Optical Memory Disks in Optical Pattern Recognition Systems

Thesis by

Mark A. Neifeld

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1991

(Submitted December 1990)

Acknowledgements

I am grateful to my thesis advisor, Professor Demetri Psaltis for his insight, his guidance, and his patience (some would say tolerance). I, like most other Ph.D. candidates, during my tenure at Caltech had more than a few moments of doubt and frustration. Demetri's support was always there. Without his creativity and his leadership, this work would not have been possible.

Very special thanks are due Mr. Seiji Kobayashi of Sony corporation and Mr. Alan Yamamura (Yama, soon to be Professor Yama) without whom the optical disk-based architectures in this thesis would be just so many nice ideas. Had it not been for Seiji, who built Rocky and brought it from Japan, and Alan who is largely responsible for the characterization of that system and from whom I gratefully borrowed more than a little data for Chapter 3, I could have spent far more time during the last five years at the beach. I thank them both for their friendship and their support.

The members of Demetri's group both past and present have provided me with an open friendly work environment in which new ideas can be freely exchanged and developed. I thank Professors Ken Hsu, David Brady, Nabeel Riza, R. Scott Hudson, Robert Snapp, Fai Mok, Jeff Yu, and Cheol Hoon Park for many fruitful and enlightening discussions. The basketball was fun too. I would especially like to thank Dr. Eung Gi Paek for spending many a late night to early morning helping me to align my optics. The present members of Demetri's group from whom I have received friendship and support include Chuanyi Ji, Steven Lin, Yong Qiao, Subrata Rakshit, Sidney Li, Donald Lie, and David Marks. Special thanks go to Charlie Stirk for encouraging my bad

habits and to Ruth Erlanson for keeping the basement in baked goods. For endless patience and support in those administrative tasks for which I am not particularly adept (this excludes very few) I must thank Mrs. Helen Carrier and Mrs. Su Mc Kinley.

Lastly and most importantly I'd like to thank my parents Malcolm and Ida Neifeld. They have supported me for 27 years with their unconditional love and their guidance. Their faith in me and the direction I have chosen has been unwavering and without their constant encouragement the following pages would surely have remained blank.

Abstract

We describe the use of optical memory disks in optical pattern recognition systems. Algorithmic and architectural issues associated with the realization of such systems are discussed. Experimental demonstrations of several optical disk-based architectures are included to aid in the understanding of system limitations and performance issues. First we discuss correlation-based pattern recognition and describe the relationship between this approach and the neural paradigm. The need for invariances in image recognition leads to the notion of the reference image library. This approach is shown to be attractive in the case of limited processor and spatial light modulator dynamic range. We characterize the optical disk as a parallel readout device. An overview of optical storage media is included. Parallel readout of data from Sony sampled format media is characterized. We identify a match between the characteristics of the optical disk and the requirements for pattern recognition systems. Four optical disk-based image correlators which may serve as building blocks in disk-based pattern recognition systems are introduced. These image correlators are experimentally demonstrated and compared in terms of speed, efficiency, and sensitivity to noise sources and disk imperfections. We discuss advantages and limitations of these systems.

We include a discussion of learning and generalization in neural networks. We present a new learning algorithm and discuss its generalization characteristics. Three disk-based systems for pattern recognition are proposed. The first is a correlation-based architecture. The performance of this system as compared with theoretical expectations is encouraging; however, data rate constraints suggest the investigation of an alternate approach. The next two systems are more

neurally inspired and realize the k-nearest neighbor and radial basis function algorithms. An evaluation of the performance of these two systems is presented with respect to the handwritten digit recognition problem.

Lastly, we present two candidates for future optoelectronic computing and pattern recognition systems. We detail the operation of these architectures and discuss the need for a better understanding of the relationship between mass memory and a general parallel processing environment.

Table of Contents

1. Introduction

1.1 Optical Computing, Neural Networks and Optical Disks

A great deal of effort today is being devoted to the understanding and development of complex multiprocessor-based systems or parallel computers.[1–5] We can expect the next generation of electronic computing technology to be based on such a parallel computing paradigm; however, the ultimate capabilities of such multiprocessor systems can be strongly dependent on effective interprocessor communications. Since the electromagnetic interaction among nearby electrons precludes the possibility of high density noninterfering electronic communication channels, optics becomes a viable candidate technology for use in these next generation computing systems. Photons do *not* interact appreciably in vacuum and therefore can provide extremely high density interconnection capabilities for both chip to chip and board level multiprocessing schemes.[6–13]

Since the principal strength of optics is in its ability to facilitate high density communications, future application of optical computing technologies will be most natural where we find communications intensive processing. A neuronal information processing system is such a communications intensive environment.[14–17] The study of neural computing or neural networks as these systems are often called, is inspired by biological computing techniques. Biological neural networks are characterized by a large number ($> 10^{10}$ in the human brain) of relatively simple processing elements or neurons, and a large degree of connectivity ($> 10,000$ connections per neuron in human brain). Connections among neurons are made through synapses which determine the influence the output of one neuron will

have on the input to another. In addition to the strength of the synaptic connections between neurons, the topology of the dendritic tree (i.e. the wires in the nervous system) is believed to play an important role in defining the interaction among neuronal signals.[18,19] It is commonly held that the *program* for our biological neural computer or equivalently, the memories that have been stored in the neural network, reside in the synapse strengths. The distributed nature of such a storage scheme together with the redundancy present in such a largely parallel system is believed to result in the apparent robustness exhibited by biological neural computers. Only recently have many details of biological neural systems become available owing primarily to improved experimental techniques, yet the vast majority of complexities associated with the detailed structure/function of these computers are still unknown or only poorly understood. In order to facilitate the study of a neural computing paradigm therefore, we grossly oversimplify the behavior of the biological components to arrive at a simple working model. These simplifications are akin to assuming a spherical cow, and only time will tell whether such models will exhibit behavior like that of their biological cousins; however, the richness of the collective behavior exhibited by even our simple models is today, worthy of further study. The details of these neural models are discussed in Chapter 2 and will not be presented here. We will point out however, that the highly connected character of neural networks make electronic realizations of such systems extremely difficult while making an optical implementation quite natural.[20–26]

Why will we study neuronal information processing? Aside from a fundamental interest in the workings of the human brain, one of the principal motivations

for studying artificial neural networks is the hope that such systems will provide useful solutions to problems that have eluded solution via conventional computing techniques. The human brain is particularly adept at solving pattern recognition and associative tasks. It is our hope that through simulation of a brainlike architecture, some of the capabilities of these biological systems will manifest themselves in our models. We hope that through *learning* in a neural framework, collective behavior will emerge that facilitates the solution to pattern recognition tasks at which *programming* and conventional computing has failed. Putting the above discussions into perspective for our purposes it is clear that *optical neural networks for pattern recognition* is a natural area of both academic as well as practical interest.[27–31]

As the title indicates, we will study the use of optical memory disk technology in the realization of these neural pattern recognition systems. Optical disks represent a mature storage technology which optical computing systems in general may use to advantage. [25–38] One issue associated with the realization of a general optical computing system is the need for an *optical* interface to the outside world (I/O) and memory. Spatial Light Modulator (SLM) technology has matured in the past few years to provide several realistic options for real-time optical I/O capabilities; however, work in this area is still required so that SLM device characteristics may be better matched to optical computing system needs. Optical disks provide the other required capability with a natural optical interface to mass memory. The conventional mode of access to optical storage is serial and represents a severe limitation in terms of data transfer rate and suitability to optical processing schemes so that the approach we will investigate

here is the parallel access mode. Large data rates and a natural mechanism for the storage of data in two-dimensional format is facilitated by the use of optical disks in this mode. By the end of Chapter 6 we will have seen how optical disk characteristics make this optical storage technology one that is well suited to the optical implementation of neural pattern recognition systems.

1.2 Thesis Overview

In this thesis we will describe the use of optical memory disks in optical pattern recognition systems. Algorithmic as well as architectural issues associated with the realization of such systems will be discussed. Experimental demonstrations of several optical disk-based architectures will be included to aid in the understanding of system limitations and various performance issues. In Chapter 2 we discuss correlation-based pattern recognition and describe the relationship between this approach and the neural paradigm. The need for invariances in image recognition leads to the use of the reference image library in such correlation-based systems. This approach is shown to be attractive from the perspective of limited processor and SLM dynamic range. The optical disk is proposed to provide the necessary storage for such an image library. At the end of this chapter we demonstrate a simple incoherent optical image correlator based on acoustooptic technology. This image correlator is used to simulate a single shift invariant neuron or perceptron.

In Chapter 3 we discuss the characteristics of optical disks. An overview of optical storage media as well as experimental characterization of some commercially available media is included here. The disk system used in most of our

work is a Sony sampled format disk drive which is capable of reading and writing both Write Once Read Many (WORM) and MagnetoOptic (MO) media. Parallel readout of data from the Sony media will be fully characterized in terms of phase uniformity, diffraction efficiency, contrast, etc. In this chapter we will see a match between the characteristics of the optical disk and the requirements for library-based pattern recognition systems. Chapter 4 introduces four optical disk-based image correlators which may serve as building blocks in disk-based pattern recognition systems. These image correlators are experimentally demonstrated and compared in terms of speed, efficiency, and sensitivity to various noise sources and disk imperfections. We discuss advantages and limitations of these systems.

In Chapter 5 we begin with a discussion of learning and generalization in neural networks. We present a new learning algorithm and discuss its generalization characteristics with respect to certain classes of problems. We then present three disk-based systems for pattern recognition. The first is a shift invariant correlation-based architecture which we demonstrate using a three-class handwritten character recognition task. The performance of this system as compared with theoretical expectations is encouraging; however, data rate constraints suggest the investigation of an alternate approach. The next two systems represent this alternative approach as they are more neurally inspired. These two disk-based systems realize the k-nearest neighbor and radial basis function algorithms for pattern classification. A comparison between the performance of these two systems and the expected performance of the algorithms is presented with respect to the 10 class 0-9 handwritten digit recognition problem. Chapter 5 concludes

with a discussion of the robustness of these optical disk-based pattern classification systems.

In the last chapter we present a discussion of potential candidates for future optoelectronic computing and pattern recognition systems. We detail the operation of two suggested architectures and discuss the need for a better understanding of the relationship between mass memory and a parallel processing environment. We see that through the incorporation of feedback and the use of Read-Write-Erase media, interesting optical disk-based systems for neural computing and pattern recognition may be realized.

2. Optical Pattern Recognition

In this thesis, *Pattern Recognition* (PR) system will refer to an automated system which for any input provides one of a finite number of classifications or labels as an output. The outputs of the PR system should correspond to the identity of some "meaningful regularity" which is present in the input environment.[39] We should note that PR in this sense differs from the related field of image understanding in that the output of a system of the later type is a *description* of the input scene including the spatial and temporal relationships among the objects present. Our PR system in contrast is designed to identify or to sometimes merely indicate the existence of familiar objects and as such may serve as a building block in a knowledge-based image understanding system. The canonical PR system consists of three parts as shown in Figure 2.1.

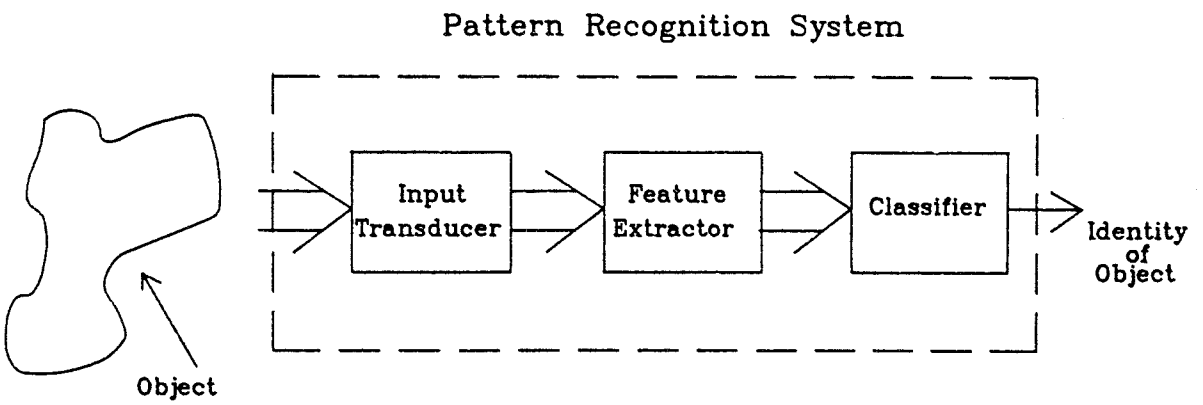


Figure 2.1 : Block diagram of canonical pattern recognition system.

The first block shown is the input transducer which converts information present in the observable world, such as electric field intensity or acoustic field amplitude, into a form suitable for the rest of the PR system, typically an electrical signal. The feature extraction stage can be considered the most important component of the PR system since it is assigned the formidable task of projecting a potentially high-dimensional input space (10^6 dimensional for a 1000×1000 pixel image) onto a manageable number of *relevant* feature dimensions. Since this task is quite problem dependent, the design of the feature extractor remains more of an art than a science although some recent neural approaches to PR have yielded interesting results (i.e. useful features) in problem domains such as speech and handwritten character recognition.[40–42] There is clearly a trade off in complexity between the feature extractor and the classifier. The classifier is responsible for operating on a feature vector to perform the desired identification of the input. This task can be made trivially simple through the use of an omnipotent feature extractor or impossibly difficult in the case of an unreliable one. Since we will be primarily interested in implementational issues associated with the realization of optical PR systems, we will make some specific assumptions regarding the three components described above. Firstly, the input transducer for our optical PR systems will almost always be some form of Spatial Light Modulator (SLM). This is because our interest here is primarily in the realization of *optical* PR systems. The SLM is a key device in such systems since it facilitates the conversion of an electrical signal or an incoherent optical signal, into an optical signal suitable for further processing. Furthermore, we will lump the feature extraction and the classifier stages into a single optical processor so that simple optical systems

may be demonstrated. For the image recognition tasks we will be describing, this approach is almost certainly suboptimal; however, the resulting systems will demonstrate the feasibility of more sophisticated architectures.

A very simple approach to pattern classification is the look up table approach. If for each possible input to our PR system we store the appropriate output response, then the process of recognizing a scene would reduce to a search of a large memory or database of inputs. It is the required memory size that makes this approach prohibitive. For input images of 1000×1000 binary pixels, the required look up table size would be 2^{10^6}. This is clearly an unrealistic solution and most of these images are meaningless for the PR task at hand. The alternative approach is to find a compact representation of the mapping to be performed. This representation can be descriptive as for the case of syntactic approaches or it can be probabilistic as with decision theoretic or statistical pattern classification.[39,43,44] In both of these approaches, the difficulty lies in providing the PR system with the necessary invariances to potentially distorted input features. PR systems most often operate on natural inputs such as a visual field or an acoustic signal. Such natural inputs can exhibit a wide degree of variability without changing their essential identity. As an example, consider adding noise to a photograph of your grandmother or changing the lighting conditions under which the photo was taken. Within reason, these variations in the input should not inhibit your ability to recognize her face. A system that can recognize an input object in the presence of such variability is said to exhibit invariance with respect to that variability. A system for image recognition for example, would do well to exhibit shift, scale, rotation and illumination invariance as well as other

potential invariances that will depend on the particular task at hand.[45–48] Some of these invariances, such as the first three mentioned above, can be described in a concise mathematical form, whereas others, such as hair length invariance in grandmother recognition, cannot. Although designing PR systems that exhibit or *learn* to exhibit such invariances is a very difficult task, it is also the most interesting problem associated with real world PR.

Perhaps the simplest form of invariance a PR system can exhibit is noise tolerance. It is well known from detection theory that the optimal PR system for a known signal in stationary noise is the correlator. In signal processing, one is often concerned with *white* noise in which case the required correlation filter becomes the so-called matched filter. The matched filter is simply a replica of the object of interest that is compared against a received signal in order to detect the presence of that object. The correlation function in one dimension can be defined equivalently in two ways :

$$c(\tilde{x}) = \int_{-\infty}^{\infty} f(x)g^*(x-\tilde{x})dx \tag{2.1}$$

$$= \mathcal{F}^{-1}\{\mathcal{F}\{f(x)\}[\mathcal{F}\{g(x)\}]^*\}, \tag{2.2}$$

where $f(x)$ is the input function, $g(x)$ is the reference function and $\mathcal{F}\{\}$ represents the Fourier transform operation. Note that although we will be dealing with two-dimensional images throughout most of this thesis, in most expressions we will present only one dimension, the second being implied. The first of the above equations indicates that the correlation function $c(\tilde{x})$ at some point $\tilde{x}$ is simply the inner product between the input image $f(x)$ and a shifted version of the

reference image, where the required shift is just $\tilde{x}$. Further, the peak of the normalized correlation function that we can define as :

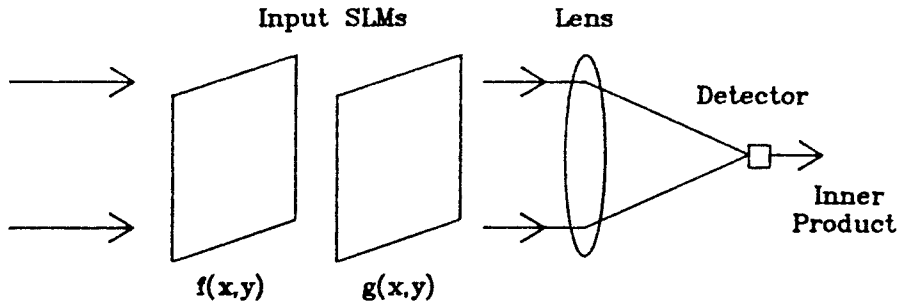
$$c_{max} = \max_{\tilde{x}}[\overline{c(\tilde{x})}] \tag{2.3}$$

$$= \max_{\tilde{x}} [\frac{c(\tilde{x})}{\sqrt{\int |g(x)|^2 dx \int |f(x)|^2 dx}}], \tag{2.4}$$

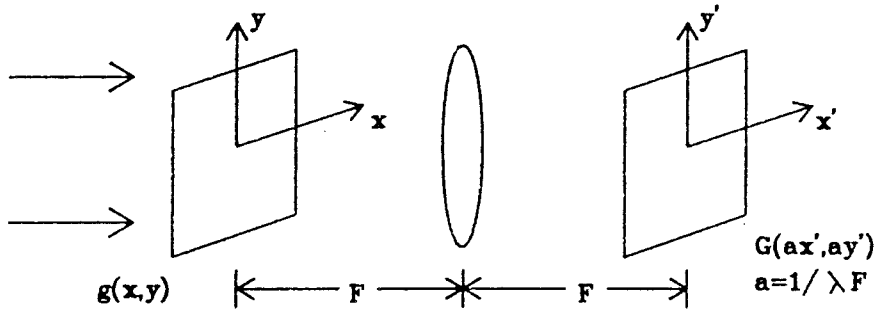
will be the largest when the input and reference images are equal. That is,

$$\max_{f(x)}[c_{max}] = 1 \qquad iff \quad f(x) = g(x). \tag{2.5}$$

So, regardless of the position of an object in the input scene, the normalized correlation function will have a peak of $\overline{c(\tilde{x})} = 1$ when the reference image is *matched* to the input object. This *shift invariance* coupled with its desirable noise tolerance, makes the correlation function a powerful tool in image recognition. Perhaps more importantly for this work, the correlation function is very easily realized optically. This is due to both the ease with which optics can perform inner products as well as the ease with which optics can perform 2D Fourier transforms. The two optical systems required to perform these two operations are shown in Figure 2.2. In Figure 2.2a we show two SLMs on which the images $f(x, y)$ and $g(x, y)$ have been recorded. The two SLMs are adjacent so that when illuminated as shown, the pixel by pixel product of $f(x, y)$ and $g(x, y)$ will appear immediately to the right of the two SLMs. This product pattern is integrated by the lens and appears as an electrical signal after the detector. Figure 2.2b depicts the optical 2D Fourier transform system in which a single lens computes the transform of the input function $g(x, y)$. The result appears in the back focal



(a)



(b)

Figure 2.2 : Operations suited to direct optical implementation.

(a) Optical system used to compute inner product.

(b) Optical system used to compute 2D Fourier transform.

plane of the lens as shown and is scaled by the factor $a = 1/\lambda f$. Until recently, optical PR has been almost exclusively correlation-based.[49–51]

The reason correlation-based optical PR systems have not seen widespread usage in the solution of real world problems is that the combinatorial explosion associated with the number of reference images required to achieve the many desired invariances has been prohibitive. That is, a large library of reference images must be compared against the input in order to realize a useful system. Implementation of this image library requires large storage capacity to provide the necessary invariances (scale, rotation, etc.), programmability to provide novel filter design capabilities, and a mechanism through which the stored images may be read into the high speed optical correlator. We will see later in this thesis that the optical memory disk is a useful technology to utilize in the realization of the required reference library.

