory to encode the stored memories. For higher order outer product memories,

Progress: Three-Dimensional Object Recognition

The match networks were used to recognize 3-D objects from stereo images. Shown at left are a stereo pair of images after edge finding. The corresponding data graphs are shown. A match network (not shown) recovers 3-D structure by stereo correspondence. The 3-D structure is itself matched against a database as shown ar right. (Each square is a match neuron.)

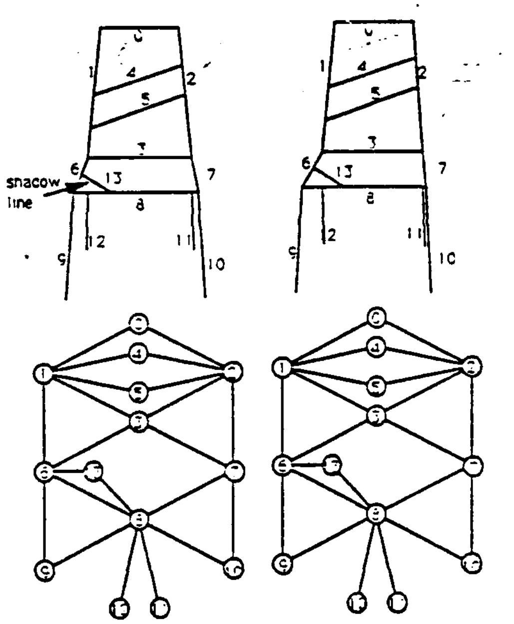


Figure 1: Top: Stereo images of two chairs. Bottom: Graphs generated from images.

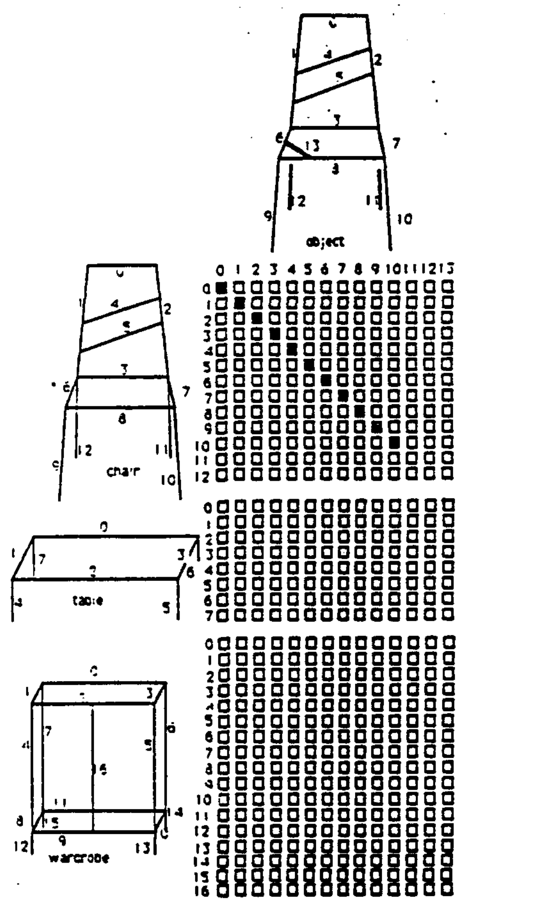


Figure 6: The data graph is matched against three model graphs.

Progress: Recognizing Industrial Parts

The match networks were used to recognize industrial parts from 2-D images. Models are stored as graphs, data compiled into graphs as shown below.

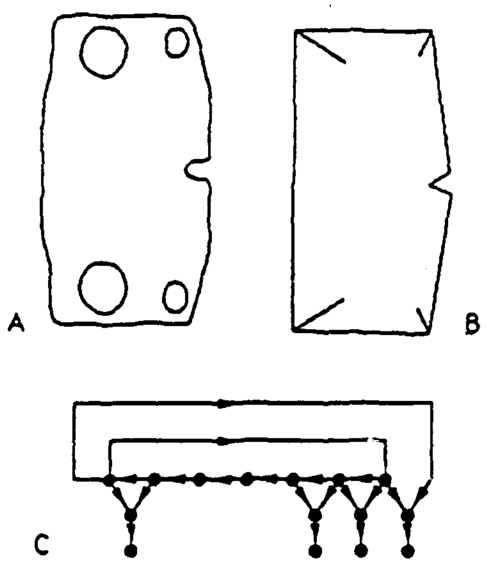


Figure 1: A: Model after edge detection. B: Segmented into sticks. C: As described by a graph.