2.1 Neural Networks

The relative ease with which humans can perform sophisticated PR tasks is one of the primary motivations for the study of artificial neural networks. Even the common house cat can surpass conventional PR systems in terms of both utility and robustness. The nervous systems of even very simple organisms however, are such incredibly complex computing engines whose structure/performance is only very primitively understood today, that the artificial network models we study are gross simplifications of their biological counterparts. Even in these simplified systems though, there is interesting behavior which emerges as a result of the collective nature of the computing process. There are three characteristics of

biological Neural Networks (NNs) that our artificial models share and which distinguish the neural computer from more conventional digital approaches. Firstly, the NN is a computational paradigm which exploits a large degree of parallelism. The human brain contains more than 10^{10} processing elements or *neurons* each of which is connected to as many as 200,000 others forming a highly interconnected network of relatively simple processors. In analogy to this biological system our NN models consist of many (e.g., $10^4 - 10^5$ or more) processing elements whose function is often taken to be that of a simple thresholding device and each of these simple processors might communicate with up to 100-1000 others. This is in contrast to conventional *parallel* computers in which relatively few complex microprocessors are arranged so that each communicates with only several others. The second departure from conventional computing arises from the analog and distributed nature of information storage in NNs. In a neural computer, stored data or memories reside in the analog connection strengths or synapses of the network and may be retrieved by observing the state of activation that is present over the array of neurons in response to some input. Since in general this pattern of activation will depend on all of the synapse strengths, the information stored in the network is said to be distributed. For this reason, NNs are robust to perturbations of or damage to individual interconnections since destroying a single synapse will not tend to destroy any particular stored memory but will likely degrade slightly the recall of all of the stored data. This is in contrast with conventional digital RAM in which damage of a single transistor could be catastrophic. The third aspect of NNs that distinguishes them from conventional computers is the aspect of learning. In NNs, the output, which is represented

as the activation over some subset of neurons (output neurons) in response to an input that is often represented as an external signal received by a subset of neurons (input neurons), is defined by the synapse strengths. Note here that the network output may be defined not only by a particular pattern of activation at one time but could also be defined as the dynamic behavior of the output neurons over some period. In either case, in order to choose the weights appropriately to define some desired I/O relationship, the network is shown examples of the mapping and a learning algorithm is used to evolve a set of connection strengths to cause the network to generate desirable outputs. This is called the learning or training phase. This is a convenient method of *programming* a machine to perform a PR task since we do not have an algorithm that defines the task to begin with. As far as we are concerned, the task is *defined* by examples. It is certainly true that we, as biological computers, *learned* to recognize our own grandmothers. The success of any training algorithm is measured in two ways. First, the network that results from the learning phase should perform well on the training data. That is, for a PR task the trained network should correctly classify those inputs on which it was trained. Also, however, we would like the NN to perform well on other inputs that are in some sense (defined by the problem at hand) similar to the training inputs. This second performance criterion is called generalization. Although good generalization is the universal goal of NN learning theory, no complete theory of learning and generalization has emerged to date.[52–60] This is primarily due to the complexity of the relationship among the network structure, the learning algorithm and its suitability to some particular class of problems, and the use of *a priori* information in the training phase.

Later in this thesis we will address some of these issues in connection with a new learning algorithm for NNs that is found to yield good generalization for certain types of problems.

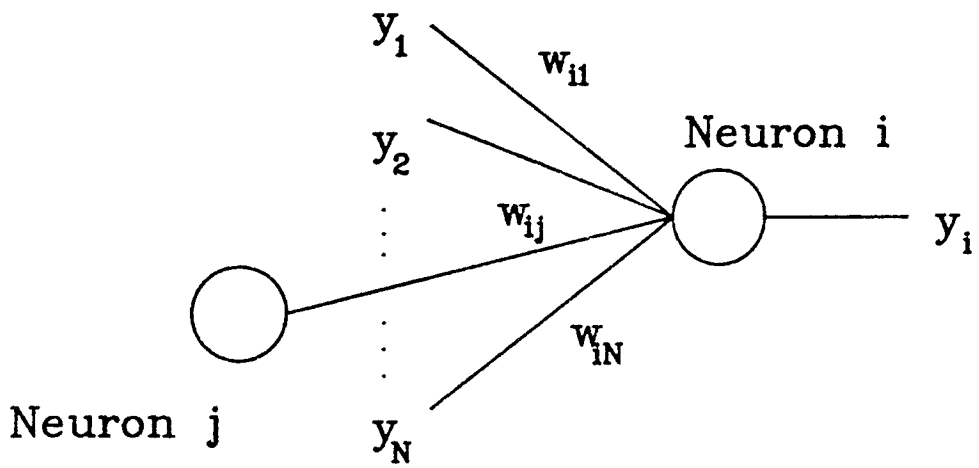


Figure 2.3 : Simple model of a single neuron.

In order to provide a brief overview of NN-based PR we will begin with the simplest of NNs, a single neuron. Our model of the neuron is shown in Figure 2.3 and consists of two parts : a weight vector and a threshold function. The synapse strengths w_{ij} determine how the output of neuron j will affect neuron i as

$$x_i = \sum_{j=1}^{N} w_{ij} y_j \tag{2.6}$$

where x_i is the input to neuron i, w_{ij} are the connection strengths shown in the figure and y_j are the outputs of the N neurons to which neuron i is connected.

The output of neuron i, y_i is determined from its input x_i as

$$y_i = \Theta(x_i), \tag{2.7}$$

where Θ is often taken to be a soft threshold function such as

$$\Theta(x) = tanh(\gamma x) \tag{2.8}$$

or

$$\Theta(x) = \frac{1}{1 + e^{-\gamma x}}. \tag{2.9}$$

The parameter γ is called the gain of the neuron. It is interesting to note that a single neuron with infinite gain (i.e. a hard thresholding device) is the optimal PR system in the Bayesian sense, for a certain class of problem. The resulting classifier is the linear discriminant function. This fact is interesting because it provides a mathematical link between neural models, which are only simple-minded approximations to the biological PR systems that we aspire to mimic, and traditional statistical PR. Beyond this connection, the single neuron network is interesting from the perspective of learning as well. The weight vector $\underline{w}$ associated with a single neuron can be trained using an algorithm such as the perceptron, which is an iterative procedure that is guaranteed to converge.[39] For this reason a single neuron is sometimes referred to as a perceptron.[61–64] The existence of the perceptron algorithm implies that if a PR problem can be solved using a single neuron, then the weight vector required to solve that problem can be found in a simple way. Problems of this sort are called linearly separable

problems. Furthermore, since the principal operation involved in computing y_i is an inner product

$$y_i = \Theta(\underline{w} \cdot \underline{y}), \tag{2.10}$$

optics is a good technology for the realization of such a PR system. In fact, since optics is particularly well suited to performing image correlations, we can define a slightly more useful PR system for image applications as the Shift Invariant Neuron (SIN). The SIN is simply a single neuron in which the inner product operation has been replaced with a full 2-D correlation followed by peak detection. The output of a SIN, therefore, is a single value representing the thresholded peak of the discrete correlation $c_{\tilde{i}}$ where

$$c_{i\tilde{i}} = \sum_{i=-N}^{N} w_{ij} y_{j-\tilde{i}}. \tag{2.11}$$

The SIN input therefore is given by

$$x_i = \max_{\tilde{i}}[c_{i\tilde{i}}], \tag{2.12}$$

and the corresponding SIN output can be written as given by equation 2.7. It should be pointed out that with this simple extension of the concept of a single neuron, we have identified a connection between NN-based PR and optimal detection theory.

Having introduced the neuron, which is the basic building block of NNs, the next level of complexity to consider is an *array* of neurons, which will provide an *array* of output signals as shown in Figure 2.4. Such an array of SINs represents,

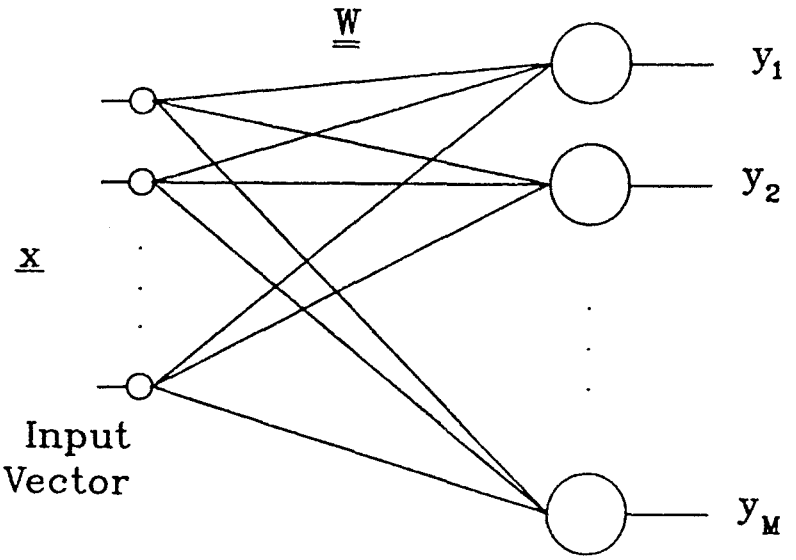


Figure 2.4 : Array of neurons with connection matrix $\underline{\underline{W}}$

in a neural framework, the correlation library-based approach discussed in the previous section. If lateral inhibition is added to the output layer of Figure 2.4, this system exactly implements a correlation-based nearest neighbor PR classifier. What is the motivation for using multiple filters as opposed to a single composite filter that we know may be calculated using some procedure akin to the perceptron algorithm? There are three reasons for considering multiple filters. First, a single neuron can only solve linearly separable problems. Since almost all interesting problems are not linearly separable, multiple neurons are required. Secondly, an optical implementation of a NN will require the use of SLMs, and most available SLMs are binary. It is possible that the filter that solves the linearly separable problem at hand requires analog weights. For this reason, formation of the optimal filter on an optical device may be impossible, thereby requiring the use of multiple filters. The third reason is also implementational in

origin and concerns the limited dynamic range available in optical storage media and optical computation itself. For descriptive purposes, consider the simple sum algorithm for computing a complex filter from M binary images drawn from two classes. The filter resulting from this algorithm is given by

$$\underline{G} = \sum_{i=1}^{M/2} \underline{g_i} \quad - \sum_{i=M/2+1}^{M} \underline{g_i}, \tag{2.13}$$

where the training vectors or prototypes $\underline{g_i}$ are N dimensional bipolar vectors distributed among the two classes Ω_1 and Ω_2 according to

$$\underline{g_i} \in \Omega_1 \quad \forall \quad i \leq M/2 \tag{2.14}$$

$$\underline{g_i} \in \Omega_2 \quad \forall \quad i > M/2. \tag{2.15}$$

Consider also an unknown input vector $\underline{f}$ which we will classify as belonging to Ω_1 or Ω_2 through the use of two different systems. System 1 is the single neuron system whose weights are given by $\underline{G}$ and system 2 is the reference library approach in which all M correlation peaks are computed and the class associated with the largest result is taken as the output. Without consideration of the actual classification performance of these two systems, we can calculate several measures of the cost to build such systems. In terms of storage, system 1 requires storage of N pixels, each with a maximum of $2M$ possible values yielding an upper bound on the storage of $\underline{G}$ equal to $S_G = Nlog_2(2M)$ bits. Similarly for system 2 we require storage of M templates, each comprising N bits yielding a storage requirement of $S_g = NM$ bits. If we consider these storage requirements as bounds on the space required to realize the associated system, then the

composite filter approach requires less space by a factor of $M/log_2(M)$ at the expense of requiring a storage medium with a dynamic range of $log_2(2M)$ bits or $2M$ levels. The time required to perform the necessary comparisons is equal for the two systems when system 2 uses spatial multiplexing to allow simultaneous calculation of all M correlations; however, the optical power requirements in this case must increase by a corresponding factor of M. Another difference between these two approaches is the required *processor* dynamic range (DR) for each. For system 1 each pixel of the output correlation pattern is the sum of $2N$ numbers, each with a DR of $4M$ levels. The optical system DR required for system 1 therefore is $log_2(8MN)$ bits. For system 2 however, each correlation peak output is due to $2N$ numbers each with a DR of only 1 bit which yields an output DR requirement of $log_2(4N)$ bits which is a factor of $log_2(2M)$ less than that required for the complex filter-based system. We know from our previous discussion that for interesting PR problems, M can be very large (≈ 1000), yielding an appreciable DR requirement for the optical processor of system 1. For an optical implementation therefore, in which SLM DR as well as the DR of the computation itself is limited, the reference library approach or equivalently, an array of neurons is preferable to an approach based on a single complex filter when we can afford a factor of $M/log_2(M)$ increase in storage requirement. We will see in later chapters that this is indeed the case for optical disk-based PR systems and therefore, the reference library approach will be adopted in preference to the composite filter approach. We should note here that a hybrid solution is also possible in which arrays of composite filters are used. This is useful for example when gray level images must be stored; however, this approach requires

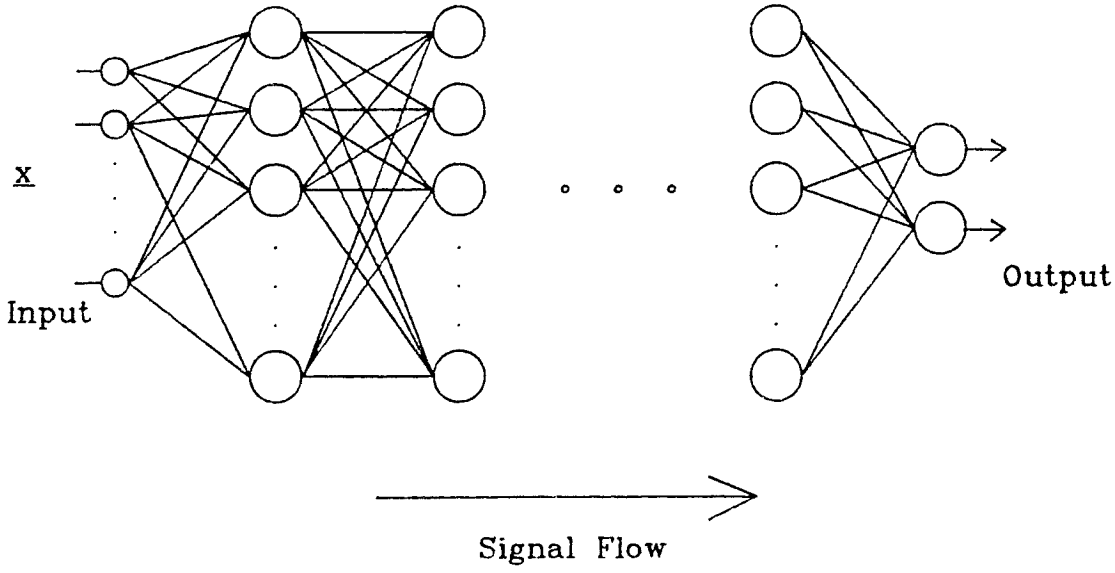


Figure 2.5 : Multilayer feedforward neural network.

both large capacity and dynamic range.

The array of neurons shown in Figure 2.4 can be thought to generate some alternate representation of the input scene on which other neurons may then operate. This process of designing layers of neurons separated by matrices of connections will result in the formation of a multilayer feedforward NN as shown in Figure 2.5. Deeper layers of such networks may consist of conventional neurons as well as SINs. In these NNs there is an identifiable signal flow through the system from left to right whereby each layer processes the output of the previous layer to arrive at a new representation of the input space until finally, at the output layer, a representation is achieved which represents the desired classification. The feedforward nature of these systems simplifies the possible network behavior considerably by eliminating the complicated dynamics that can arise in feedback

networks; however, learning in multilayer networks remains a formidable problem. Although recently there have been numerous heuristic algorithms proposed for training such networks, very few analytic results analogous to the perceptron convergence proof exist for these new training procedures.[15,57,65] Despite the mathematical difficulty of studying such algorithms, many of these have proven successful in training networks which have sometimes yielded interesting generalization results. Once again in connection with correlation-based approaches, if we construct the first layer of a multilayer network from SINs, then this layer can be viewed as a reference library with deeper layers serving to postprocess the resulting correlation peaks.

2.2 LED Bipolar SIN

Now we will demonstrate a simple correlation-based optical PR system that is modeled after the SIN approach described above. This system utilizes a single complex correlation filter formed using the perceptron algorithm. The task is to classify images viewed using a TV camera into one of two classes. One issue that arises in the optical implementation of such a PR system is the representation of inhibitory or negative synaptic strengths. Although a coherent optical system may represent bipolar signals using phase modulation, these systems must use interferometric detection if the bipolar signals are to be measured. This approach is somewhat cumbersome and a technique of preference is to operate the optical system with unipolar signals and to use electrical bias subtraction after detection to reveal the result of the equivalent bipolar computation.

A class of image correlators that are implemented with minimal use of 2-

D spatial light modulators and therefore maximize the use of mature device technologies and make the practical implementation of such systems easier, has been previously reported.[66–68] These architectures typically use acousto-optic (AO) input devices in time and space integrating systems to achieve the desired two dimensional correlation function.

In this section we describe a 2-D incoherent acousto-optic image correlator operating in the bipolar mode. In addition to the advantages associated with using the well developed AO technology, the present system is attractive in the sense that it is fully programmable. The programmability of this system provides a convenient method by which pattern recognition and classification algorithms may be realized. We begin with an overview of the bipolar correlator system operation. What follows is a discussion of the effect of input bias on bipolar correlation outputs. This sub-section is followed by experimental results obtained using the bipolar system. The final sub-section describes the results of using the bipolar system to implement a single SIN or shift invariant Linear Discriminant Function (LDF).

2.2.1 Bipolar System Operation

The basic incoherent correlator system architecture shown in Figure 2.6 consists of a 16-element vertical LED array in the reference plane, an acousto-optic device (AOD) in the input plane, and a CCD camera in the output plane. The associated optical system is shown in Figure 2.7. In the horizontal dimension this system is essentially responsible for imaging the LED array into the Bragg cell shown, and imaging the diffracted light once again onto the output CCD.

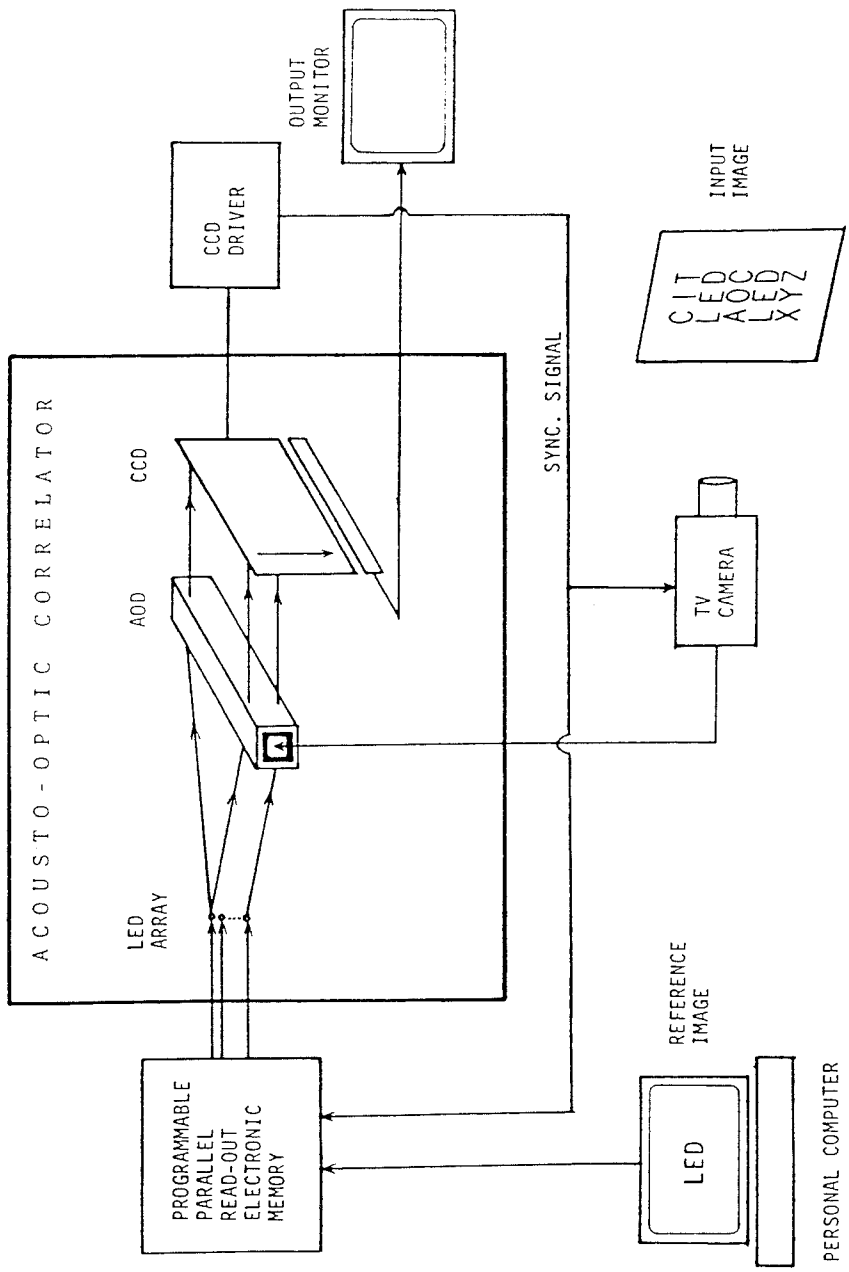


Figure 2.6 : Basic incoherent correlator architecture.