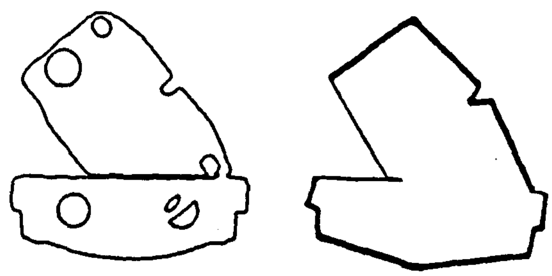
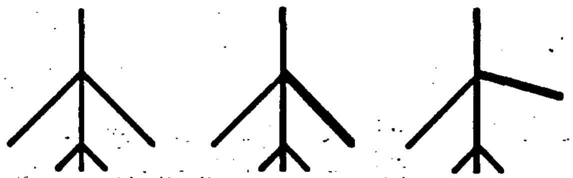


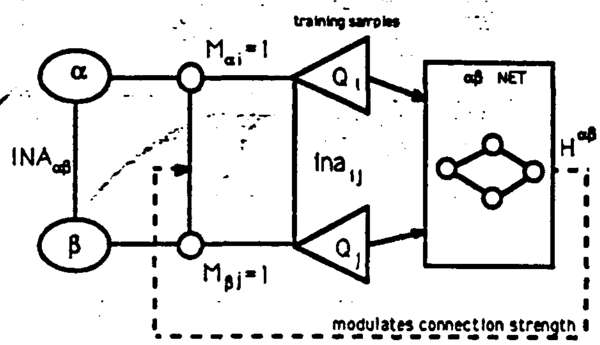
Figure 4: Overlapping pieces. Fat sticks were matched correctly.

Learning

- Match Metrics can be learned from example, instead of being hand designed.
- Learning may be accomplished by presenting labelled positive and negative examples to the learning network, of the form:



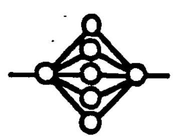
- The network, including the subnetwork responsible for learning the match metrics, takes the general form:



(a)

The output of the learning net is used to modulate the connection strengths among match neurons.

- The actual network responsible for learning takes the form:



with the equations of motion:

$$
\begin{aligned}
g(M(c_i, \kappa_i, x_i) &= \frac{c_i}{(1 + e^{\kappa(x - x_i^2)})} \\
E(M) &= \sum_i \sum_j g(M_i(c_i, \kappa_i, x_i)(g(M_i(c_i, \kappa_i, x_i) - data_j)
\end{aligned}
$$

where g is a Gaussian function. Learning takes place by arranging the gaussian neurons with respect to their magnitude c_i and center x_i.