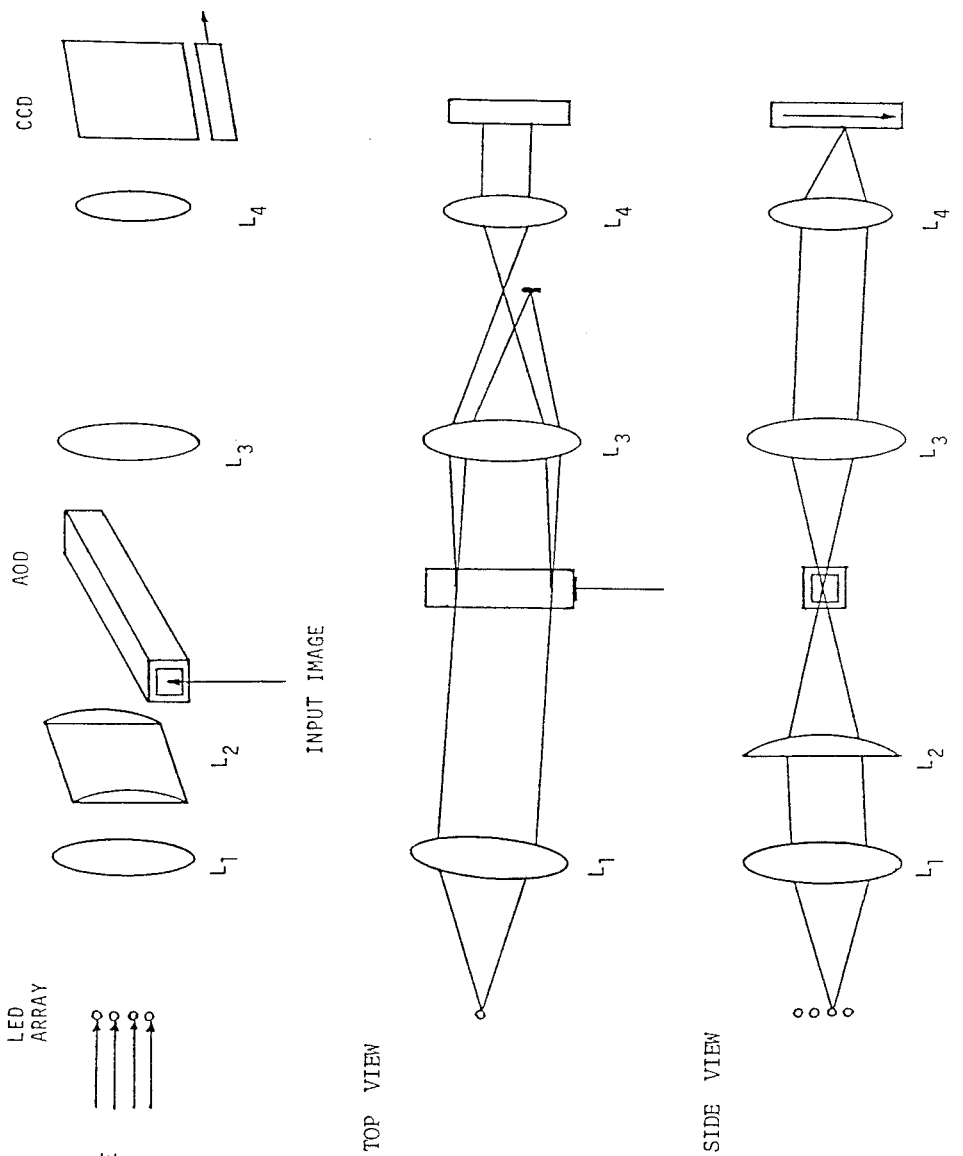


Figure 2.7 : Optical system used for incoherent image correlator

In the vertical direction, the light emitted from each LED is collimated so as to correspond to an entire line on the output CCD. The input scene is detected by a TV camera, and the video signal is applied to the AOD. An IBM PC is used to load reference image data into an electronic memory with multichannel readout capability. The data is read out to the LEDs, 16 nibbles at a time in such a way as to modulate the intensity of the ith LED with consecutive pixel values of the ith reference image line. A 2 bit D-A converter preceding each LED allows for representation of up to 4 distinct intensity levels per pixel. In the system reported here, three of these levels are used to represent an unbiased bipolar signal (-1,0,1) as a biased unipolar signal (0,1,2). The memory is read out cyclically in synchronism with the video signal from the TV camera. At the beginning of each horizontal TV scan, the readout of the memory is initiated. The optical system shown in Figure 2.7 is a multichannel time integrating correlator that forms on the 2-D CCD detector the correlation between each of the signals applied to the LED array and the signal applied to the AOD.[69–71] The charge stored on the CCD is transferred vertically by one pixel during the blanking interval of the TV camera. As a result, the correlation between the current video line and a particular line of the reference image is added to the correlation between the previous video line and the adjacent reference line. This procedure results in the formation of the 2D correlation between input and reference images on the CCD detector. More precisely, we can say that the ith LED is intensity modulated by

$$A_i(t) = [1 + a_i(t)]\text{rect}(\frac{t - T_2 - T_1/2}{T_1}), \tag{2.16}$$

where $a_i(t)$ represents a bipolar signal associated with the ith line of the reference

image $1 \leq i \leq M$, M is the number of lines in the reference image, and T_1 is the duration of the reference image. The delay time T_2, is equal to the video horizontal line time plus blanking interval and is necessary so as to ensure that the entire video line has filled the Bragg cell before the reference image readout is triggered. The jth line of the input image results in an intensity modulation of the diffracted light after the AOD (assuming uniform incident intensity) that is given by

$$B_j(t - x/v) = [1 + b_j(t - x/v)]^2 \text{rect}(\frac{t - x/v - T_2/2}{T_2}), \tag{2.17}$$

where x is the horizontal coordinate measured from the AOD transducer, v is the acoustic velocity of the signal in the AOD, $b_j(t)$ is the bipolar signal associated with the jth line of the input image $1 \leq j \leq N$, and N is the total number of lines in the input image. Using this notation, the *new* charge accumulated on the ith line of the CCD during the jth horizontal video line is given by

$$\Delta C_i(x) = \int_0^{T_1+T_2} A_i(t) B_j(t + x/v) dt, \tag{2.18}$$

where the change of sign in the argument of B_j is due to the imaging optics between the AOD and CCD planes. So, the total charge that accumulates in the Mth row of the CCD after the nth horizontal video line is equal to the charge accumulated in the $M - 1$ previous CCD rows plus the charge generated in the

Mth row during the nth input line. That is

$$\begin{aligned} C_n(x) &= \sum_{i=0}^{M-1} \int_0^{T_1+T_2} A_{M-i}(t) B_{n-i}(t + x/v) dt \\ &= \sum_{n'=n-M+1}^{n'=n} \int_0^{T_1+T_2} A_{n'+M-n}(t) B_{n'}(t + x/v) dt, \\ &= \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} [1 + a_{n'+M-n}(t)][1 + b_{n'}(t + x/v)]^2 dt, \\ &= T_n(x) + 2 \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} a_{n'+M-n}(t) b_{n'}(t + x/v) dt, \end{aligned} \tag{2.19}$$

where $T_n(x)$ is the bias signal on the nth output line and for simplicity, we have assumed that the Bragg cell aperture T is equal to the sum of input and reference image durations (i.e. $T = T_1 + T_2$).

For zero mean signals $a_i(x)$ and $b_i(x)$ we have that $< a_i(x) >=< b_i(x) >= 0$, where $<>$ indicates the expected value over x and i and we can write the correlation plane bias signal as

$$T_n^0(x) =< T_n(x) >= 2M \int_{T_2}^{T_1+T_2} \text{rect}(\frac{t - x/v - T_2 - T_1/2}{T_2}) dt. \tag{2.20}$$

In this special case, we can calculate the worst case signal to bias ratio (SBR) as

$$SBR = \frac{1}{T_1 M} \sum_{n'=1}^{n'=M} \int_{T_2}^{T_1+T_2} a_{n'}(t) b_{n'}(t) dt, \tag{2.21}$$

where we have assumed that the correlation signal reaches a maximum at $x = 0$, $n = M$. So we can see that a great deal (more than $1/2$ in general) of the output

signal dynamic range is devoted to representing the bias signal $T_n^0(x)$. In fact, the SBR reaches a maximum value of SBR=1 only when the input and reference images are exactly matched, $a_i(x) = b_i(x)$. This effect can be costly when trying to perform image recognition in the presence of noise, so we shall demonstrate a method of obtaining the true correlation $a_i(x) * b_i(x)$ by removing the bias term from Equation 2.19. It is important to note that under the zero mean condition the bias signal $T_n^0(x)$ is independent of the input and reference signals. In the next section we will examine the more general case of signal dependent bias.

2.2.2 Bias Considerations

In order to represent a bipolar signal with light intensity, a constant bias is added to the signal before modulation. The presence of this bias term in $A_i(t)$ and $B_j(t)$ results in the bias term $T_n(x)$ in the system output. Removal of $T_n(x)$ is therefore required before the true correlation may be observed. Shown in Figure 2.8 is a typical line of the signal independent bias $T_n^0(x)$. This signal was obtained by providing a black background to the TV camera and setting the reference image equal to a constant level of 1. The bias signal is approximately triangular, characteristic of the correlation between the two rectangular functions associated with $A(t)$ and $B(t)$.

Removal of the bias function $T_n^0(x)$ is achieved by first generating constant reference and input images as described above. A recording of the 2-D output correlation of these two images is exactly $T_n^0(x)$. Thereafter, in order to retrieve the true correlation from the system output, $T_n^0(x)$ is subtracted from $C_n(x)$ on

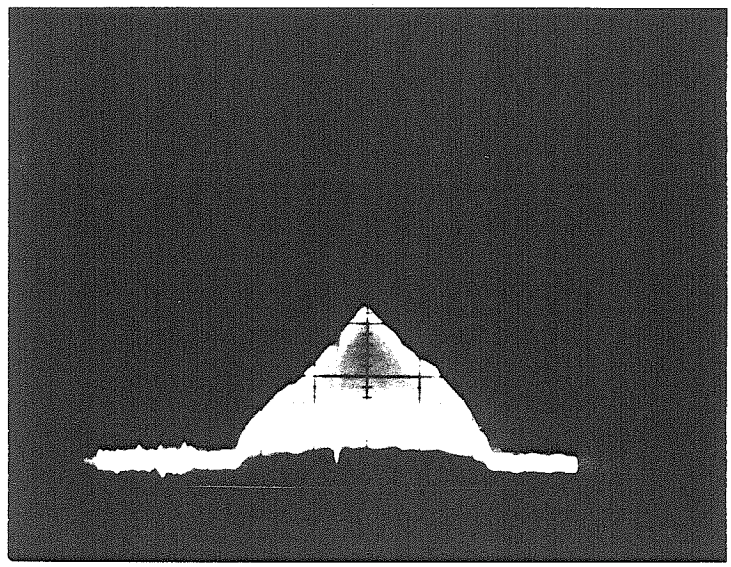


Figure 2.8 : Triangular bias signal measured from optical system

a pixel-by-pixel basis. This yields the desired signal :

$$a_n(x) * b_n(x) = C_n(x) - T_n^0(x). \tag{2.22}$$

If the zero mean condition is not satisfied, then the bias term $T_n(x)$ becomes signal dependent. This is generally the case for two reasons: first, although it is possible in principle to electrically preprocess an arbitrary input image and generate a zero mean signal, this procedure is difficult to realize in practice. Second, the finite extent of the input and reference images as required by the system, results in effective zero padding outside the image region. This zero padding changes the effective image statistics near the edges of the correlation plane. It is easy to show that the bias signal for arbitrary input and reference

images is given by :

$$T_n(x) = 2M \int_{T_2}^{T_1+T_2} \text{rect}(\frac{t+x/v}{T_2})dt$$
$$+ 2 \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} [a_{n'+M-n}(t) + b_{n'}(t+x/v)]dt, \quad (2.23)$$
$$= T_n^0(x) + 2 \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} [a_{n'+M-n}(t) + b_{n'}(t+x/v)]dt.$$

We see that the first term is the typical triangular bias; however, the second term is in fact signal dependent. Various schemes for removing the second term from Equation 2.23 may be envisioned. One particularly straightforward method would be to simply record the two signals

$$signal1_n = \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} a_{n'+M-n}(t)dt \quad (2.24)$$

and

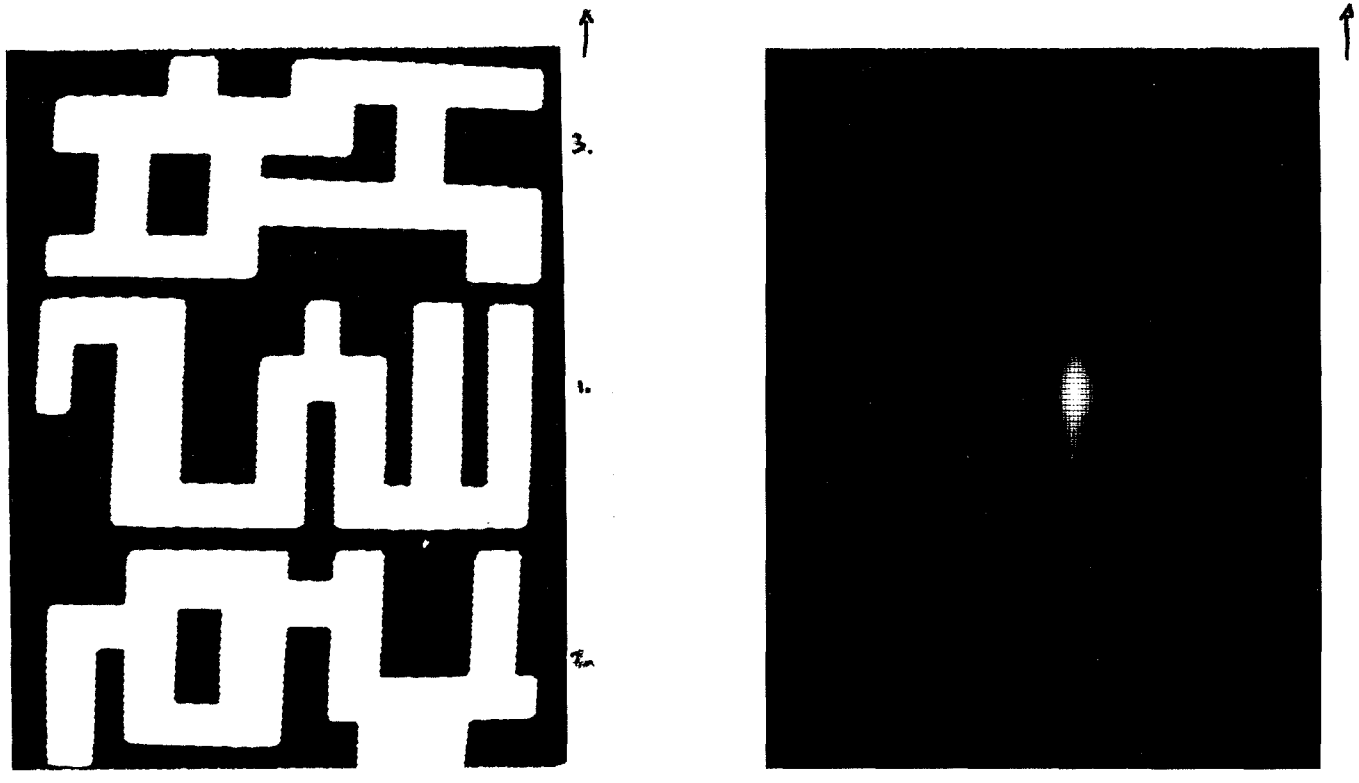
$$signal2_n(x) = \sum_{n'=n-M+1}^{n'=n} \int_{T_2}^{T_1+T_2} b_{n'}(t+x/v)dt, \quad (2.25)$$

and perform a pixel by pixel subtraction off line as we did with the signal independent term.

2.2.3 Bipolar Correlation Results

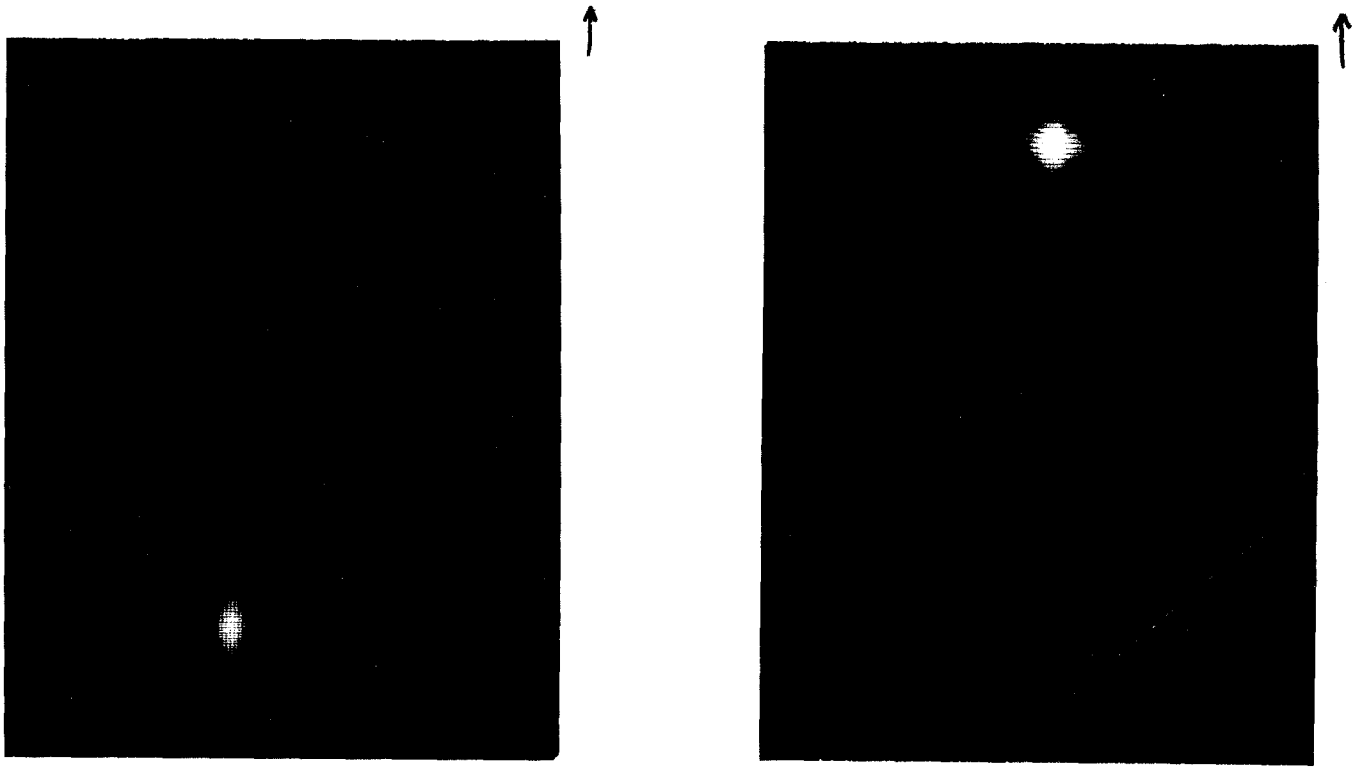
We have implemented the above system and have obtained correlation results as shown in Figure 2.9. The AOD used in our system was a TeO_2 (Crystal Technology # 4050s) device with a center frequency of 50MHz, a 35MHz bandwidth and a 70μs aperture. This was more than adequate to accommodate one standard video line of duration 63μs with a bandwidth of 5MHz. In our experiments, the three images shown in Figure 2.9a were presented to the TV camera as the input scene. These images were chosen heuristically and satisfy the zero mean condition. The TV camera signal was highpass filtered in order to remove any constant bias due to background illumination or camera electronics. After also removing the horizontal and vertical sync pulses the video signal was amplified and a constant level was added. This level was adjusted so as to ensure that the input signal was strictly positive. This signal was used to amplitude modulate a 50MHz RF carrier, which was then applied to the AOD.

The LED array used in these experiments was fabricated by Honeywell and comprised 180 elements, each capable of transmitting 5μW of optical power. Various reference images were generated using an IBM PC. Each reference image was 16 lines high by 32 pixels deep, however by virtue of the bipolar encoding scheme employed, a data array of 32 bits by 32 bits was necessary to represent one image. At the beginning of each input image horizontal line time, the 32 lines of the reference image were read out to the LEDs. Preceding each LED was a 2 bit resistive ladder network used to convert two bits of digital data into one of the three appropriate analog levels. The output signal from the correlator was captured using an IBM PC equipped with a frame grabber board



(a) (b)

Figure 2.9 : Bipolar correlation results after bias subtraction.
(a) Input scene [top:IMG3, middle:IMG1, bottom:IMG2].
(b) Correlation output for reference image = IMG1.



(c) (d)

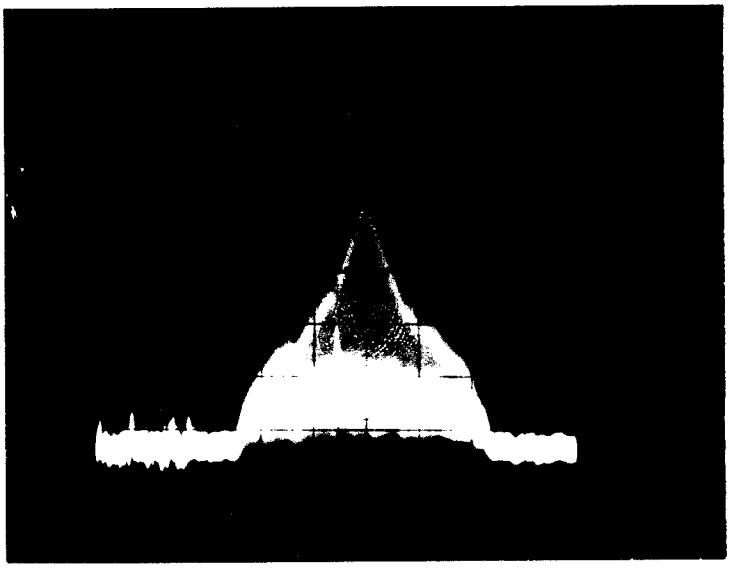
Figure 2.9 (cont.)

(c) Correlation output for reference image = IMG2.

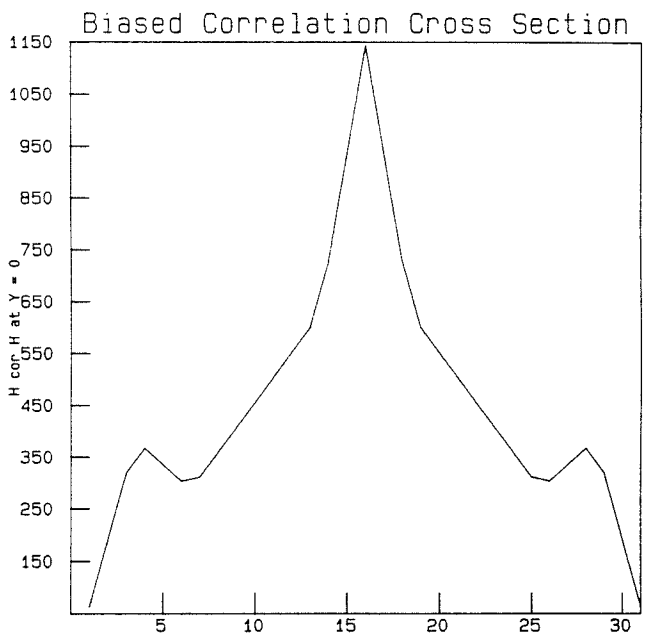
(d) Correlation output for reference image = IMG3.

and software. Using this postprocessing equipment the signal independent bias was first captured and stored as described earlier. Then, for each 2D correlation performed, an image subtraction operation was necessary to obtain the desired bipolar correlation signal. The bipolar correlations obtained using this procedure are shown in Figure 2.9(b-d) for various reference images. As can be seen from the figures, the system exhibits good cross correlation suppression and strong autocorrelation peaks.