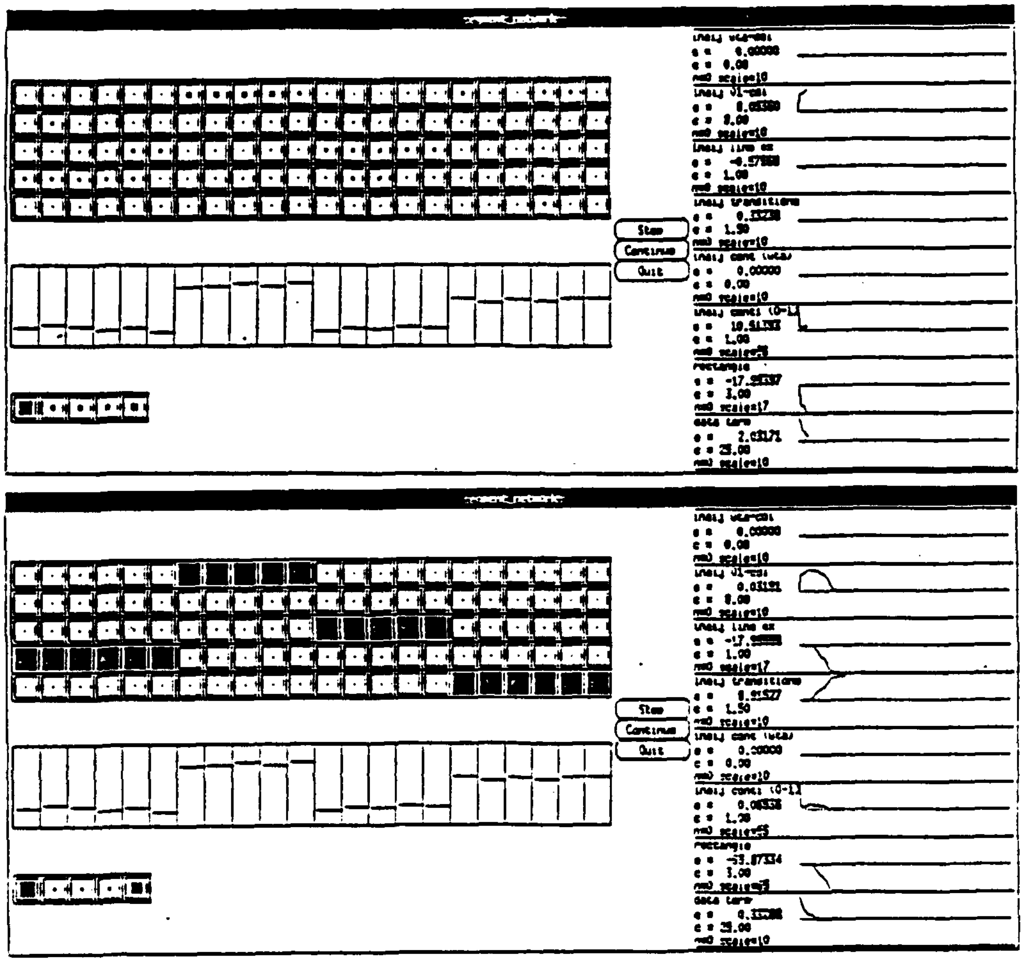


Figure 10: **Segmentation of 1-D data** into piecewise constant regions. Upper image shows results early in the optimization while lower one shows the fixpoint results. The data is indicated by a bargraph; the *ina* matrix is depicted above, and the evolution of various terms in the objective function is shown at right. The net allocates 4 parts out of a maximum of 5 for its segmentation. The neurons at lower left of each display show the value of P_i, the approximating constant. At the fixpoint, P_i equals the mean of the pixels chosen for segment i, i.e. row i of the *ina* matrix.

the need for large numbers of internal "product units" is eliminated.

In the previous section, an approach to learning in Frameville was discussed but no committment to a given learning algorithm was mentioned. Reseach on learning, though somewhat peripheral to the immmediate goals in this contract, was conducted. In [14] [15], a learning algorithm in whuich a system learns to approximate mappings by constructing an interpolating lookup table on a lattice of points in the input space.

4.4 Work on optical implementation

This section describes work on optical implementation performed as a subcontract at the University of Arizona. A paper [11] by the subcontract investigators is re[produced here and summarizes the optical progress. Figure numbers refer to figures within this report.

Design and Demonstration of an Opto-Electronic Neural Network using Fixed Planar Holographic Interconnects

Paul E. Keller and Arthur F. Gmitro
Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

1. Introduction

A key element of most neural network systems is the massive number of weighted interconnections used to tie relatively simple processing nodes together in a useful architecture. The inherent parallelism and interconnection capability of optics make it a likely candidate for the implementation of the neural network interconnection process. While there are several optical technologies worth exploring, we are looking at the capabilities and limitations of using fixed planar holographic interconnects in a neural network system and have implemented an initial test system using planar holograms and opto-electronic nodes.

2. System

All neural network systems consist of nodes (simple non-linear elements crudely imitating biological neurons) and weighted interconnections (synapses) between nodes. The basic system we have looked at employs optical interconnects and electronic nodes in a feedback architecture. A prototype is shown in Figure 1.

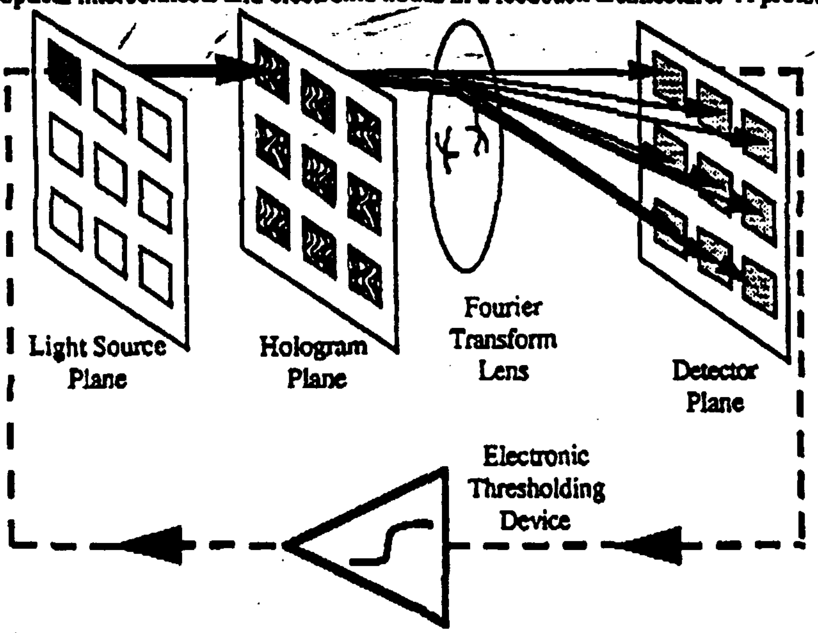


Figure 1: Prototype opto-electronic neural network system.

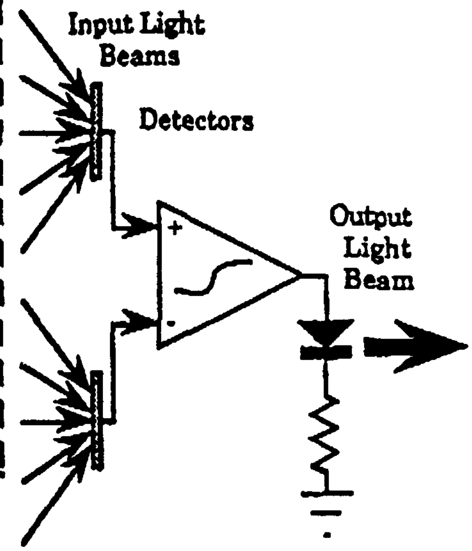


Figure 2: Single node in the system.

Each node is composed of an input summing port, non-linear transfer device, and an output port. In an opto-electronic system, a differential pair of detectors is operated as an input to the node; signals with positive (excitatory) weights arrive at one detector, and *signals with negative (inhibitory) weights arrive at another detector.* These detectors sum up the intensity of each optical signal arriving at the node. A threshold operation is electronically applied to the detected signal to produce an output signal. The output signal of the node modulates an optical source. Figure 2 illustrates an idealized node.

An individual node drives an optical beam that illuminates a single subhologram. Each subhologram stores the connection weights between that node and all other nodes. A subhologram is designed as a Fourier transform hologram and used in a coherent optical system so that the diffracted connection pattern is independent of subhologram position.

3. Design

The Hopfield[1] auto-associative memory model was chosen as a means to test the interconnect capability of planar holographic optical interconnects in the experimental opto-electronic neural network. This neural network tries to associate each pattern presented to it with a pattern that it was trained on during an initial batch training process. Using the Hopfield outer product formulation, a training set of patterns was used to construct the fixed interconnection weights. The Hopfield model is a globally interconnected neural network; all nodes are connected to all other nodes with a deterministic strength or weight. These weights were then encoded in an array of binary amplitude subholograms.

Much effort went into the construction of the holograms used in the interconnect process. After examining a variety of computer generated hologram (CGH) techniques for accuracy of reconstructed interconnect weights, computation time, required space bandwidth product, and diffraction efficiency, two techniques, error diffusion and random search, were found to satisfy many of these criteria.

Both techniques were used to produce binary amplitude holograms. The general design techniques are as follows. The weights connecting a single node output to all node inputs are represented as an intensity pattern in the detector plane. The optical amplitude at the detector plane is given by the square root of the intensity. To reduce the dynamic range required to encode a hologram, a random phase function is added. Since amplitude holograms must produce a hermitian diffraction pattern, the mapped weights are shifted off the optic axis and a hermitian conjugate is added. To compensate for the sinc function roll-off in the connection pattern due to finite sized hologram pixels, the connection weights are multiplied by an inverse sinc function weighting. This predetermined diffraction pattern is then inverse Fourier transformed, and the transformed data values are normalized using the extreme amplitude values for the entire set of subholograms. Lexicographically scanning these sampled values, each sample is binarized. Since the data is continuous valued, an error is produced by the binarization. This error is propagated to the adjacent unbinarized pixels of the hologram; this is the error diffusion process.[2] The net effect of the error diffusion process is to reduce the total quantization error across the entire hologram. What remains is a high frequency binarization error that manifests itself as diffracted light far off the optic axis in the detector plane. The location of this diffracted light is controlled by the method used to distribute the binarization error in the hologram. To improve interconnection weight accuracy and to confine the diffracted spots of light to the center of each detector cell, each hologram is replicated 4-times vertically and horizontally (16 replicas). The error diffusion algorithm has shown the best performance for a non-iterative CGH design process.

Random search is an iterative process used to improve the connection accuracy of the error diffusion holograms. Starting with an error diffusion hologram, this method determines whether a perturbation (flipping a pixel in the hologram from opaque to transparent or vice versa) improves the accuracy. A perturbation is kept only if the accuracy is improved. This process is repeated until convergence. The main disadvantage of the random search process is the massive computation required. This process is related to the simulated annealing process except that no annealing takes place. The simulated annealing process allows accuracy degrading perturbations of the hologram to be kept with a probability modeled by the Maxwell-Boltzmann distribution.[3] Simulated annealing, in theory, is able to find the globally optimum solution; in practice, limited computation requires compromises that may or may not produce good results. We have found that the random search algorithm produces holograms with almost the same performance as simulated annealing but requiring far less computation time.

For a large scale problem, electron beam fabrication would be required to produce the array of subholograms. For the small scale problem that was implemented, a photolithography process was used; the hologram mask was printed onto a sheet of film using a laser film writer and photographically reduced onto a holographic plate.

4. Experiment

The experimental opto-electronic neural network is illustrated in Figure 3. An initial pattern of 8 by 8 pixels is fed into the system by a computer; this pattern represents the initial state of the neural network. The pattern is written onto a Hughes Liquid Crystal Light Valve SLM using a high-intensity projection television. This binary pattern is polarization encoded onto the coherent optical laser beam by the SLM. The polarization beam splitting cube reflects only the vertical component of this polarized signal so that a binary amplitude pattern illuminates the hologram array. Each pixel of the pattern illuminates an individual subhologram. There are 64 nodes with 4096 bipolar interconnections in the experimental system.