In Figure 2.10 we show the $Y = 0$ cross section of an autocorrelation of the letter **H** as obtained before bias subtraction using the bipolar system as well as by computer simulation. The simulations show, as we would expect, superior peak signal to sidelobe level ratio in the case of bipolar scenes as compared with the binary (unipolar) case. This is an advantage not only because of the SBR issue discussed earlier but also because recognition performance is often a function of the peak to sidelobe ratio. The peak to sidelobe ratios obtained in the three cases shown in Figure 2.10(b-d) are 3.3, 2.75, and 6.1 for the biased bipolar, unipolar and bipolar images respectively. We also see that this performance criterion places the biased bipolar system on an approximately equal level with the unipolar case. This is an expected result since we can consider the biased bipolar system simply as an extension of the binary system to accommodate a third, higher pixel value. The oscilloscope trace is seen to be of approximately the same form as these simulation plots and exhibits a peak to sidelobe ratio of 2.8.



(a)

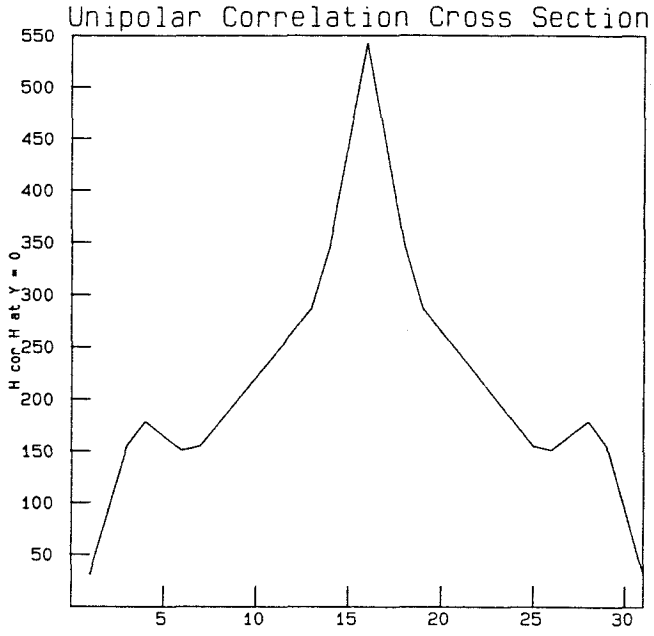


(b)

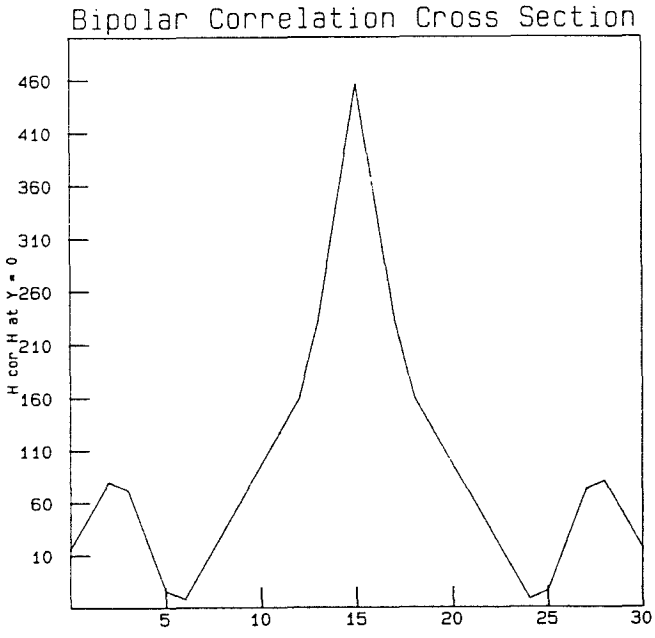
Figure 2.10 : Y=0 slice of the 2D autocorrelation of the character **H**.

(a) Signal obtained from optical system before bias subtraction.

(b) Computer simulation of optical system output before bias subtraction.



(c)



(d)

Figure 2.10 (cont.)

(c) Computer simulation of unipolar autocorrelation slice.

(d) Computer simulation of bipolar autocorrelation slice.

2.2.4 Implementation of a Linear Discriminant Function

Taking advantage of the flexibility afforded by the computer generated reference images, it is possible to form arbitrary linear combinations of images to generate more powerful filters. These filters may in turn be used to perform more complex pattern recognition tasks as described in the previous section. In this work, a linear discriminant function (LDF) based on three images was generated using the perceptron learning algorithm. The perceptron algorithm may be written as

$$w^{k+1} = w^k + b_k f_i, \tag{2.26}$$

where f_i are the images to be classified as elements of either class Ω_1 or class Ω_2, w^k is the classification filter or weight vector after the kth iteration of the algorithm and

$$b_k = \begin{cases} 1 & \text{if } f_i \text{ is in } \Omega_1 \text{ and } w^k \cdot f_i < w_0 \\ -1 & \text{if } f_i \text{ is in } \Omega_2 \text{ and } w^k \cdot f_i > w_0 \\ 0 & \text{if } f_i \text{ is classified correctly,} \end{cases} \tag{2.27}$$

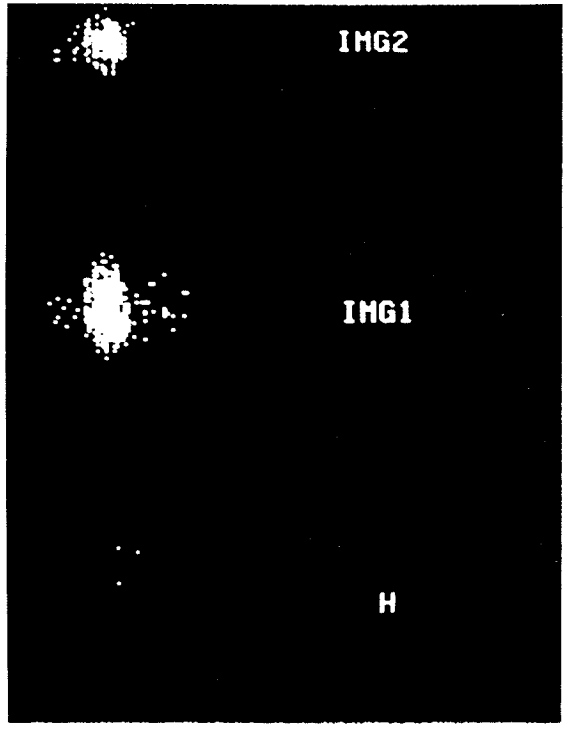
where w_0 is the correlation peak threshold level. A classification is performed by evaluation of the 2D correlation between an unknown input image and the stored weight vector. If the peak of the correlation pattern is above w_0 then the input is assigned to class Ω_1 otherwise it is assigned to class Ω_2. In our experiments the perceptron algorithm was initialized with the filter

$$w^0 = \sum_{f_i \text{ in } \Omega_1} f_i - \sum_{f_i \text{ in } \Omega_2} f_i. \tag{2.28}$$

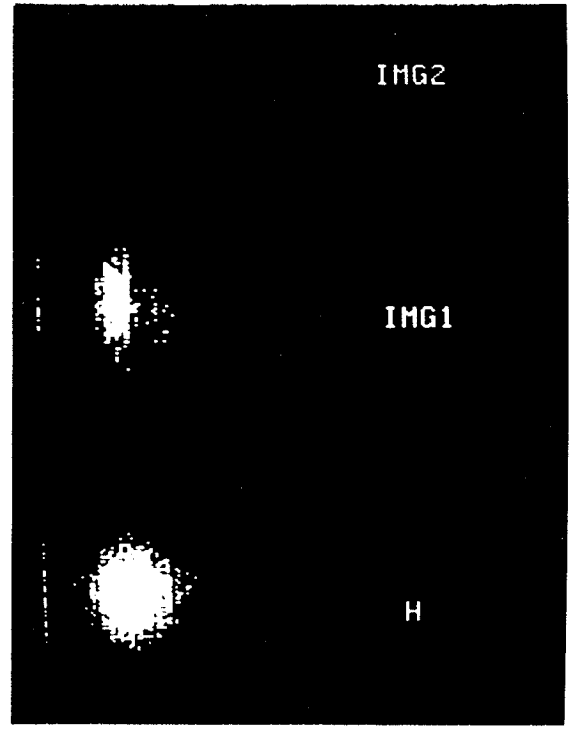
Since only three levels may be represented in our system, a threshold arithmetic

must be implemented during training. That is, when executing the above algorithm we have that $1 + 1 = 1$ and that $-1 - 1 = -1$. We say that the algorithm has converged when all images are classified correctly. Although there is no guarantee of convergence in the bipolar case, we observed convergence for all possible dichotomies. The LDF as found through the above procedure, was used as the reference image while the three seed images were placed in the input plane of the optical system. The results for three different class assignments are shown in Figure 2.11. Specifically, in Figure 2.11a we have formed a filter designed to classify $IMG1$ and $IMG2$ as elements of class Ω_1 while classifying H as Ω_2. The two other dichotomies are shown in Figure 2.11b and 2.11c. We see that successful classification was achieved for all three of the nontrivial dichotomies possible.

This system is a simple example of the correlation-based approach to optical PR in which a composite filter is used as a reference image. As we know from the previous discussion, this approach is suboptimal for an optical implementation due to DR limitations. In the remainder of this thesis we will describe the optical disk as it may be used in optical PR systems and we will demonstrate architectures based on the reference library approach that take advantage of the characteristics of optical memory disk technology.

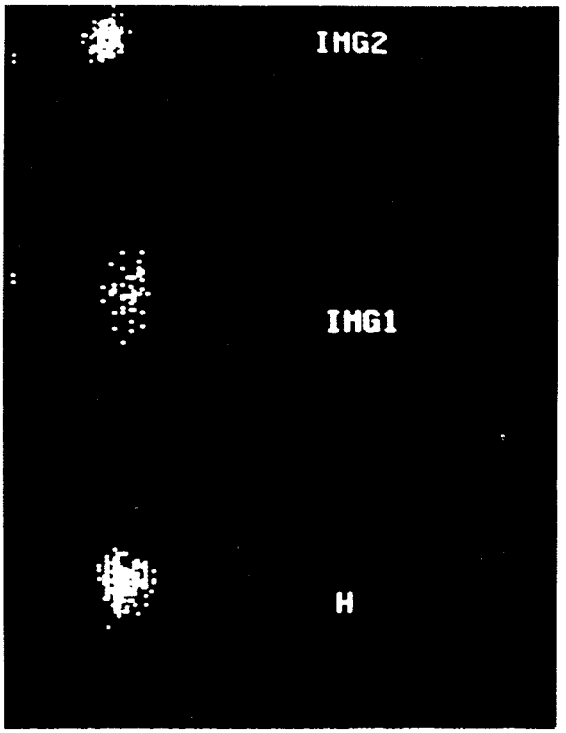


(a)



(b)

Figure 2.11 : Optical discriminant function results.
(a) Correlator output for filter classifying $\Omega_1 = IMG1 + IMG2$ and $\Omega_2 = H$.
(b) Correlator output for filter classifying $\Omega_1 = IMG1 + H$ and $\Omega_2 = IMG2$.



(c)

Figure 2.11 (cont.)

(c) Correlator output for filter classifying $\Omega_1 = IMG2 + H$ and $\Omega_2 = IMG1$.

3. Optical Disk Characterization and Parallel Readout

Ever since the introduction of the videodisc system in the late 1970s and the compact audio disc player in the mid 1980s, optical disk technology has been maturing at a rapid pace. Both write once read many (WORM) and magneto optic read/write disk drives are presently available for high density storage on mainframes and personal computers.[72–76] The conventional mode of both reading and writing used in present optical disk systems is serial. Specifically, a laser source will write one bit of data at a time on the disk, typically through a thermal mechanism. Readout is achieved by using a lower power beam to illuminate the location of each bit on the disk individually, and based on the reflected or transmitted intensity detected, the bit is decoded as a logical 1 or 0. Although serial readout is well suited to conventional computers, the optical disk itself is naturally a parallel readout device.[32] To see this, consider illuminating a large portion of the disk with a collimated beam. The reflected or transmitted light contains all of the data originally recorded in the illuminated area and a simple imaging system makes this data available to a detector array. This parallel access capability can be attractive when trying to solve memory access and contention problems in parallel computing architectures or when trying to implement an intelligent memory search procedure as with database machines.[4,5,33,77,78] Further, the optical disk represents a high resolution, computer controllable, spatial light modulator (SLM) that may be used in various optical computing architectures. For example, images stored on an optical disk may serve as a library of references in an optical image correlator as we proposed in the previous chapter, or holograms stored on the disk may serve as interconnect patterns for hybrid

optical/VLSI-based neural networks. All of the applications we discuss throughout the remainder of this thesis are designed to combine the parallelism and interconnectivity of optics with a mature optical disk technology to result in feasible optical systems that perform useful computational tasks.

We begin this chapter with a brief overview of optical storage media. Preliminary data on some ablative media from Optotech will be included here for comparison with the Sony system to be discussed later. We will then go on to characterize the disk system used in our work, a Sony prototype sampled format drive with both WORM and magneto-optic media. In the last section we discuss parallel optical readout of 2-D blocks of data such as images. Issues of contrast and diffraction efficiency will be analyzed in that section.

3.1 Overview of Optical Storage Mechanisms

Recently a great deal of energy has been invested in the science of optical data storage. This fact is evidenced by the growing number of papers presented annually at conferences devoted to this field. This effort has not gone unrewarded. Optical storage systems for applications such as providing access to mass produced information databases, archiving large amounts of data or realtime high density storage and retrieval for PC systems are all commercially available. Although we will be dealing here with the disk format, optical tape is also available for archival purposes and many of the mechanisms we will discuss are equally well suited to a tape format. The principal motivation for the development of optical data storage systems is their potentially large capacity with relative mechanical simplicity of the reading/writing mechanism. While magnetic disks can

achieve storage densities comparable to their optical counterparts, the mechanical complexity associated with such systems precludes the use of a removable format. The removability of the high capacity optical disk is one of the principal attractions of this new technology. Wavelength limited optical recording can approach a density of 10^8 bits/cm^2 as compared with conventional removable magnetic recording techniques, which achieve densities of 10^5 bits/cm^2.[79–82] A second motivation for considering optical storage is potentially higher speed. Although this advantage has not yet been realized for random access owing to the relatively large masses associated with conventional serial optical readout heads, burst data rates from optical memory can approach 5-10 Mbytes/sec. Because optical recording mechanisms are most often thermal, *recording time* is limited by two factors: minimum heating pulse width and cooling time. The practical limitation is cooling time. For a typical metal, cooling times are on the order of 0.1μs yielding a maximum recording rate of 10 Mpixels/sec making the recording speeds for optical storage systems roughly equal to those of their magnetic counterparts.

A wide variety of materials and recording mechanisms have been proposed for use in optical disks. An excellent review of these may be found in Reference[79]. Here, we will briefly overview several of the more common optical storage schemes giving only an abbreviated account of each mechanism in order to provide some insight into the breadth of the optical storage field.

Optical storage systems may be classified as one of three types; factory replicated read-only memory or CD-ROM, permanent user recordable WORM, and erasable or read-write-erase (RWE) type. Into the first category falls the most

successful commercial application of optical storage technology, the audio compact disc or CD. Also in this group are videodisks. These mass produced disks comprise a glass or more often plastic base layer or substrate on which is deposited a thin layer of metal, most often aluminum. The data to be recorded on the disk is then embossed onto the Al layer through a stamping process using a metal master disk. Finally, a protective transparent layer is deposited. The recorded data may then be read out optically as variations in the reflectivity of the Al layer.

There are two basic mechanisms which hold promise for use in RWE disk systems: magneto-optic (MO) and chalcogenide. In the first mechanism MO readout is accomplished through the detection of the polarization of light that has been reflected from or transmitted through the MO medium.[82–86] The MO medium causes the polarization of an incident electric field to rotate on reflection via the Kerr effect or on transmission via the Faraday effect. The direction of the induced rotation is determined by the magnetization of the MO medium. Typically, an output polarizer is used to convert the induced rotation into an intensity signal, which is interpreted as the stored data. In this way the magnetic state of the MO disk is detected optically. In order to characterize a particular magnetic material with respect to optical storage there is the MO figure of merit, which relates the induced polarization rotation and the optical absorption. For some of the most attractive MO media, this figure approaches $2F/\alpha = 1°$, where F measures the induced rotation in degrees per cm and α is the optical absorption coefficient in nepers/cm. Since this figure of merit is quite small even in the best of material systems, readout SNR in MO storage systems is rather poor and was

until recently the primary deterrent to the successful commercialization of these systems.

There are two basic MO recording mechanisms. In Curie point writing, an incident laser is focused into the MO medium, heating it above its Curie temperature. If we assume that the region surrounding the heated spot is all magnetized in the same direction (say 'up'), then the closure flux in the heated region is in the opposite direction and results in a force being exerted on the heated spot. Magnetic domains within the heated spot therefore tend to align in the 'down' direction. Often, an auxiliary magnetic field is applied to facilitate this alignment also. On cooling, the region that was illuminated retains the reverse magnetization, thereby recording a single bit. The other MO recording mechanisms take advantage of the temperature dependence of the coersive field (H_c) exhibited by some materials. The application of a heating laser pulse can be used to reduce H_c thereby making it relatively easy for an external field to switch the magnetization of the heated spot. In both modes, typical writing energies are $\approx 0.1nJ/\mu m^2$ and since submicron sized domains may exist in many magnetic materials, media resolutions can easily be made >1000 lp/mm.[84]

Another reversible mechanism that may someday lead to erasable optical storage media is photocrystallization and photoinduced amorphization in chalcogenide films.[87–89] In these so-called *ovonic* materials, heating can cause a phase transition to occur between amorphous and crystalline states with the heating and cooling rates determining the direction of the transition. An incident laser pulse then can be used to induce such a phase change with the pulse energy and duration determining the direction. Since different material states exhibit

different optical properties, index (phase) modulation or reflectivity (intensity) modulation may be used as the readout mechanism. Once again we have that typical writing energies for the least energetically favorable direction in these materials are roughly $0.1nJ/\mu m^2$ and resolutions can be >1000 lp/mm. Ovonic material systems have been demonstrated to yield an 8:1 change in reflectivity between amorphous and crystalline states, which is more than sufficient for optical storage applications; however, limited cycling between phases due to material fatigue remains a serious limitation.

In the WORM category of optical storage media there are many potential mechanisms. Two approaches that show promise for future optical storage systems are photopolymerization and photochemical hole burning.[90–92] The most common WORM mechanism and the approach taken in nearly all commercial WORM systems is the ablative type. The recording layer of an ablative optical disk is most often a thin layer of some metal (e.g., tellurium thickness $\approx$ 100Å), although recently, various organic materials have also been studied in this regard.[93,94] The metal recording layer may be either deposited directly onto the substrate or it may be stretched between plastic edge anchors resulting in the air sandwich configuration. In either case an incident laser heats a spot on the thin recording layer ablating the material and recording a hole. The debris material is either left as clumps on the substrate surface or is redistributed over the recording layer tending to bunch around the edges of the recorded pixels. On readout, the variation in reflectivity between metal film and ablated hole is detected either in reflection or transmission and can yield contrasts in the range 10:1 to 100:1. Writing energies and resolutions here are comparable with

previously described mechanisms.

During a preliminary characterization phase of this thesis we initially studied some commercially available ablative recording media from Optotech. The recording layer of the Optotech disks consists of a 100-200Å layer of Te in an air sandwich configuration with air gaps of approximately 0.3mm on either side. The required recording energy for these disks was given by the manufacturer as $0.7nJ/\mu m^2$ corresponding to a $70ns$, 8.5mW laser pulse heating a $1\mu m$ diameter circular spot. In order to avoid complications in parallel readout owing to the presence of formatting data, we purchased EA-09-004 unformatted, ungrooved, single-sided media. Unfortunately, there is no commercially available device for recording data on this media; therefore, we constructed a recording system comprising two computer controlled motor driven micrometer stages, an Ar^+ laser, and a $40\times$ objective lens for focusing. The optical system shown in Figure 3.1 was used for alignment of the optical disk in the z-direction. Notice that when the disk is in the focal plane of lens L1, a sharp focused spot will appear on the alignment CCD. Since this optical system achieved a diffraction limited spot size of $10\mu m$ diameter, a factor of 100 increase in writing power was necessary to achieve recording. This results in a required Ar^+ laser power of 1W. Also, as the media was optimized for recording and readout at laser diode wavelengths ($\approx 800nm$), an exposure time of 1-10ms was required in order to compensate the increased absorption exhibited by the substrate and coating materials at the Ar^+ green line of 514nm. Using this computer controlled recording system several patterns were recorded on the Optotech media. An example of five pixels recorded with center to center spacings of $10\mu m$ is shown in Figure 3.2.