The Fourier transform (Fraunhofer diffraction pattern) of the hologram array is produced at the back focal plane of the lens. To reduce scatter, the low frequency information of the diffraction pattern is filtered out. A relay lens is used to image the filtered Fourier plane onto a video camera. The light beams (diffraction from the hologram plane) arriving at the detector plane constitute the input to the node plane. In a practical opto-electronic neural network, each electronic node would take the difference between the signal on its positive-weight detector and its negative-weight detector, threshold the result, and drive an optical source such a laser diode to be either on or off. For our experimental test system, a video camera is used to detect the optical input signals. The video signal is fed into the computer where it is digitized by a video frame buffer. The computer splits up the video frame into a grid and sums up the intensity in each cell to simulate a detector array. The difference and thresholding operations are performed digitally and the output stored in a video frame buffer, where the video output represents the next iteration of the network. This forms the new network state, which illuminates the hologram plane, and the process continues until the network converges to a stable state.

As this experimental system was described, a node can take on two values; a value of 0 is represented by a dark pixel, and a value of 1 is represented by a light pixel. The performance of a Hopfield style neural network is significantly improved by using bipolar node values instead of unipolar node values. As an experimental test of bipolar nodes, a two step process was used. During the first step, a pattern was projected onto the SLM, and the detected pattern on the video camera was stored. During the second step, the inverse of the pattern was projected onto the SLM, and its detected pattern on the video camera was subtracted from the first detected pattern. Polarization encoding is another method for constructing bipolar nodes. A node value of +1 is encoded as horizontally polarized light, and a node value of -1 is encoded as vertically polarized light. With bipolar weights and bipolar state values, four detectors and two polarizers are used in the input summing port of the node.

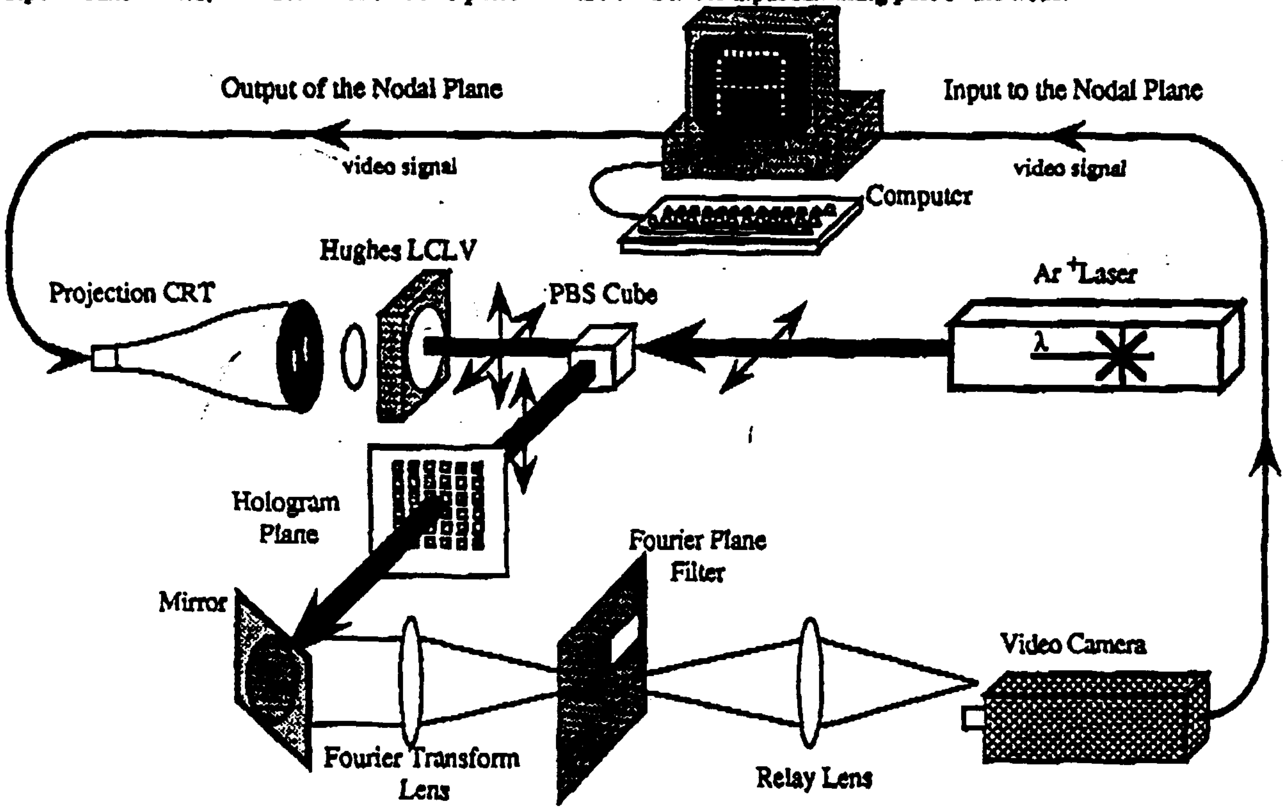


Figure 3: Experimental opto-electronic neural network used to test and evaluate the performance of planar holographic interconnects.