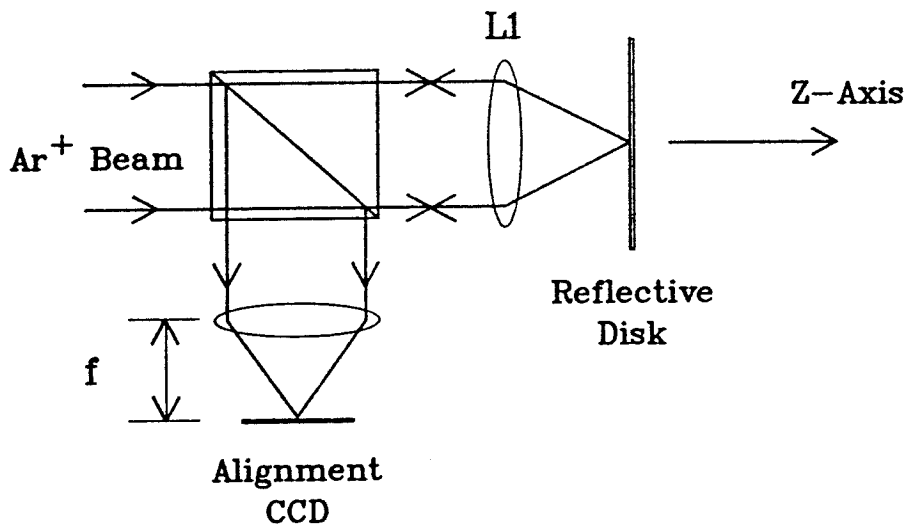


Figure 3.1 : Alignment system used for automated Optotech disk recording system.

This photograph was taken using an optical microscope in transmission mode at a magnification of $\times 2000$. It is interesting to observe the distorted edges of the recorded pixels and the variations in texture of the recording layer in the neighborhood of the holes. These effects are indicative of the redistribution of ablated material on cooling. Shown in Figure 3.3 is the first disk-based image recorded for parallel readout. The acronym CIT was recorded on the Optotech disk using $10\mu m$ spots and is photographed once again in transmission at a magnification of $\times 500$ making the total horizontal image extent $240\mu m$. As we can see from the photo, the readout image contrast is quite good, suitable for use in an optical information processing system; however, coherent optical systems will require good phase uniformity of the disk as well. Using a Jamin interferometer, a 1.25cm diameter region around the recorded CIT was measured

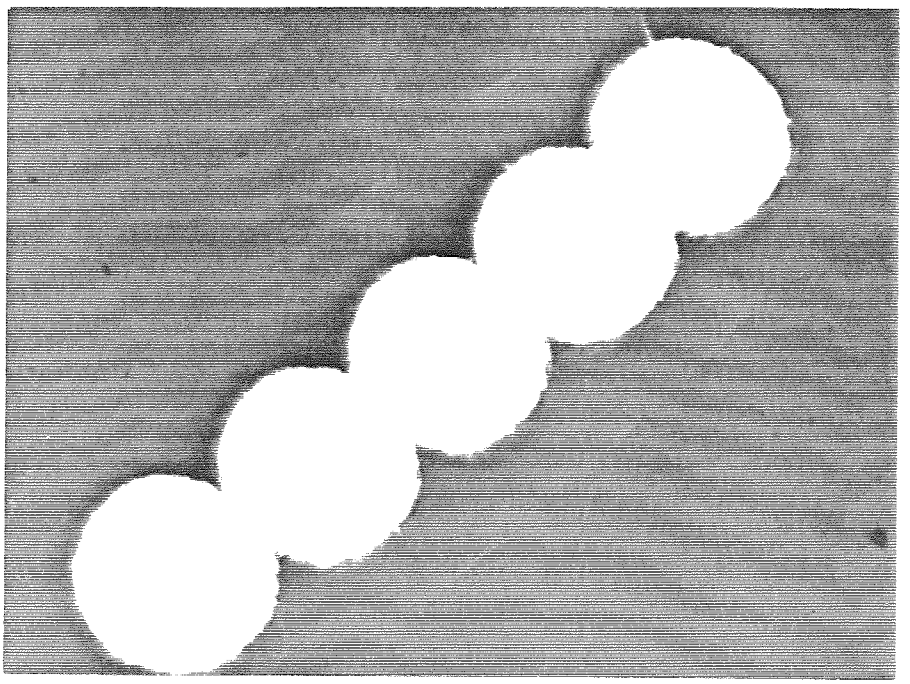


Figure 3.2 : Five data pixels recorded on Optotech disk showing distribution of debris material over recording layer.

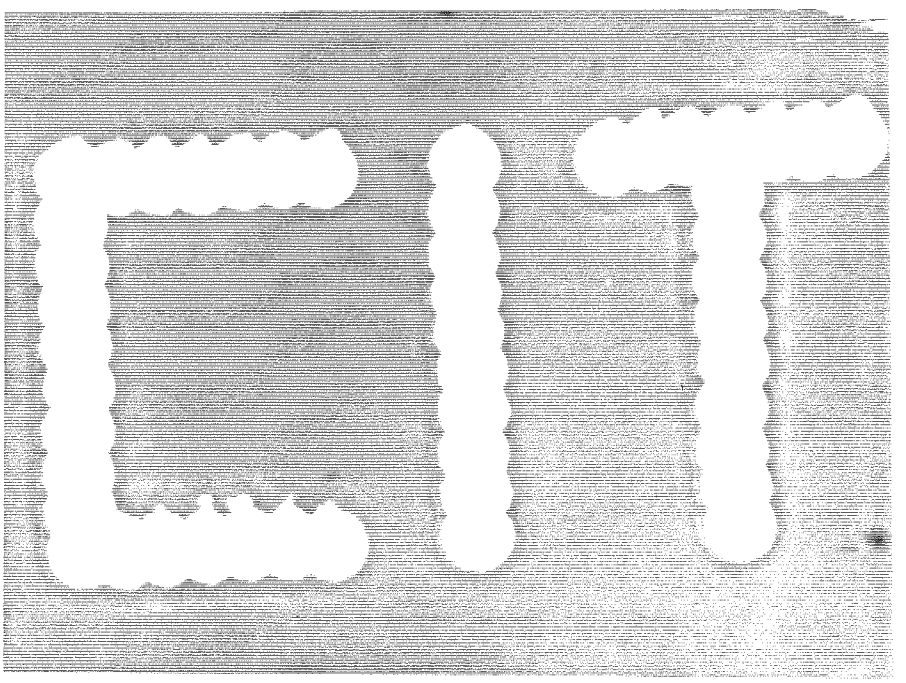


Figure 3.3 : Acronym CIT recorded on Optotech disk for parallel readout.

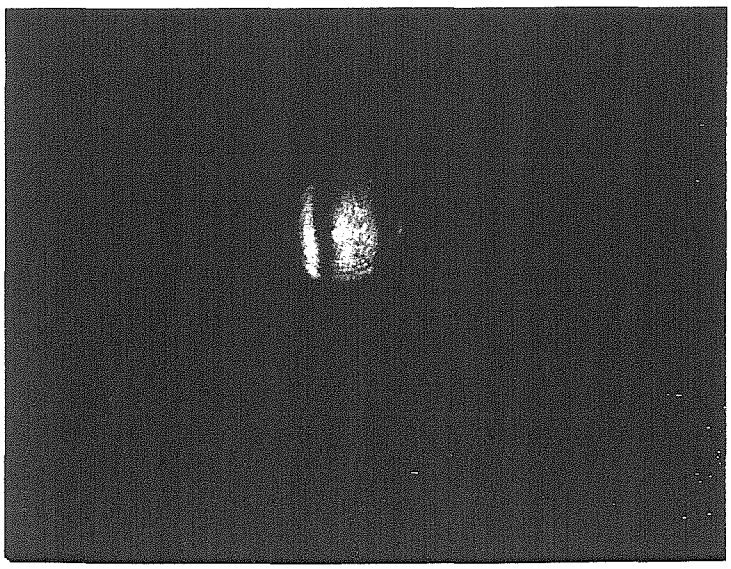


Figure 3.4 : Interferogram of Optotech disk in 1.25cm diameter circular region around recorded CIT.

for optical flatness. A photo of the interferogram thus obtained is shown in Figure 3.4.

The photo in Figure 3.4 reveals $\approx 2.5\lambda$ variation in plastic thickness in the radial direction over the illuminated 1.25cm. The illumination wavelength in this case was $\lambda = 633nm$. This result implies that square regions of 1.25mm on a side exist, over which phase variations less than $\lambda/4$ occur. One such region is large enough to record a 1250×1250 pixel image making parallel readout from these disks an attractive option for coherent optical processing systems.

Later in this chapter we will discuss parallel readout from optical disks and the implications this technique might have in confronting parallel access problems in conventional computing. Parallel access in general can refer to parallel write

as well as parallel read operations. For available optical disks however, parallel write is not a realistic operation to consider owing to power limitations. All of the mechanism discussed in this section require roughly 0.1nJ/pixel writing energy. In the attempt to write many pixels simultaneously, exposure time will be limited by thermal diffusion to $\approx 100ns$ in all cases except photopolymer-based systems. This implies that for a 1000×1000 pixel image, we require an incident laser power of 10^3 watts. This is clearly an unrealistic requirement for any commercially feasible computing system. It is for this reason that throughout the remainder if this thesis we will consider primarily parallel read *out* schemes based on optical disk technology.

3.2 Characterization of Sony System

The prototype Sony disk system used in most of our work (Figure 3.5) can read and write both write-once and magneto optic 5" reflective optical disks. The system records data as circular pixels $1\mu m$ in diameter along a spiral on the disk with 20,000 turns between an inner radius of $3cm$ and an outer radius of $6cm$. The disk is divided into 32 sectors, and each loop of the spiral from the beginning of sector 0 to the end of sector 31 is called a *track*. Because of the gradual change in the radius of the tracks with angle, we often model the tracks as concentric circles separated by $1.5\mu m$. Pixels are recorded with a constant angular separation of .001°. This corresponds to an along track pixel-to-pixel separation that varies between $.5\mu m$ and $1\mu m$ depending on radial position on the disk. This pixel recording density yields a storage capacity of over 7×10^{10} bits on each side of the disk. The system is interfaced to a personal computer

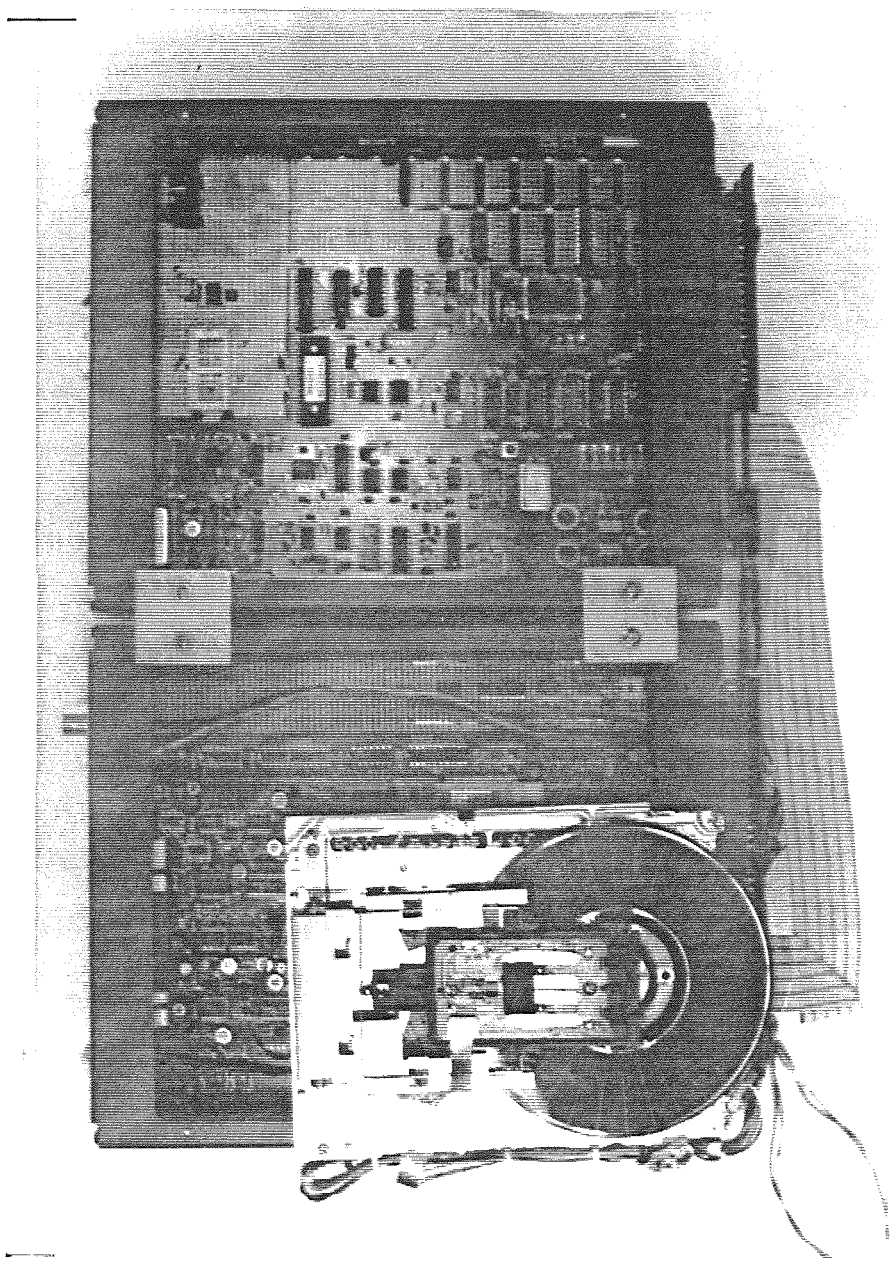


Figure 3.5 : Photograph of Sony optical disk drive used in our work.

(PC) that provides serial read/write access to the disks. The system can read or write up to 15 million bits-per-second and consequently, a 1000×1000 pixel image can be entered on the disk in 1/15th of a second. Since we can only write one line of the image per revolution, about 30 seconds are required to record the image in two-dimensional format on the disk. Notice that a thousand images using the same tracks could also be written during the same time.

Since it differs from any of the mechanisms discussed above, we will briefly describe the *alloy mode* recording mechanism employed in the Sony write-once disk that we use in our experiments.[95,96] The disk contains the four thin metal films Sb_2Se_3 (300Å), Bi_2Te_3 (150Å), Sb_2Se_3 (1400Å), and Al (1000Å) formed by sputtering deposition on a glass or plastic substrate. The thicknesses of the various layers are chosen so that they form a low (5%) reflectivity interference filter. During the recording stage, a focused laser beam heats a spot of the Bi_2Te_3 layer through absorption. The Bi_2Te_3 and Sb_2Se_3 then form a four-element alloy by diffusion, eliminating the sharp interfaces between the layers. The low reflectivity interference filter is thus destroyed, increasing the reflectivity of the medium to 12%. This reflectivity difference is detected during readout and decoded as a logical 1 or 0. The reflectivity of an interference filter is wavelength dependent. The thicknesses of the layers are chosen to maximize the change in reflectivity for the laser diode wavelength of $830nm$. An interesting effect occurs in the Sony media as a result of the wavelength dependence exhibited by the interference filter structure. A computer model of this structure predicts that for some wavelengths (i.e., 650nm $< \lambda <$ 750nm) a contrast reversal will occur wherein the off state reflectivity will become greater than the on state reflectivity.

This effect may have interesting implications for optical information processing systems based on the Sony media. Without using additional storage area on the disk, both a binary image and its complement may be retrieved from the disk simultaneously using a two wavelength readout scheme. Using this technique, the storage of 2D *bipolar* data on the *binary* Sony storage media becomes possible.

The Sony system can also read and write on magneto-optic media. Since the mechanism of recording and readout for the Sony MO media is essentially the same as described in the previous section, only a few details specific to the Sony media are summarized here. The Sony magneto optic disk contains a rare-earth transition-metal alloy of TbFeCo. During the recording stage, the laser heats a spot on the disk above $180°C$, the Curie temperature of the material. As the spot cools below this temperature, the material within the spot retains the magnetization of an external field applied perpendicular to the disk surface. The polarization of a low power readout laser rotates upon reflection from the spot by an angle of $\pm.15°$, due to the magneto optic Kerr effect. The reflectivity of the magneto optic disk is 17%, and the sign of the rotation angle depends on the direction of magnetization in the spot. This rotation is detected through a crossed polarizer and decoded as a logical 1 or 0. Depending on the setting of the polarizers, the amplitude of the light corresponding to the two states can be either on/off or plus/minus.

Throughout the remainder of this thesis we will describe architectures and demonstrate systems which utilize the Sony WORM media operating in a parallel readout mode. One potential disadvantage to this approach is that realtime modification of the recorded data is not possible thereby limiting the possibilities

of realizing adaptive optical systems. Such systems would be possible using MO disks; however, the extremely small rotation angle induced by the MO media results in poor light efficiency in systems that use these devices. Although careful engineering of such systems may yield realistic adaptive parallel access optical systems based on MO technology, we will not discuss such systems here.

Current disk systems use either continuous or sampled format schemes to maintain the position of the head over data in a track. Continuous format systems use a return signal either from a guide groove embossed on the disk or the recorded data itself to constantly monitor and correct the position of the head relative to the data in a track. In contrast, sampled format systems such as the one used in our experiments, use tracking and timing information embossed along radial lines on the disk to *periodically* monitor and correct the head position. These lines of tracking and timing information appear every 270 pixels. Each radial line of tracking and timing data consists of a pattern of three embossed pits repeated in all 20,000 tracks as shown in Figure 3.6. The first two pits provide tracking information. They are displaced an equal distance from the center of the track, one toward the inside of the disk and the other toward the outside. If the head is exactly over the center of the track, the readout signal strength of the two pits will be equal; otherwise, the signal returning from one of the pits will be stronger than the other, thus indicating the direction to move the head. The third pit provides timing information used to synchronize the system clock and the disk.

There are two byproducts of the sampled format scheme that facilitate the parallel readout of data. First, the across track alignment of tracking and timing information combined with the synchronization between recorder clock and disk

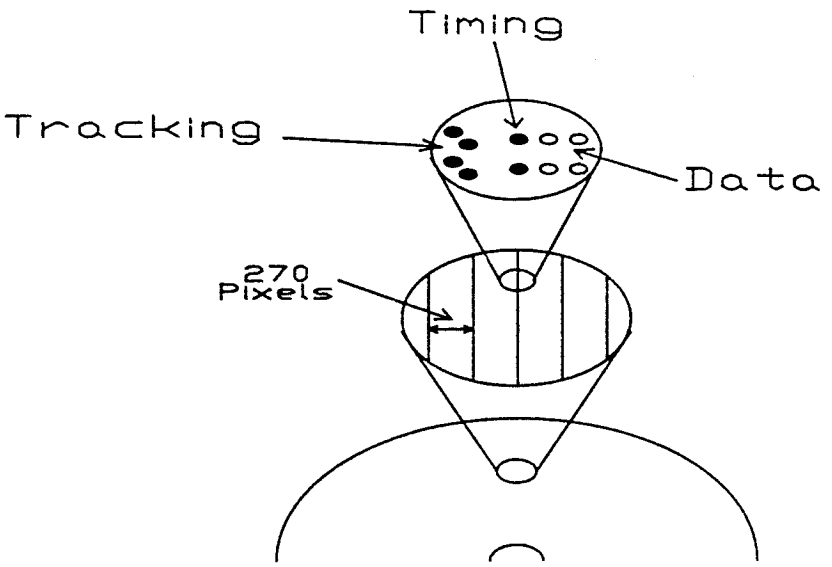


Figure 3.6 : Schematic of Sony optical disk showing servo areas which provide tracking and timing information.

rotation allows us to specify the position of individual pixels with submicron accuracy in any one of more than one billion locations. This provides us with *across track coherence*, the ability to radially align pixels across different tracks. In contrast, with continuous format systems, the position of pixels in different tracks can drift by several pixel widths within a single sector. Second, the absence of guide-grooves allows us to retrieve high contrast images from low contrast media through Schlieren imaging as described in the next section.

When we consider using the disk as a spatial light modulator, a number of additional performance issues arise. The SLM resolution is determined by the track spacing in the radial direction ($1.5\mu m$) and the minimum spot size in the azimuthal direction ($.5\mu m$ to $1\mu m$). Notice that there is an inherent sampling in

the radial direction due to the tracks. At the outer tracks, where the recorded pixels do not overlap, the image is also sampled in the azimuthal direction. We will see later that we can make use of image diffraction caused by this sampling. The reflectance function of the disk is basically binary both for the write-once and the magneto optic disks. We have observed some dynamic range in the reflectivity of the write-once disks, controllable by varying the exposure for each pixel. We have not characterized fully the grey scale capability of the system; however, we have demonstrated that area modulation can be used to code multiple reflectivity levels for superpixels consisting of several bits. For example, turning on n out of N pixels in a superpixel can be used to represent the integer value n. Various superpixel coding techniques have been investigated in the past.[109] We have implemented an area modulation scheme which uses a stochastic procedure to determine the position of on-pixel locations within each superpixel. This scheme improves the dynamic range of regions of low spatial frequency by stochastically selecting the value to be recorded in each element of an array of superpixels. This procedure along with experimental results demonstrating its success is the topic of another thesis and can be found elsewhere. The contrast of the light reflected from the disk is low for the Sony write-once disks (2:1). For magneto optic disks, the polarization of the modulated light is orthogonal to the polarization of the incident light, and the use of orthogonal polarizers in conjunction with the carrier encoding method discussed in the next section yields excellent contrast, limited primarily by the quality of the polarizers.

Many of the processing architectures we propose use coherent processing techniques requiring phase uniformity across the surface of the disk. We have

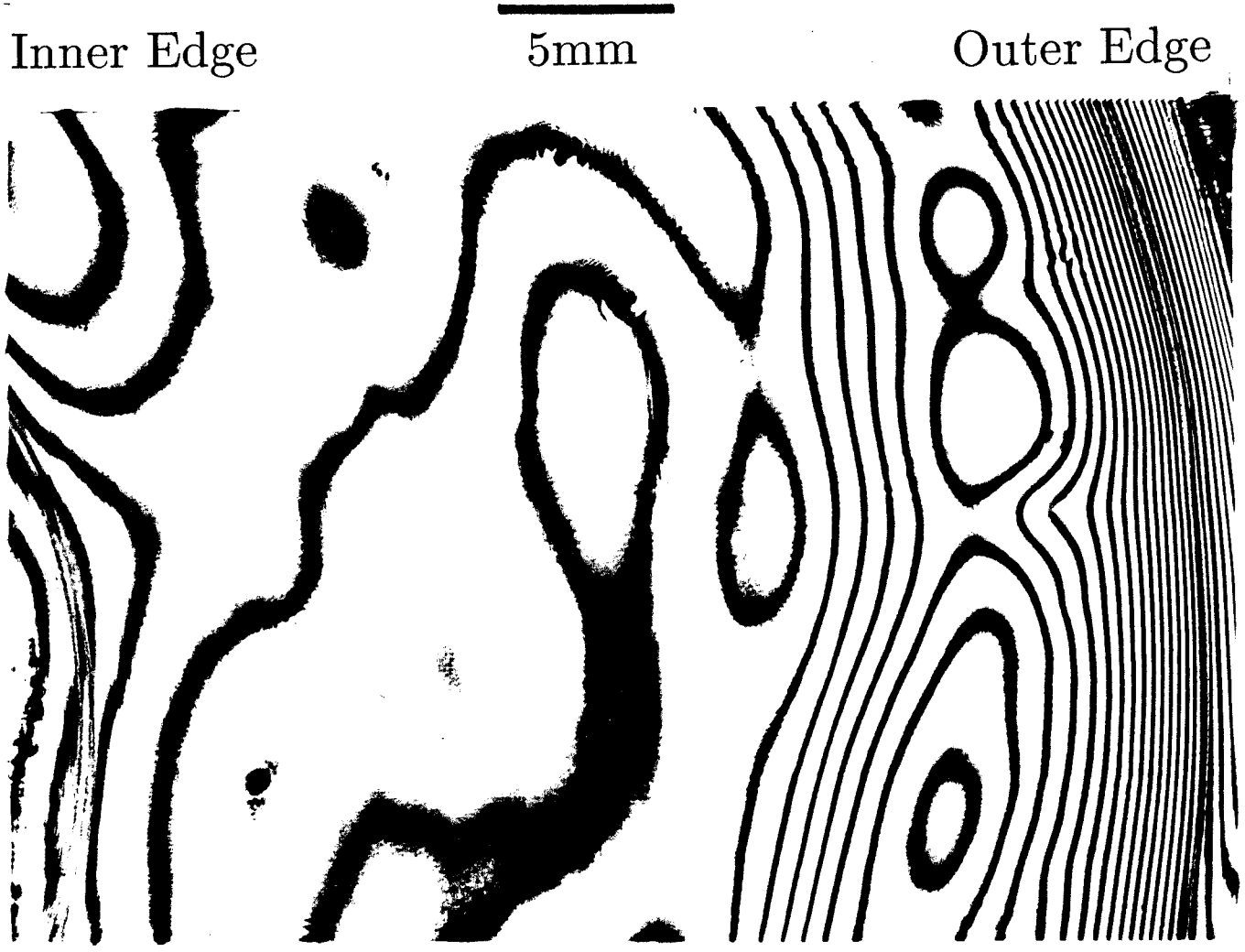


Figure 3.7 : Fizeau interferogram of Sony glass covered disk.

used a Fizeau interferometer to measure the phase uniformity of the Sony disks. Figure 3.7 shows a Fizeau interferogram of a $24\times36mm$ area of a glass-covered Sony write-once disk taken with a HeNe laser source providing illumination. The figure shows numerous regions on the disk with optical thickness variations of less than a wavelength (λ=$633nm$) over distances of several millimeters. These regions are sufficiently large to contain images or holograms thousands of pixels on a side. The greater phase variation toward the outer edge of the disk is most likely caused by index variations due to stresses induced during manufacturing. We have also tested plastic-covered disks, which generally show greater phase variation than the glass-covered ones.

In most applications, it would be most convenient if the optical disk system recorded pixels on a Cartesian grid. However, as noted earlier, our system actually writes pixels along curved tracks. We can neglect this curvature if we restrict attention to a small area of the disk. Consider a portion of the disk, centered a distance R from the disk center. As shown in Figure 3.8, we establish Cartesian coordinate axes with x tangent at the origin with the azimuthal or along track direction and y in the radial or across track direction.

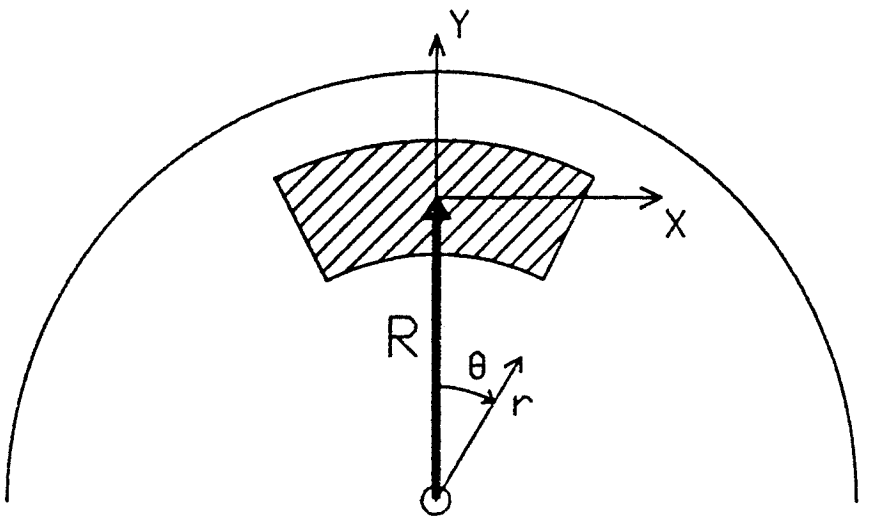


Figure 3.8 : Geometry used to calculate positional error arising from Cartesian assumption.

Equation 3.1 converts the polar coordinates of the disk to the Cartesian coordinates in the region of interest:

$$\left\{ \begin{array}{l} x = r\,sin(\theta) \\ \\ y = r\,cos(\theta) - R \end{array} \right\}. \tag{3.1}$$

The center-to-center spacing of the pixels in the radial dimension is $\delta_r = 1.5\mu m$, and the angular separation between adjacent pixels is $\delta_\theta = 0.001°$ in azimuth. We now superimpose a Cartesian grid on this pixel structure with x and y spacings as follows:

$$\left\{ \begin{array}{l} \Delta_x = R\delta_\theta \\ \\ \Delta_y = \delta_r \end{array} \right\}. \tag{3.2}$$

This choice for Δ_x and Δ_y provides the best match between the pixels recorded

on the disk and pixels near the origin on the superimposed Cartesian grid. We now calculate the deviation of the actual pixel locations from their presumed Cartesian locations. The presumed coordinates of the points on the Cartesian grid are

$$\left\{ \begin{array}{l} x' = n\Delta_x \\ y' = m\Delta_y \end{array} \right\}, \tag{3.3}$$

whereas the actual location of the pixels on the polar grid are

$$\left\{ \begin{array}{l} r = R + m\delta_r + n\dfrac{\delta_\theta \delta_r}{2\pi} \\ \theta = n\delta_\theta \end{array} \right\}. \tag{3.4}$$

The actual Cartesian coordinates of the recorded pixels therefore are

$$\left\{ \begin{array}{l} x = (R + m\delta_r + n\dfrac{\delta_\theta \delta_r}{2\pi})sin(n\delta_\theta) \approx n\Delta_x + nm\dfrac{\Delta_x \Delta_y}{R} \\ y = (R + m\delta_r + n\dfrac{\delta_\theta \delta_r}{2\pi})cos(n\delta_\theta) - R \approx m\Delta_y + n\dfrac{\Delta_x \Delta_y}{2\pi R} - n^2\dfrac{\Delta_x^2}{2R} \end{array} \right\}. \tag{3.5}$$

We calculate the deviation between the actual pixel position and the presumed location on the Cartesian grid by subtracting Equation 3.5 from Equation 3.3.

$$\left\{ \begin{array}{l} \epsilon_x = x' - x \approx nm\dfrac{\Delta_x \Delta_y}{R} = \dfrac{x'y'}{R} \\ \epsilon_y = y' - y \approx n\dfrac{\Delta_x \Delta_y}{2\pi R} - n^2\dfrac{\Delta_x^2}{2R} = x'\dfrac{\Delta_y}{2\pi R} - \dfrac{x'^2}{2R} \end{array} \right\}. \tag{3.6}$$

For an array of 1000×1000 pixels on the Sony disks, the worst case pixel placement error is 1.25% of the array size (12.5 pixels) in the x-direction (at R=3cm, Δ_x=.5μm, and Δ_y=1.5μm) and .14% of the array size (1.4 pixels) in the y-direction (at R=6cm, Δ_x=1μm, and Δ_y=1.5μm). In applications where this kind of positional error is not tolerable, we need to compensate the track curvature through optical techniques and/or the recording geometry.

Diffraction efficiency is a key parameter in determining overall system efficiency since many of the optical systems presented in the following chapters use light diffracted from the disk. Given an accurate model of the surface reflectivity of the disk and how it will be used in an optical system, we can calculate the expected efficiency of the disk in that application. As an example consider recording *all* pixel locations over some extended 2D region of the disk near the center. Since pixels on the innermost radii overlap considerably in the along track direction we may ignore the azimuthal variation of the 2D recorded pattern. Under this assumption we may consider the recorded region as a rectangular grating with a period of $1.5\mu m$ in the radial direction and a duty cycle of 0.67. Ignoring the effect of track curvature and finite aperture for the present, we can write an expression for the reflectance function of the disk recording layer as

$$r(x,y) = r_0 + (r_1 - r_0) rect(y/\delta_y) * comb(y/\Delta_y), \tag{3.7}$$

where δ_y is the pixel diameter $\delta_y = 1.0\mu m$, Δ_y is the track spacing as before, r_0, r_1 are the unwritten and written *field* reflection coefficients of the recording media respectively and $*$ here represents the 1D convolution operation. For our case we have that $r_0 = \sqrt{0.05} = 0.22$ and $r_1 = \sqrt{0.12} = 0.35$. Given the pattern $r(x,y)$ on the recording layer of our disk we may consider illuminating this disk with a normally incident, unit amplitude, plane wave whose polarization vector is perpendicular to the plane of incidence. Since the glass protective coating over the recording layer has an index of refraction different from that of air, the incident beam will undergo partial transmission at the air-glass interface. The efficiency with which the incident field will enter the glass layer is given by the

Fresnel formula $\eta_1 = (2/n+1)^2 = 0.64$ for n=1.5. This then is the intensity of the readout field that illuminates the recording layer. The recorded grating will diffract the incident plane wave into multiple orders. The intensity of the first order diffracted beam is obtained by computing the Fourier Transform of the reflectance $r(x,y)$ and choosing the appropriate term. The term of interest I', is given by

$$I' = \eta_1(r_1 - r_0)^2 sinc^2(\delta_y/\Delta_y), \tag{3.8}$$

and is propagating at an angle of $\theta' = sin^{-1}(\lambda/n\Delta_y)$ inside the glass layer. As usual θ' is measured with respect to the surface normal and in our case corresponds to a first order diffracted angle of $\theta = sin^{-1}(\lambda/\Delta_y) = 25°$ in the air for an illumination wavelength of $\lambda = 633nm$. The first order diffracted intensity in the air I, is obtained from the intensity in the glass I', as $I = \eta_2 I'$ where η_2 is the Fresnel transmission coefficient for propagation from glass to air at an incident angle of $\theta' = 16.3°$ and is given by the Fresnel formula

$$\eta_2 = 1 - \frac{sin^2(\theta' - \theta)}{sin^2(\theta' + \theta)}, \tag{3.9}$$

which yields $\eta_2 = 0.95$ in our case. We then find the overall first order diffraction efficiency of the track grating to be

$$\eta = \eta_1\eta_2(r_1 - r_0)^2 sinc^2(\delta_y/\Delta_y) = 0.09\%. \tag{3.10}$$

This computed value agrees well with a measured value of 0.11%. We estimate that the magneto optic disk will be almost 1000 times less efficient than the write-once disk in most applications. This large loss in efficiency was also experimentally observed.

3.3 Parallel Readout

The fact that data can be retrieved in parallel from optical disks creates the possibility for eliminating some of the bottlenecks that currently exist in computers due to the mismatch between mass storage media and semiconductor memories. A parallel *random access* memory would be one possible way to construct a parallel readout optical memory. In this case, M out of the N bits stored on a disk could be specified and retrieved simultaneously. In such a system, the apparatus that would scan the memory to realize this parallel retrieval scheme would have to be capable of being set in $\binom{N}{M} \approx (N/M)^M$ distinct states in order to arbitrarily select any M-tuple. With N=10^{10} and M=1000, we obtain about 10^{7000} distinct states. It is clearly not practical to realize an optical scanning mechanism that can do this. Therefore, we conclude that we must somehow *structure* the stored data to reduce the complexity of the access mechanism. The most straightforward way to impose such structure is to arrange the stored data in 2-D blocks, M bits each, that are retrievable in parallel. This reduces the number of choices the access mechanism is faced with to a practical $\binom{N/M}{1} = 10^7$, for the previously quoted numbers. Notice that even in the case of structured storage, the entire disk capacity or Space Bandwidth Product, SBP=10^{10} represents too severe a requirement for simultaneous parallel readout owing to the cost of the associated optical system. This fact implies that some scanning mechanism must be combined with the structured data storage approach in order that all the disk SBP may be accessed using a simple optical system. In this section, we discuss several methods, including holographic, for recording and retrieving 2-D blocks of data from optical disks.

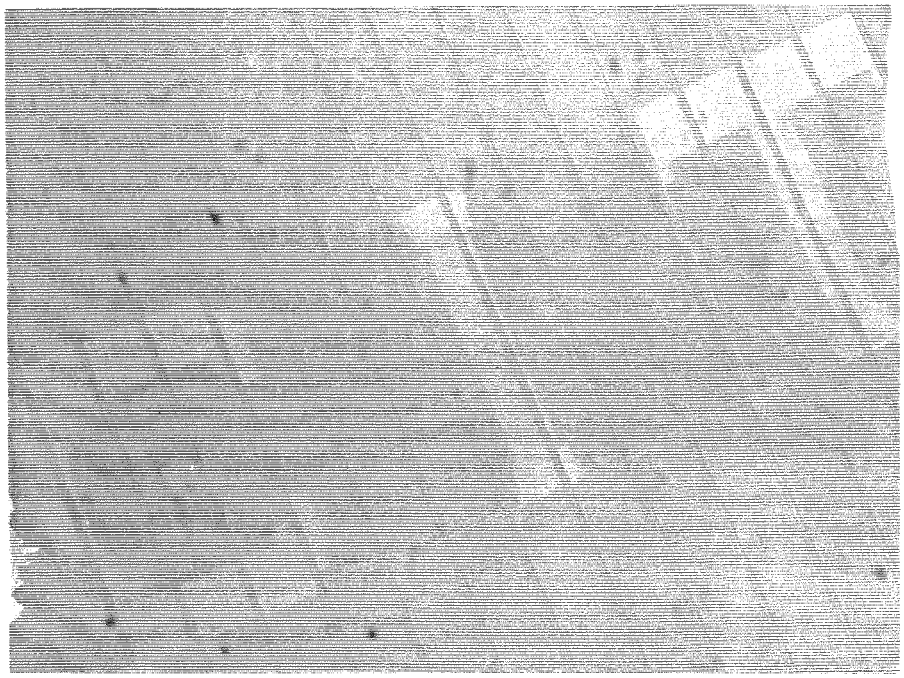


Figure 3.9 : Photograph of image (CIT) written on Sony disk as viewed through a microscope.

Figure 3.9 is a photograph of a binary image written on the Sony write-once disk as viewed through a microscope. This image consists of 4024 $\times$ 512 pixels recorded on a polar grid. Note that the track curvature is not visible. The parallel lines, evident in the figure, are the radial strips of tracking and timing information described in the previous section. There are 270 pixels between each pair of these lines. Individual pixels are not discernible in this figure, but they are perfectly aligned in the radial direction resulting in the accurate recording of the letters in the figure. Notice the poor contrast in Figure 3.9. This is due to the fact that the disk has an off-state (unwritten) reflectivity of 5% while the on-state (written) reflectivity is only 12%. This large background and low differential reflectivity results in the poor contrast seen in the photo. Tanguay *et.al.* used an interferometric technique to eliminate the background light obtained on reflection

from a stamped optical disk. We use an alternative means for improving the contrast of the retrieved image. Since the binary image $b(x,y)$ to be recorded on the disk is sampled by a polar grid, light reflected from the disk will be diffracted into many orders or sidebands whose center frequencies will be determined by the grid spacing. A Schlieren imaging system that captures only one such sideband will exhibit high contrast since the background reflectivity is not sampled by the polar grid. This can be seen by once again considering for simplicity an inner portion of the disk on which we have recorded the binary image $b(x,y)$. At the outer radii the same analysis will hold except that we would need to include the effects of diffraction into x-diffracted orders. This would reduce slightly the overall diffraction efficiency but would otherwise not alter the flavor of this analysis; therefore, as per our earlier discussion, we will treat the disk sampling as comprising a rectangular grating in the radial direction only. An expression for the disk reflectivity in this case is given by

$$r(x,y) = r_0 + b(x,y)(r_1 - r_0)rect(y/\delta_y) * comb(y/\Delta_y), \tag{3.11}$$

where $b(x,y)$ is sampled by the track grating. Consider the parallel data retrieval system shown in Figure 3.10, which we have drawn with a transmissive disk for simplicity. The DC block shown in the Fourier plane will result in an image of only the ± 1 orders appearing in the output plane. That is to say that the intensity in the output plane is given by

$$f(x,y) = 2\eta_1\eta_2(r_1 - r_0)^2 sinc^2(\delta_y/\Delta_y)b(x,y). \tag{3.12}$$

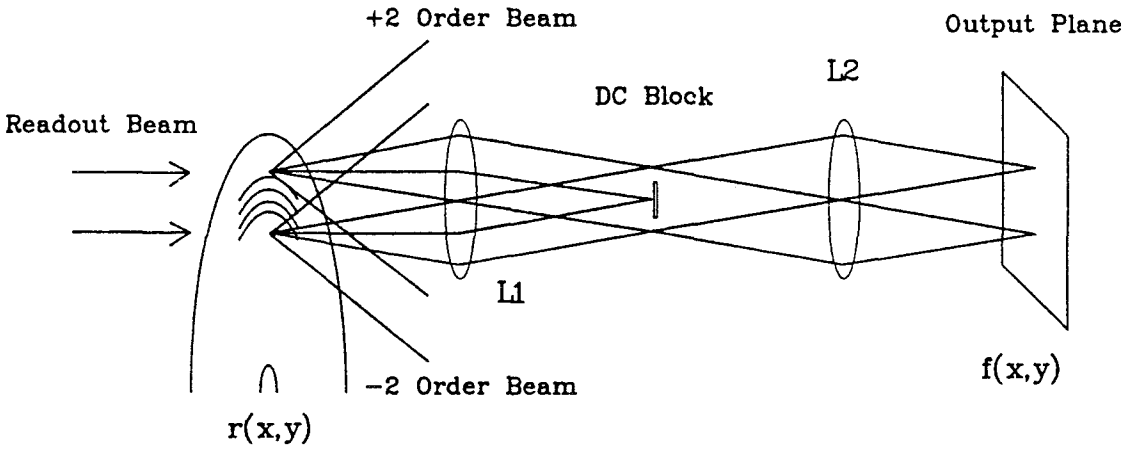
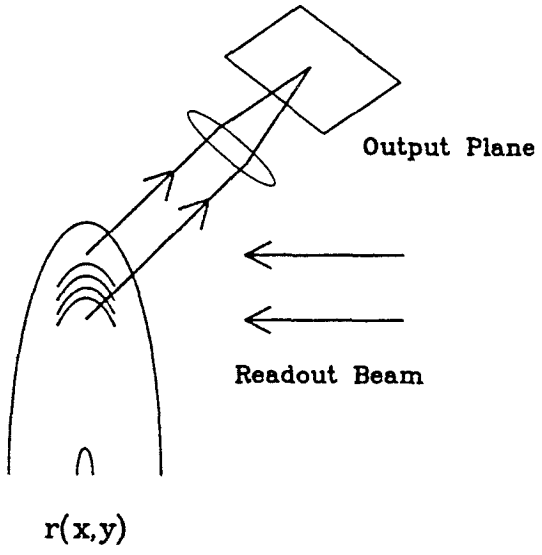
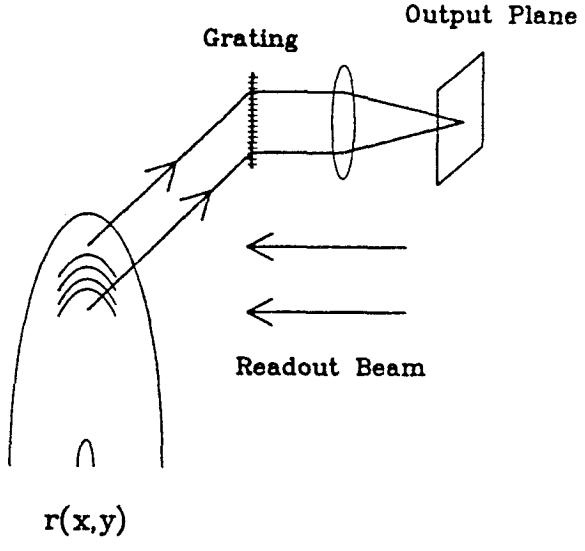


Figure 3.10 : High contrast diffractive parallel readout system.