5. Results

The best performance of the associative memory neural network would come from a network storing randomly generated patterns, but since patterns of distinct structure (vertical lines, horizontal lines, diagonal lines) are generally encountered in vision and pattern recognition tasks, it was decided to use a set of ordinary typewriter characters (letters, numbers, symbols) to construct the test network. Using the Hopfield outer product formulation, a training set of three patterns, A B X, was used to determine the interconnection weights.

A prime feature of auto-associative memory neural networks is the convergence of the network to the ideal stored pattern when the input pattern is corrupted. By randomly flipping the pixels of the training set, a test set of corrupted patterns was generated. These patterns were presented to the experimental opto-electronic neural network, a computer simulation of the opto-electronic neural network, and a computer simulation of the ideal neural network. From the simulation, it was found that the auto-associative neural network constructed with random search holograms performed almost identically to the same neural network with ideal interconnect weights. The experiment, while not performing quite as well as the simulation, did come close for both unipolar and bipolar state values. The results with error diffusion holograms were not as good as with the random search holograms but show that error diffusion based holographic interconnects are a good trade-off between system performance and CGH computation time for bipolar state values. Figure 4 illustrates the performance of the experimental opto-electronic neural network with a test set composed of corrupted versions of the letter B. This figure is a graph of the probability that the network converges to the original letter B as a function of the number of corrupted pixels in the input pattern. The total number of pixels in the input is 64. From this graph, it is apparent that when the number of incorrect pixels in the input pattern becomes too large, the network does not converge to the ideal pattern. Similar responses were found with the other stored patterns.

The small differences between the experimental results and the simulation results were caused by aberrations in the Fourier transform lens and relay lens, non-uniformity of the video camera, high frequency roll off in the holograms due to loss of resolution during the hologram fabrication process, and RF interference in the electronics produced by the argon ion laser's plasma discharge tube.

The experimental opto-electronic neural network system along with its computer simulation shows that a planar hologram can be used to implement the interconnect weights of a neural network. The results we have found with the experiment agree well our analytic calculations of neural network performance.[4]

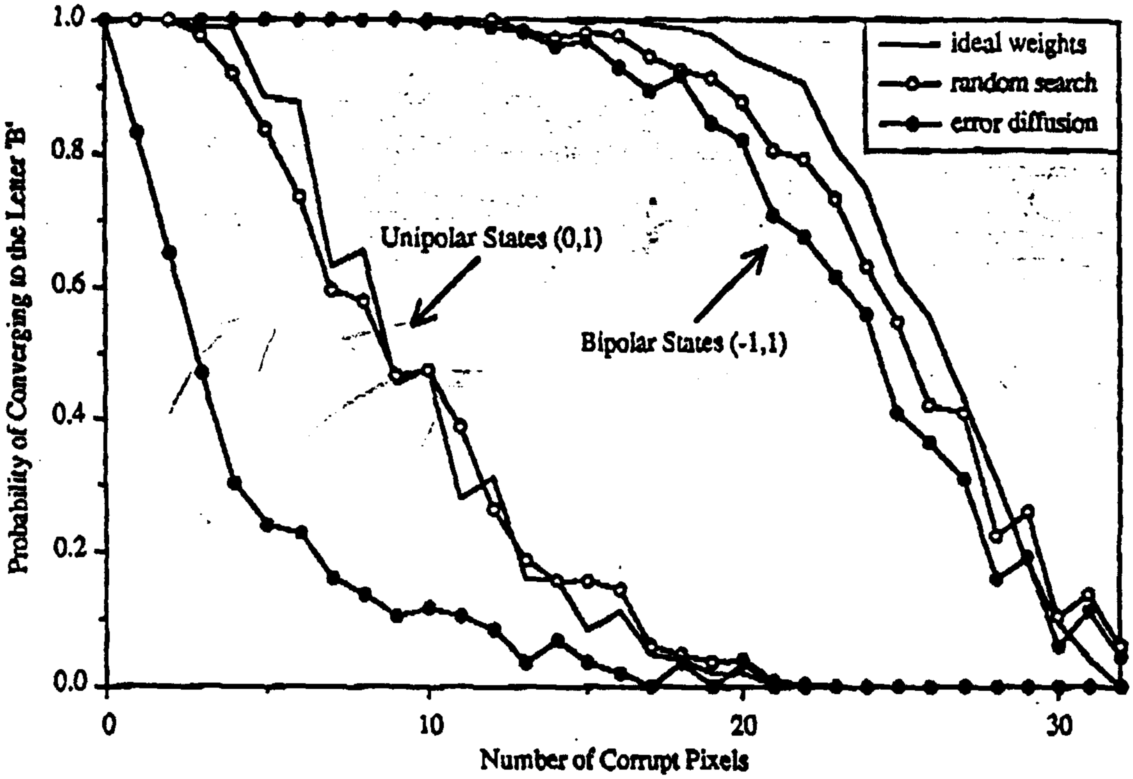


Figure 4: Performance of the ideal associative memory and the opto-electronic implementations. The graph plots the probability of convergence of the network to the correct state versus the number of corrupted pixels.