We see that the output plane contains a replica of the sampled binary image $b(x, y)$ without the bias term due to the background reflectivity r_0. This image then is a high contrast version of the disk reflectivity function $r(x, y)$ attenuated by the diffraction efficiency we calculated earlier. The optical retrieval system shown in Figure 3.10 requires a numerical aperture of $N.A. = sin\theta_{max} = sin(\theta_D + \theta_b)$ where θ_D is the angle of the first order diffracted carrier and θ_b is the angular bandwidth of the binary image $b(x, y)$. We have $\theta_b < \theta_D/2$ by the sampling theorem so that $N.A. = 0.61$. This is a fairly stringent requirement for a simple imaging system. Since the N.A. calculated above is rather large and since we will want to work in reflection for the Sony media, we use the readout system shown in Figure 3.11a. The system shown in Figure 3.11a captures only one diffracted order resulting in a factor of 2 loss in light efficiency; however, the numerical aperture of this system has been reduced to $N.A. = 0.22$. Strictly



(a)



(b)

Figure 3.11 : Low N.A. diffractive readout schemes.

(a) Single sideband diffractive readout system.

(b) Optical system of (a) using a grating to correct for trapezoidal distortion.

speaking, the output image obtained from the system of Figure 3.11a will exhibit trapezoidal distortion owing to the fact that the object plane is not perpendicular to the optical axis. This distortion may be compensated through the use of an auxiliary prism or a grating as shown in Figure 3.11b. For parallel readout from optical disks we found that since the recorded images are quite small ($\approx 1mm$ on a side), the trapezoidal distortion resulting from the use of Figure 3.11a was undetectable; for simplicity, we will choose to use this system throughout the remainder of this thesis. An example of a high contrast image obtained by imaging the first diffracted order is shown in Figure 3.12. The light diffracted by the tracks was selected to form this image.

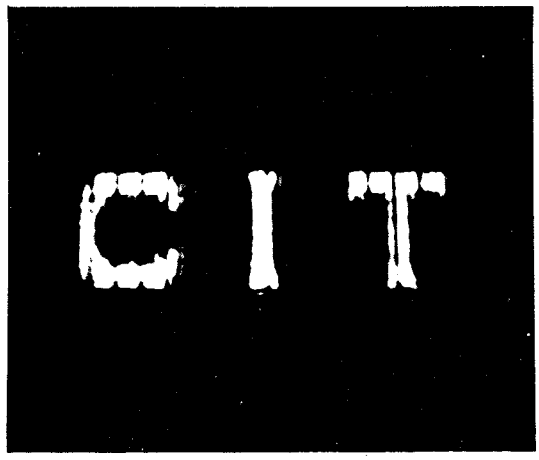


Figure 3.12 : High contrast image obtained from disk using diffractive readout scheme of Figure 3.11a.

As described in the previous section, when we assume that pixels are written on a Cartesian grid, the presence of track curvature leads to positional errors

given by Equation 3.6. In the Schlieren imaging system described above, the positional error of a recorded spot can lead to amplitude and phase errors in its contribution to the reflected field. We neglect the amplitude error since it only becomes significant when the position error of recorded pixels is comparable to the distance between the disk and the image plane. We must consider the phase error, however. Let us write the expression for a track grating written over some set of inner radii now including the effect of track curvature.

$$r(x,y) = r_0 + (r_1 - r_0)rect(r/\delta_y) * comb(r/\Delta_y), \tag{3.13}$$

where we have simply replaced the cartesian approximation for the radial coordinate with the true disk radial coordinate $r = \sqrt{x^2 + (y-R)^2}$. Once again we consider a normally incident plane wave readout beam. The first order diffracted components arise from the fundamental sinusoidal component of $r(x,y)$, which is given by

$$r_{\pm 1}(x,y) \propto sin(r/\Delta_y), \tag{3.14}$$

where for the purposes of this discussion the amplitude of this component is not material. So we have that

$$r_{\pm 1}(x,y) = sin(\sqrt{x^2 + (y-R)^2}/\Delta_y). \tag{3.15}$$

If we assume that the image position R is large compared to the image extents in both x and y (i.e., $R >> x_{max}, y_{max}$) then we have

$$\begin{aligned} r_{\pm 1}(x,y) \approx & sin[\frac{1}{\Delta_y}(y-R)(1 + x^2/2(y-R)^2)] \\ \approx & sin[(\frac{y-R}{\Delta_y}) + \frac{x^2}{2\Delta_y R}]. \end{aligned} \tag{3.16}$$

So, finally we have that

$$r_{+1}(x,y) \propto e^{i(y-R)/\Delta_y} e^{ix^2/2R\Delta_y} \tag{3.17}$$

and

$$r_{-1}(x,y) \propto e^{-i(y-R)/\Delta_y} e^{-ix^2/2R\Delta_y}. \tag{3.18}$$

We see that for the ± 1 order diffracted fields, the track curvature introduces a phase error of $\Phi_{\pm 1} = \pm x^2/2R\Delta_y$. This phase error can be modeled as a cylindrical lens at the disk plane with focal length f=$\pm R\Delta_y/\lambda$. For typical experimental parameters R=4.5cm, Δ_y=1.5μm, and λ=633nm, the cylindrical focal length is 10.7cm. This distortion can be corrected by an illumination system containing a cylindrical lens of focal length F. The product of the incident wavefront and the reflectance function of the disk in this case is

$$E_r(x,y) = E_i(x,y)\tilde{r}(x,y), \tag{3.19}$$

where $E_r(x,y)$ is the reflected field, $E_i(x,y) = e^{j\alpha x^2}$ is the incident field corrected by the cylindrical lens, and $\tilde{r}(x,y)$ is the apparent reflectance function of the disk surface including the phase error. The illuminating optics should be chosen so that $E_r(x,y) = r(x,y)$, which yields $\alpha = -k/2f$. With this value for α, the incident illumination is given by $E_i = exp(-jkx^2/2f)$, which can be generated by a line source located a distance f in front of the disk. A cylindrical lens with focal length F, at a distance $F + f$ in front of the disk, can be used to form the line source.

Correcting for this cylindrical lens effect does not, however, account for the positional errors of pixels due to the polar grid. Earlier we found that the position error of pixels in large images can exceed the pixel spacing. Since some applications require pixel position errors less than the interpixel separation, in order to minimize these errors we must make the interface to the disk conform to this polar recording format. For example, in order to accurately record an image sensed by a television camera, the camera should be modified to scan along curved lines matching the shape of the tracks on the disk.

In addition to the recording and retrieval of images, the optical disk is an ideal medium for the storage of computer generated holograms (CGHs).[97–99] The diffractive imaging technique described above may be thought of as simply the reconstruction of an image plane hologram. Any other computer generated hologram can just as easily be stored on the optical disk. We have investigated various techniques for the calculation and recording of CGHs on the optical disk. Once again however, such techniques are not the focus of this thesis and therefore will not be discussed in detail here. Holographic storage on optical disks has been discussed in References [100-103].

4. Optical Disk Based Image Correlators

As we discussed in Chapter 2, optical correlation for pattern recognition has long been considered a promising application for optical processing, particularly in the area of pattern recognition. One of the reasons such correlators have not been used in practice is that in many practical applications a single filter is not sufficient to produce reliable recognition. In Chapter 2 we saw that this was one of the primary motivations for considering synthetic discriminants and multilayer networks for pattern recognition. In the language of image correlators, a straightforward solution to this problem is the use of spatial and temporal multiplexing to search through a library of filters in order to perform nonlinearly separable dichotomies. The optical disk-based correlation architectures we describe in this chapter provide an extremely efficient method for performing such a search since they combine in a single device the large memory required for storage of the library of reference images, the spatial light modulator needed to represent the reference in the optical correlator, and the scanning mechanism to temporally search through the library.[104,105]

The 2D correlation function $c(\tilde{x},\tilde{y})$ is defined as

$$c(\tilde{x},\tilde{y}) = \mathcal{F}^{-1}\{F(\omega_x,\omega_y)G^*(\omega_x,\omega_y)\} \tag{4.1}$$

$$= \int\int f(x,y)g(x-\tilde{x},y-\tilde{y})dxdy, \tag{4.2}$$

where $f(x,y)$ and $g(x,y)$ are two real images, $F(\omega_x,\omega_y)$ and $G(\omega_x,\omega_y)$ are their respective 2-D Fourier transforms and $\mathcal{F}^{-1}\{\}$ is the 2-D inverse Fourier transform operator. It is well known that $c(\tilde{x},\tilde{y})$ is sharply peaked at the point (x_0,y_0) when

$f(x,y) = g(x-x_0, y-y_0)$. This property makes the correlation function useful for pattern recognition because, regardless of the position of the input image $g(x,y)$, $c(\tilde{x},\tilde{y})$ will have a peak if $f(x,y)$ and $g(x,y)$ are matched. Since in general there are many versions of an image $g(x,y)$ that we would like to recognize, a reliable image recognition system should provide invariance to multiple object attributes such as scale and rotation. Often the best way to achieve this invariance is to use a large number of reference patterns $f(x,y)$ against which to compare $g(x,y)$ in order to obtain reliable recognition.

Optical image correlators based on Fourier Transform (FT) holograms were proposed by Vander Lugt in 1964.[50] In order for optical correlation to be a realistic approach to image recognition, we require a memory device sufficient to store a large reference image library, an SLM that interfaces with this memory in real time, and a scanning or addressing mechanism that allows interrogation of the entire reference library in a reasonable amount of time. The optical disk provides these three characteristics in one device. In this chapter, we describe several optical disk-based image correlation architectures and present experimental results taken from selected systems. We will examine critical parameters associated with each architecture and evaluate each system in terms of power and speed.

4.1 Rotating Mirror Correlator

We first discuss two systems that perform 2-D correlations based on Equation 4.2. These systems are referred to as image plane correlators. In this architecture the correlation function is generated by calculating an inner product for every relative shift between the input and reference images. Since these shifts will be

generated sequentially, the correlation will appear as a 1-D signal representing a *raster* version of the desired 2-D correlation pattern. As we will see, these systems sacrifice correlation rate for operational simplicity without alignment criticality, while at the same time relieving source coherence requirements. As with any incoherent correlator however, the present system is a unipolar architecture and some bias removal mechanism is necessary in order to retrieve a bipolar or high SNR correlation signal. Such bias removal techniques have been discussed at length in the literature as well as in the previous chapter.

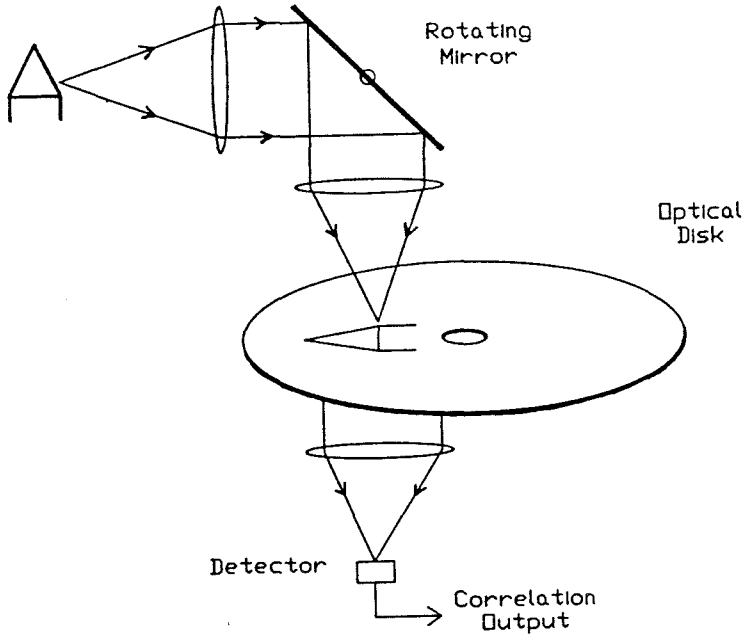


Figure 4.1 : Rotating mirror image plane correlator.

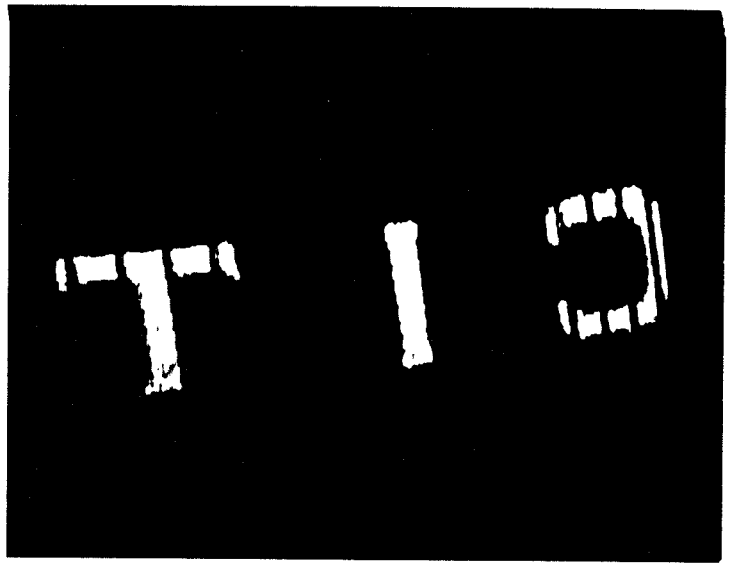
A simple image plane correlator is shown in Figure 4.1. An image of the input scene is formed at the optical disk on which a library of reference images resides. The total transmitted or reflected light is collected by a detector at the output. The rotation of a polygon mirror causes the input image to scan the disk radially

while the disk rotation itself provides scanning in the orthogonal direction. The detected light therefore, represents the instantaneous inner product between the input and a shifted version of one of the references. All relative shifts between input and reference images are generated using these two scanning mechanisms and as a result, the light collected at the output is exactly a 1-D raster signal of the desired 2-D correlation. In fact, the *rotation* of the disk does not provide the *translation* required to calculate the true correlation signal; however, the error induced arises from a reference image rotation equal to the reference image azimuthal extent. After superimposing the original reference image and a rotated version of this image, we see that the rotation induced error is given by $\epsilon = N_\theta tan(N_\theta \delta_\theta)$ pixels, where N_θ is the number of reference pixels in the azimuthal direction and δ_θ is once again the angular separation between reference pixels. If the reference image comprises $N_\theta = 100$ pixels in the azimuthal direction then it subtends an angle of $0.1°$ which is equivalent to less than one pixel of error ($\epsilon = 0.2$ pixels) in a 100×100 pixel image and may therefore be neglected. An example of the output obtained from the rotating mirror correlator is shown in Figure 4.2 along with a computer simulation of the desired autocorrelation function. In this demonstration of the rotating mirror correlator, the optical system consisted of a single imaging lens with a focal length of $10cm$ and an aperture of $5cm$, which provided a magnification of $1/20$ from the input plane to the disk plane. Figure 4.2a is the reference image written on a Sony write once disk. This image has up to 6912 pixels along track and comprises 1024 tracks. The actual pixel size in this reference ($\approx 100\mu m \times 40\mu m$) is quite large compared to both the disk resolution and the diffraction limited resolution of the imaging

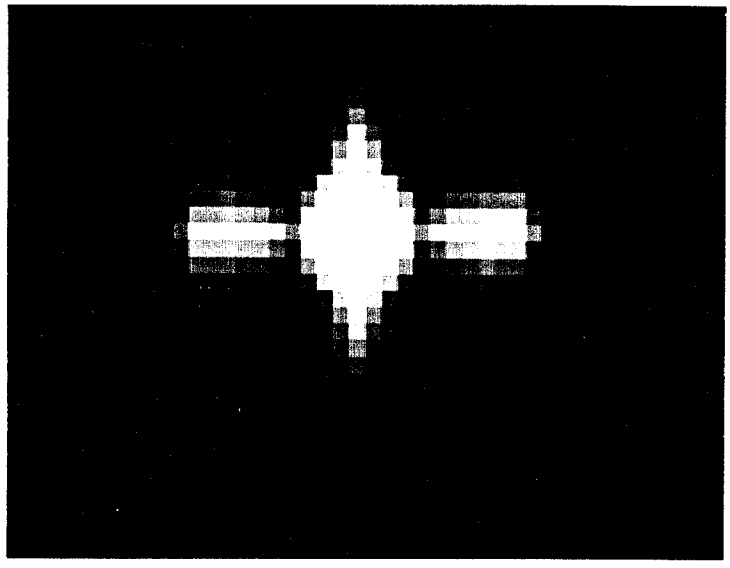
system, so effects due to these limitations were not observed. The input to the system was provided by a transparency of the acronym CIT illuminated by a HeNe laser ($\lambda = 633nm$). Figure 4.2d shows the correlation signal generated by the optical system for this input. The asymmetry in the optical autocorrelation is due to a slight mismatch between the input and reference images.

This system can operate with incoherent light, which simplifies the illumination source requirements and eliminates speckle. However, one critical limitation is its speed. The speed in this architecture is dictated primarily by the rates of the relative scanning mechanisms. Specifically, in order to generate an accurate correlation signal, the radial scan time (i.e., the time for the input image to shift radially past the reference image) must be less than the time it takes the reference to rotate by 1 pixel. The correlation rate is thus limited by factors such as reference image pixel size in the azimuthal direction Δ_θ (m), radial scan rate $\sigma(s^{-1})$, disk rotation speed $\omega(s^{-1})$, and ultimately by illumination level and disk efficiency. These factors may be used to calculate the radial scanning limited correlation rate (ρ_c). As stated above, we require that the time to complete one radial scan, $t_s = 1/\sigma$, be less that the time required to rotate the disk by one reference pixel t_r where t_r is given by the ratio of the time to complete one full rotation to the number of pixels in a circumference. So we have that $t_r = \Delta_\theta / 2\pi r_0 \omega$ where r_0 is the disk radius at which the reference image is written. We also have that t_s and t_r must satisfy $t_s \leq t_r$, which implies that

$$\sigma \geq 2\pi\omega r_0/\Delta_\theta, \tag{4.3}$$



(a)

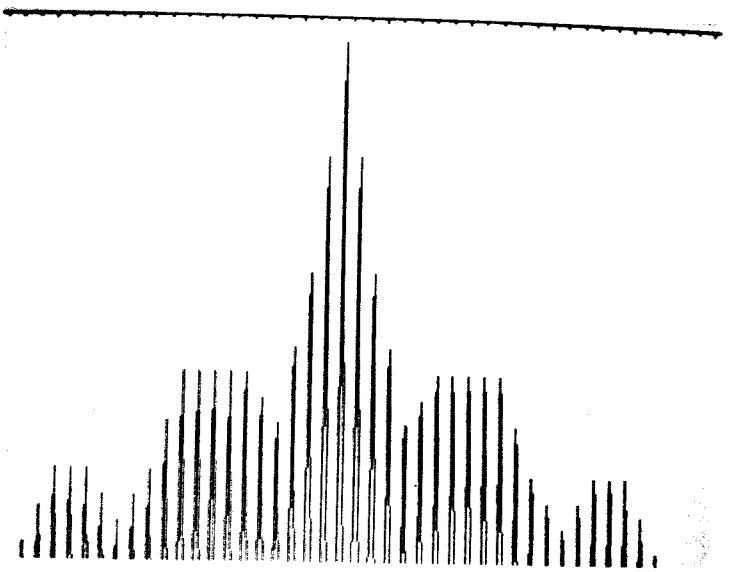


(b)

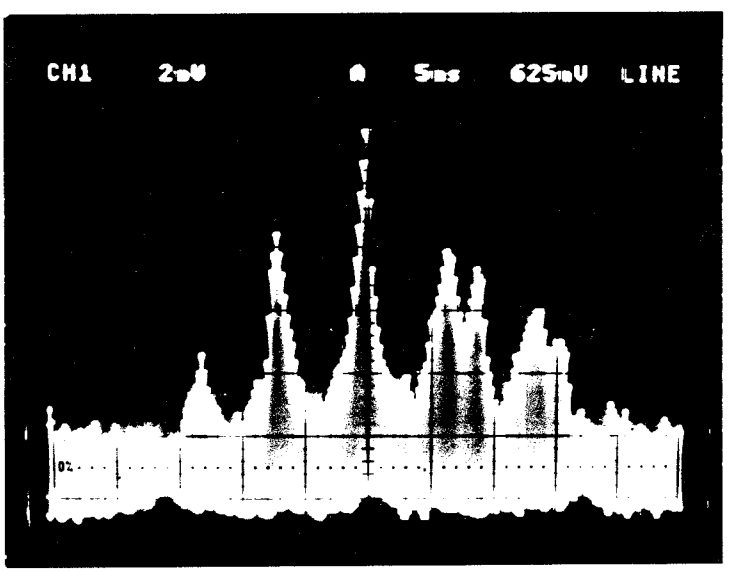
Figure 4.2 : Rotating mirror correlator results.

(a) Reference image stored on Sony WORM disk.

(b) 2D autocorrelation pattern of (a) by computer simulation.



(c)



(d)

Figure 4.2(cont) : Rotating mirror correlator results.

(c) Raster signal of 2D pattern shown in (b).

(d) Optical system output.

or equivalently

$$\omega \leq \sigma \Delta_\theta / 2\pi r_o. \tag{4.4}$$

These expressions yield the rotating mirror rotation rate σ required given the various disk parameters, or equivalently, the limit on disk rotation rate for a fixed mirror scan rate. In our system we utilized a rotating polygon mirror system from Lincoln Laser Company, which incorporates a 10-sided mirror capable of rotating at 400Hz yielding a radial scan rate of $\sigma = 4kHz$. In this experiment, the speed of the rotating mirror is the limiting factor, and it results in a maximum correlation rate of 40 correlations/sec for 100×100 pixel images.