6. Conclusions

We have demonstrated that a system employing planar holographic optical interconnects can be used to implement a neural network architecture and that the performance of an optically implemented Hopfield style network comes close to that of an arbitrary system employing ideal interconnect weights.

This research is supported by the Optical Circuitry Cooperative at the University of Arizona.

7. References

1 J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. Natl. Acad. Sci. USA 79, 2554-2558 (1982).

2 S. Weissbach, F.Wyrowski, and O. Bryngdahl, "Quantization Noise in Pulse Density Modulated Holograms," Opt. Comm. 67, 167-171 (1988).

3 M. Kim, M. Feldman, and C. Guest, "Optimum encoding of binary phase-only filters with a simulated annealing algorithm," Optics Letters 14, 545-547 (1989).

4 A. Gmitro, P. Keller, and G. Gindi, "Statistical Performance of Outer-Product Associative Memory Models," Appl. Opt. 28, 1940-1948 (1989).

5 Summary of Personnel and Publications

Below is a list of professional personnel associated with this grant:

- G.R. Gindi, Associate Prof. Diagnostic Radiology and Electrical Engineering,Yale U.
- A.F. Gmitro, Assistant Prof. Diagnostic Radiology and Optical Sciences, U. of Arizona
- John Moody, Associate Research Scientist, Depertment of Computer Science, Yale University
- Joachim Utans, Graduate Student, Department Electrical Engineering, Yale U.
- Volker Tresp, Graduate Student, Department Mechanical Engineering, Yale U.
- Kannan Parthasarathy, Graduate Student, Department of Electrical Engineering, Yale U.
- Paul Keller, Graduate student , Optical Sciences Center, University of Arizona.
- Tony Zador, Graduate Student, Section of Neuroanatomy, Yale U.
- Grant Shumaker, Medical Student, Yale University School of Medicine

Below is a list of publications resulting from this grant:

- G Gindi, E Mjolsness, P Anandan: "Neural Networks for Model-Based Recognition", in *Neural Networks: Concepts, Applications and Implementations*, Prentice – Hall (in press).
- V Tresp and G Gindi: "Invariant Object Recognition by Inexact Subgraph Matching with Applications in Industrial Part Recognition", Proc. Intl Neural Net Conference, Kluwer Academic Publishers, Dorrecht/Boston/London pp. 95 – 98, 1990.
- G.R. Gindi, A.F. Gmitro, and K. Parasarathy, "Hopfield Model Associative Memory with Nonzero Diagonal Terms in Memory Matrix", Applied Optics, **37**, 129-135 (1988).
- A.F. Gmitro, P. E. Keller, and G.R. Gindi, "Statistical Performance of Outer-Product Associative Memory Models", Applied Optics, **28**, p.1940 - 48 (1989).
- G Gindi, E Mjolsness, and P Anandan: "Neural Networks for Model Matching and Perceptual Organization", in *Advances in Neural Information Processing Systems I*, pp. 618-626, D. Touretzky, ed. Morgan Kaufman, San Mateo,1989.
- Paul Keller and A.F. Gmitro, "Design and Demonstration of an Opto-Electronic Neural Network using Fixed Planar Holographic Interconnects", to appear OSA Topical Meeting an Optical Computing 1990.
- T. Zador, G. Gindi, E. Mjolsness, P. Anandan, "Neural Network for Model based Recognition - Simulation Results", Proc. Intl. Conference on Neural Networks, p.532, Boston 1988.
- E. Mjolsness, G. R. Gindi, P. Anandan, "Objective Functions for Visual Recognition: A Neural Network that Incorporates Inheritance and Abstraction", Proc. Snowbird Conference on Neural Networks for Computing, Salt Lake City, Utah 1988.

- E Mjolsness, G Gindi, and P Anandan: "Optimization in Model Matching and Perceptual Organization", *Neural Computation* (invited paper) vol. 1, pp. 218 – 229, 1989.
- E. Mjolsness, G.R. Gindi, and P Anandan, Objective Functions in Recognition",AAAI Spring Symposium on on Physical and Biological Approaches to Computational Vision" (Stanford March 1988) pp. 105-107

Below is a list of abstracts and technical reports associated with this grant:

- J Utans, G Gindi, E Mjolsness and P Anandan: "Neural Networks for Object Recognition within Compositional Hierarchies: Initial Experiments" Yale University Dept Electrical Engineering, Center for Systems Science, TR 8903, February 1989.
- E. Mjolsness, G.R. Gindi, and P Anandan, "Optimization in Model Matching and Perceptual Organization" Technical Report YALEU/DCS/RR-634 Yale U. Dept Computer Science, June 1988
- A.F. Gmitro and P. Keller, "Space-Variant Optical Interconnects via Multifaceted Planar Holograms", presented at Optical Socciety of America Annual Meeting, Santa Clara CA, October 1988.
- J. Moody and C. Darken, "Learning with Localized Receptive Fields", Tech Report YALEU/DCS/RR-649, Yale U.Dept Computer Science, Sptember 1988.
- J. Moody and C. Darken, "Fast Learning in Networks of Locally-Tuned Processing Units", Tech Report YALEU/DCS/RR-654, Yale U.Dept Computer Science, Sptember 1988.
- J. Moody, "Optimal Architectures and Objective Functions for Associative Memory: A Summary" internal report (available on request) Yale U.Dept Computer Science, January 1989.
- G Shumaker, G Gindi, E Mjolsness and P Anandan: "STICKVILLE: A Neural Net for Object Recognition via Graph Matching", Yale University Dept Electrical Engineering, Center for Systems Science, TR 8908, April 1989.
- G. Shumaker, "A Stickville Capable of Rotational Projection: Rotatorberg ", Technical Report, Yale University Center for Systems Science

References

[1] G Gindi, E Mjolsness, P Anandan: "Neural Networks for Model-Based Recognition", in *Neural Networks: Concepts, Applications and Implementations*, Prentice – Hall (in press).

[2] G.R. Gindi, A.F. Gmitro, and K. Parasarathy, "Hopfield Model Associative Memory with Nonzero Diagonal Terms in Memory Matrix", Applied Optics, **37**, 129-135 (1988).

[3] V Tresp and G Gindi: "Invariant Object Recognition by Inexact Subgraph Matching with Applications in Industrial Part Recognition", Proc. Intl Neural Net Conference, Kluwer Academic Publishers, Dorrecht/Boston/London pp. 95 – 98, 1990.

[4] A.F. Gmitro, P. E. Keller, and G.R. Gindi, "Statistical Performance of Outer-Product Associative Memory Models", Applied Optics, **28**, pp. 1940 - 48, (1989).

[5] G Gindi, E Mjolsness, and P Anandan: "Neural Networks for Model Matching and Perceptual Organization", in *Advances in Neural Information Processing Systems I*, pp. 618-626, D. Touretzky, ed. Morgan Kaufman, San Mateo,1989.

[6] T. Zador, G. Gindi, E. Mjolsness, P. Anandan, "Neural Network for Model based Recognition - Simulation Results", Proc. Intl. Conference on Neural Networks, p.532, Boston 1988.

[7] E. Mjolsness, G. R. Gindi, P. Anandan, "Objective Functions for Visual Recognition: A Neural Network that Incorporates Inheritance and Abstraction", Proc. Snowbird Conference on Neural Networks for Computing, Salt Lake City, Utah 1988.

[8] E. Mjolsness, G.R. Gindi, and P Anandan, "Optimization in Model Matching and Perceptual Organization" *Neural Computation* Vol. 1 , pp. 218 - 229, May 1989

[9] E. Mjolsness, G.R. Gindi, and P Anandan, "Objective Functions in Recognition",AAAI Spring Symposium on on Physical and Biological Approaches to Computational Vision" (Stanford March 1988) pp. 105-107

[10] J. Utans, G. Gindi, E. Mjolsness, and P. Anandan "Neural Networks for Object Recognition within Compositional Hierarchies", Technical Report, Yale University Center for Systems Science, TR 8903, February 1989.

[11] Paul Keller and A.F. Gmitro, "Design and Demonstration of an Opto-Electronic Neural Network using Fixed Planar Holographic Interconnects", to appear OSA Topical Meeting an Optical Computing 1990.

[12] E. Mjolsness, G.R. Gindi, and P Anandan, "Optimization in Model Matching and Perceptual Organization" Technical Report YALEU/DCS/RR-634 Yale U. Dept Computer Science, June 1988

[13] A.F. Gmitro and P. Keller, "Space-Variant Optical Interconnects via Multifaceted Planar Holograms", presented at Optical Society of America Annual Meeting, Santa Clara CA, October 1988.

[14] J. Moody and C. Darken, "Learning with Localized Receptive Fields", Tech Report YALEU/DCS/RR-649, Yale U.Dept Computer Science, September 1988.

[15] J. Moody and C. Darken, "Fast Learning in Networks of Locally-Tuned Processing Units", Tech Report YALEU/DCS/RR-654, Yale U.Dept Computer Science, September 1988.

[16] J. Moody, "Optimal Architectures and Objective Functions for AssociativeMemory: A Summary" internal report (available on request) Yale U.Dept Computer Science, January 1989.

[17] G Shumaker, G Gindi, E Mjolsness and P Anandan: "STICKVILLE: A Neural Net for Object Recognition via Graph Matching", Yale University Dept Electrical Engineering, Center for Systems Science, TR 8908, April 1989.

[18] G. Shumaker, "A Stickville Capable of Rotational Projection: Rotatorberg", Technical Report, Yale University Center for Systems Science