Another limitation of the rotating mirror correlator results from the behavior of the image plane during scanning. Ideally, the optical disk *is* the image plane of the optical system responsible for imaging the input onto the reference image. In reality however, the image plane tilts out of the disk plane as the mirror rotates. In order to determine the effect of this behavior on the correlation signal we will consider the geometry shown in Figure 4.3. The relationship between mirror rotation angle θ and image plane tilt α is found to be linear

$$\alpha = 90° - 2\theta, \tag{4.5}$$

so that as we would expect, when $\theta = 45°$ the image plane and the optical disk plane coincide. For $\theta \neq 45°$ however, the parameter of import is the worst case image plane displacement Δ shown in Figure 4.3. For a given reference image extent W, Δ is determined by the image plane tilt obtained for the *maximum* scan angle required as shown in Figure 4.4a. The maximum scan angle will be

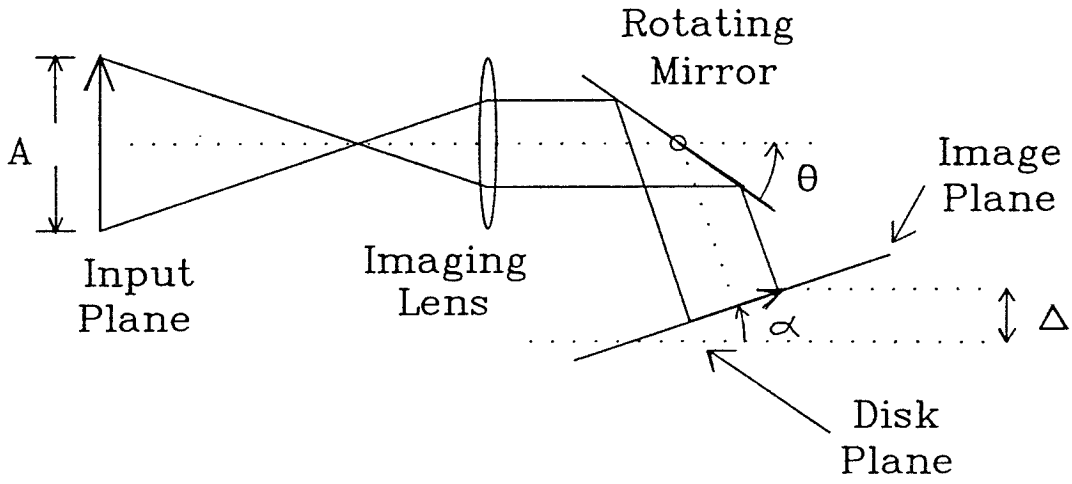


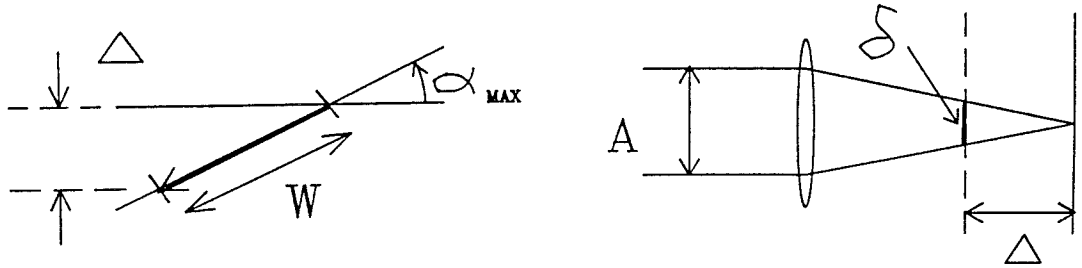
Figure 4.3 : Geometry used to determine the effect of image plane tilt on correlation signal quality.

minimized when the distance from the mirror to the image plane is maximized yielding a *best* case maximum scan angle of $\theta_{max} = 45° + \frac{1}{2}tan^{-1}(W/i)$ where i is the image distance of the imaging system and is approximately equal to the lens focal length f, since the system realizes a large demagnification. From this we can derive that

$$\Delta = W sin(\alpha_{max}) \tag{4.6}$$

$$= \frac{W^2}{\sqrt{W^2 + f^2}}. \tag{4.7}$$

Going one step further, we can find the image blurring, which arises from this image plane error as $\delta = A\Delta/f$, where δ and A are defined in Figure 4.4b and we have once again made use of the approximation $i \approx f$. Notice that δ is



(a) (b)

Figure 4.4 : Image plane tilt geometry.

(a) Maximum image plane displacement Δ.

(b) Worst case image blurr arising from image plane tilt.

inversely proportional to the diffraction limited spot size $\delta_D = 1.22\lambda f/A$. Using some typical system parameters such as $f = 10cm$, $W = 1mm$ and $A = 5cm$ we arrive at a diffraction limited spot size of $\delta_D = 1.5\mu m$ and an offset $\Delta = 10\mu m$, which corresponds to a blurred spot size of approximately $\delta = 5\mu m$. This is *not* an attractive result since the optical disk itself provides a resolution of $1\mu m$ yet we cannot take advantage of all this available SBP due to imperfect imaging. This limitation was not seen in the experimental results presented here since the effective reference pixel size was much greater than $5\mu m$. We should note here that in a practical implementation of the rotating mirror correlator, a multiple lens systems can be designed to eliminate image plane tilt.

The last issue that we will address is power efficiency. Although it is not

likely that power will be a limitation in this type of system due the relatively long integration times available, we will include a brief power budget here for completeness and for comparison to later systems. There are several sources of loss in the present system. The most dominant loss mechanism arises from low disk diffraction efficiency (η_D) and is due to the relatively poor disk contrast as discussed in Chapter 2. If the incident intensity is given by I_0 and we are primarily interested in detecting the autocorrelation *peak* obtained using the optical correlator considered in this section, then the autocorrelation peak *energy* is the most important system characteristic. If we further assume that we are using random, unipolar, binary input and reference images that are characterized by an equal number of ones and zeros on average, then we have that the expected autocorrelation peak energy E_P is given by

$$E_P = I_0 \eta t_D / 2, \tag{4.8}$$

where η is the optical system efficiency and t_D is the peak dwell time given by $t_D = 1/(N_r \sigma)$ where N_r is the number of reference pixels in the radial direction. For 100×100 images, we previously derived a correlation rate of 40 images/sec, which yields a corresponding peak dwell time of $t_D = 2.5\mu s$ or a detector bandwidth requirement of $\approx 0.8MHz$. Assuming $\eta = \eta_D$ we obtain an autocorrelation peak energy of $E_P = 12.5pJ$. This peak energy is equivalent to a photon count of $n_p = 4 \times 10^7$ photons. As predicted, this is an easily detectable signal making the rotating mirror correlator the most power efficient of the correlation architectures we will discuss.

4.2 Acousto-Optic Correlator

The speed limitation of the previous system is due primarily to the mass of the rotating mirror. Since the speed of commercially available polygon mirror-based scanners is limited to about 40kHz, it is natural to consider the use of an alternative radial scanning mechanism in order to improve the performance of the image plane correlator. In this section we will describe an image correlator based on the same mutual scanning approach taken in the previous section; however, in the present system we will replace the rotating mirror with an acousto-optic radial scanning device that can achieve scan rates up to 10MHz. The AO device provides the high speed scanning we require. Once again we obtain an advantage in this architecture in that it operates on light intensity and consequently the requirements for phase uniformity are greatly relaxed. As a result it is possible to implement this architecture as well as the previous one with most existing disk systems. The Acousto-Optic Correlator (AOC) described here suffers from more acute wavelength sensitivity than its rotating mirror counterpart owing to the diffractive nature of the imaging system used; however, the use of a quasi-monochromatic source is possible and still eliminates the need for laser illumination.

A schematic diagram of the AOC is shown in Figure 4.5. In this architecture, a chirp signal propagating in the AO device generates a moving cylindrical lens with power in the horizontal dimension. This moving cylindrical lens becomes part of the system that images the input onto the disk. Consequently, as the AO lens moves horizontally, the image formed on the disk is scanned radially. This behavior will be described in more detail below. A second lens, which provides

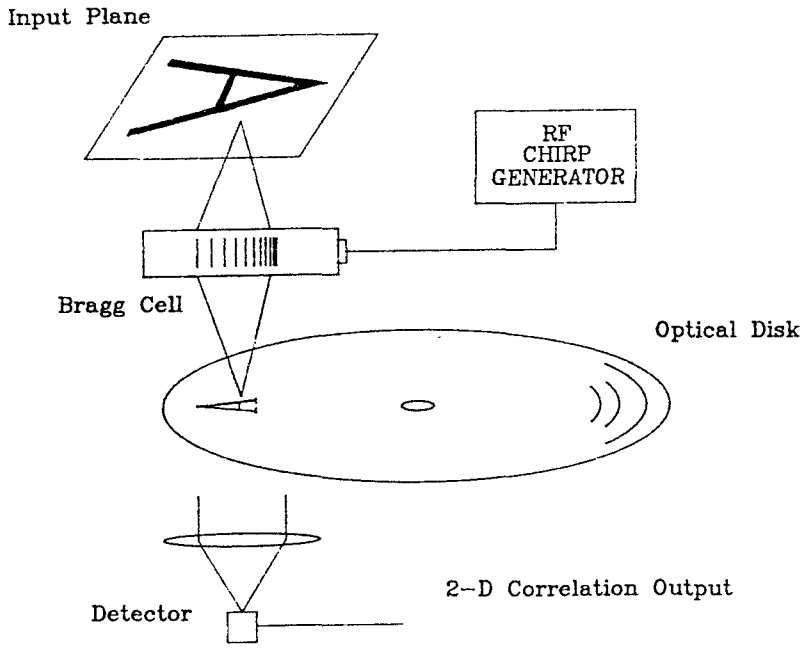
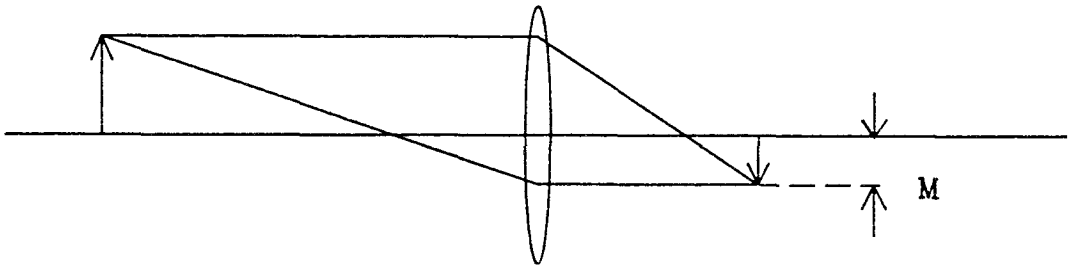


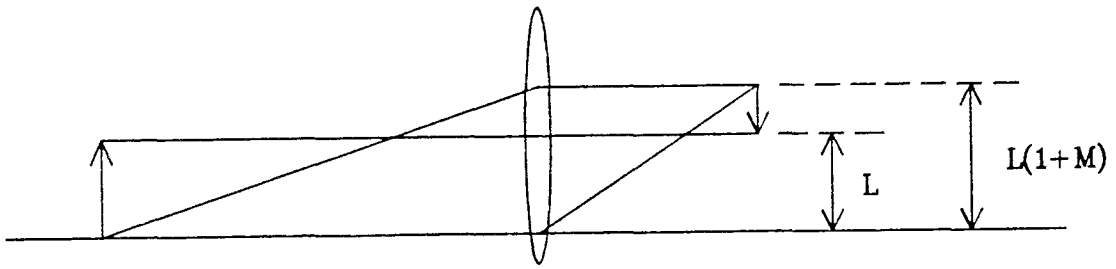
Figure 4.5 : Schematic diagram of disk-based AOC.

imaging in the other direction, is omitted from the drawing for clarity. As before, the scanner provides the relative displacement in the radial direction between the input and reference images that is necessary to calculate the correlation function and the disk rotation provides scanning in the orthogonal direction. The reflected or transmitted light is once again collected (integrated) on a single detector, which produces as its output a temporal video signal of the 2-D correlation.

In order to understand the operation of the AO moving lens scanner, consider the imaging systems shown in Figure 4.6a and Figure 4.6b. Neglecting finite aperture effects, and under the paraxial approximation, the effect of a shift of the lens L1 in a plane perpendicular to the optical axis, is a corresponding shift in the image. The magnitude of the image shift is equal to the magnitude of the lens shift multiplied by the absolute value of the magnification $m = i/o$ and the image



(a)



(b)

Figure 4.6 : Scanning imaging system using a moving lens.

(a) Lens on axis.

(b) Lens off axis yields shifted image.

shift is in the same direction as the lens shift. This is shown in Figure 4.6b. For our application, we would prefer that our moving lens have a very small mass so that its velocity might be maximized. This is accomplished through the use of an Acousto-Optic device (AOD) or Bragg cell. A Bragg cell is a device consisting of an optically transparent photoelastic crystal to which a piezoelectric transducer is attached. An electrical signal applied to the transducer results in a mechanical deformation proportional to the applied field. The deformation of the transducer in turn causes a pressure wave to propagate in the photoelastic medium resulting in a propagating refractive index variation via the photoelastic effect. This index variation can be used to modulate an incident optical field.[106–108] In this way, an electrical signal applied to the AOD can influence light. The details of this AO interaction will not be discussed here; however, in order that the operation of the AOC might be better understood, we will derive the behavior of the AO lens.

Consider the system shown in Figure 4.7, which comprises an AOD to which we have applied an RF chirp signal with chirp rate b. An RF chirp is an electrical signal whose frequency ramps linearly for time t_c (the chirp time) from a starting frequency f_1 to an ending frequency $f_2 = f_1 + bt_c$. This applied signal induces a moving chirp grating in the AOD. We have "frozen" this propagating index variation in time and will consider the effect of this stationary grating on a plane wave incident from the left. We first consider diffraction from the leading edge of the chirp signal. This portion of the chirp grating will be determined by the chirp starting frequency f_1 and corresponds to an index grating of spatial frequency $u_1 = f_1/v_a$ where v_a is the acoustic velocity of the index grating in the AOD.

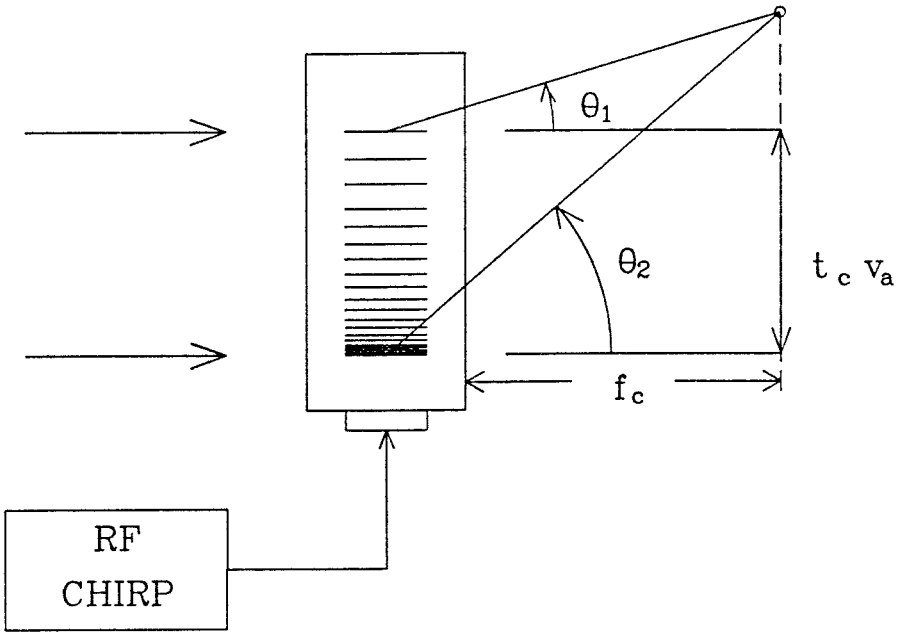


Figure 4.7 : Derivation of AO lens focal length.

Light incident on the leading edge of the chirp grating will therefore be diffracted at an angle of $\theta_1 = sin^{-1}(\lambda u_1)$ as shown in the figure. Similarly, the trailing edge of the index grating is characterized by a spatial frequency of $u_2 = f_2/v_a$, which results in a diffracted component at an angle of $\theta_2 = sin^{-1}(\lambda u_2)$. From the figure we can see that the chirp grating acts like a lens whose focal length is determined from the angles θ_1 and θ_2. Clearly, if $\theta_1 = \theta_2$ then we have only a single frequency component traveling in the AOD and no focusing will occur; however, if $\theta_1 \neq \theta_2$ then the chirp grating will behave like a lens whose focal length is given by

$$f_c = v_a^2/\lambda b. \tag{4.9}$$

This is a well-known result in AO signal processing and is obtained by a simple consideration of the geometry shown in Figure 4.7. The preceding analysis has not

included the effect of Bragg mismatch on the AO lens behavior. In general, Bragg cells are made *thick* so as to increase diffraction efficiency by virtue of a *volume* optical interaction. In our case however, we would like to ensure good lenslike behavior at the expense of diffraction efficiency. This means that we would like to calculate the AOD thickness required to ensure that *both* ends of the AO lens are simultaneously Bragg matched. In order to perform this calculation we must make use of the fact that for a volume hologram, the angular bandwidth $\Delta\theta$, over which an incident field of wavelength λ will remain Bragg matched, is given by

$$\Delta\theta \leq sin^{-1}(2\lambda/\pi L), \tag{4.10}$$

where L is the hologram thickness. For small $\Delta\theta$ we can derive the maximum thickness for which the AO lens analysis is valid as

$$L_{max} \leq 2v_a/\pi b t_c. \tag{4.11}$$

For parameters typical of the application being discussed here, we have $t_c = 7\mu s$, $b = 4MHz/\mu s$, and $v_a = 1mm/\mu s$ yielding maximum thickness of $L = 23\mu m$ for good Bragg matching of the lens edges. For this reason the remaining analysis in this section will assume that the moving chirp grating is thin. While this assumption is not entirely accurate for commercial devices, a more detailed analysis would be required in order to exactly model the acousto-optic interaction in a real device. To further complicate this situation commercial devices make use of acoustic beam steering in optically anisotropic crystals in order to achieve decreased Bragg sensitivity. The result of such techniques is that relatively thick

devices *can* be used to realize a thick moving lens; however, for our purposes the thin moving lens assumption will suffice. We will see later using our simple model, how the AOD SBP and the chirp bandwidth combine to determine the resolution available in the disk plane of the AOC. It should be noted here that the above analysis of the AO lens assumed that a stationary chirp grating was present in the AOD. For a moving grating as would appear in the present application, we merely generate the desired moving lens together with a slight doppler shift of the diffracted light. This doppler shift will not affect the operation of the AOC and will not be discussed further.

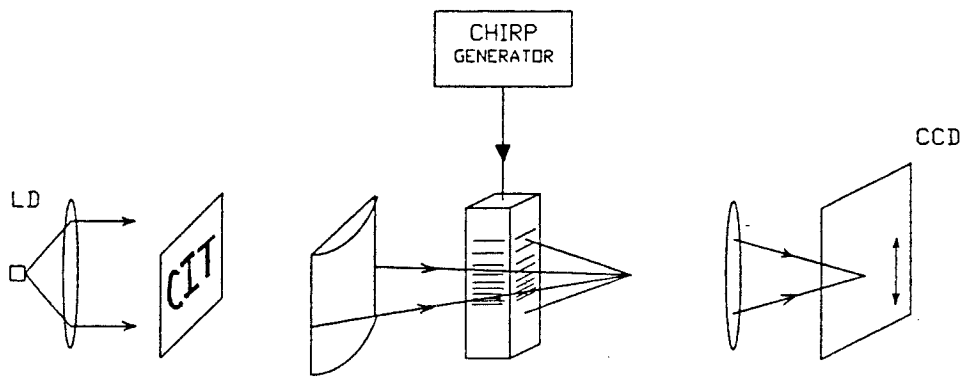
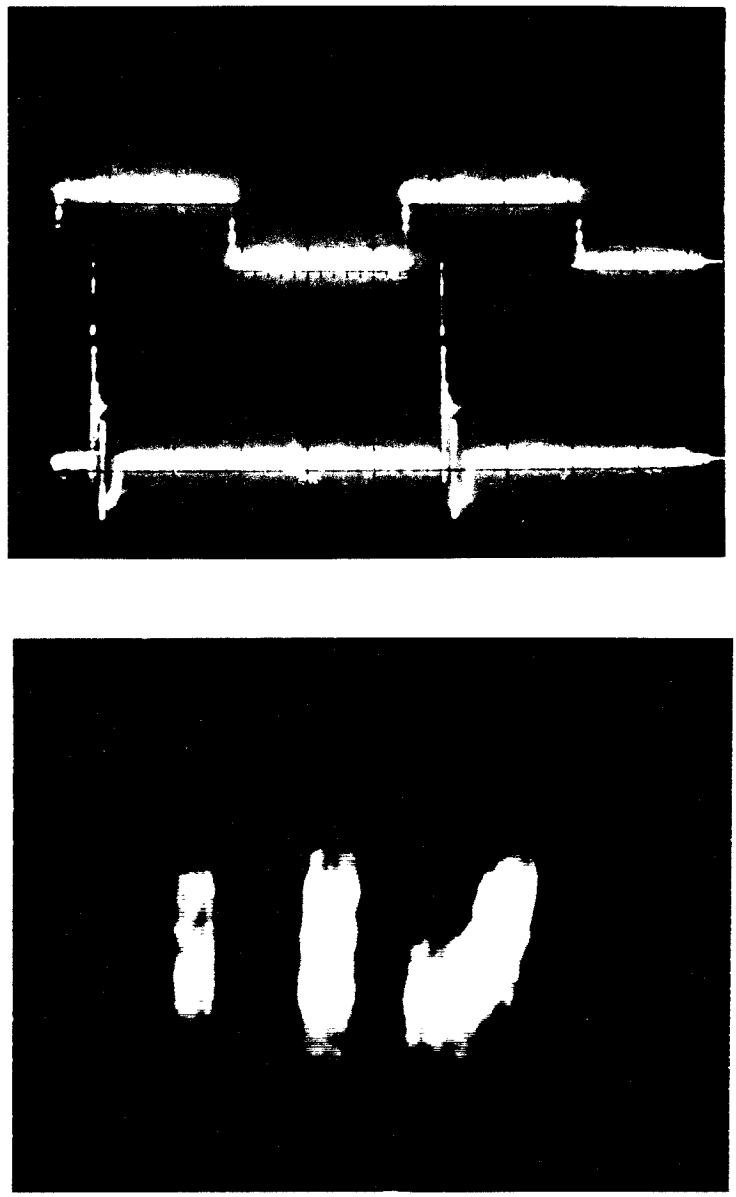


Figure 4.8 : AO lens impulse response measurement setup.

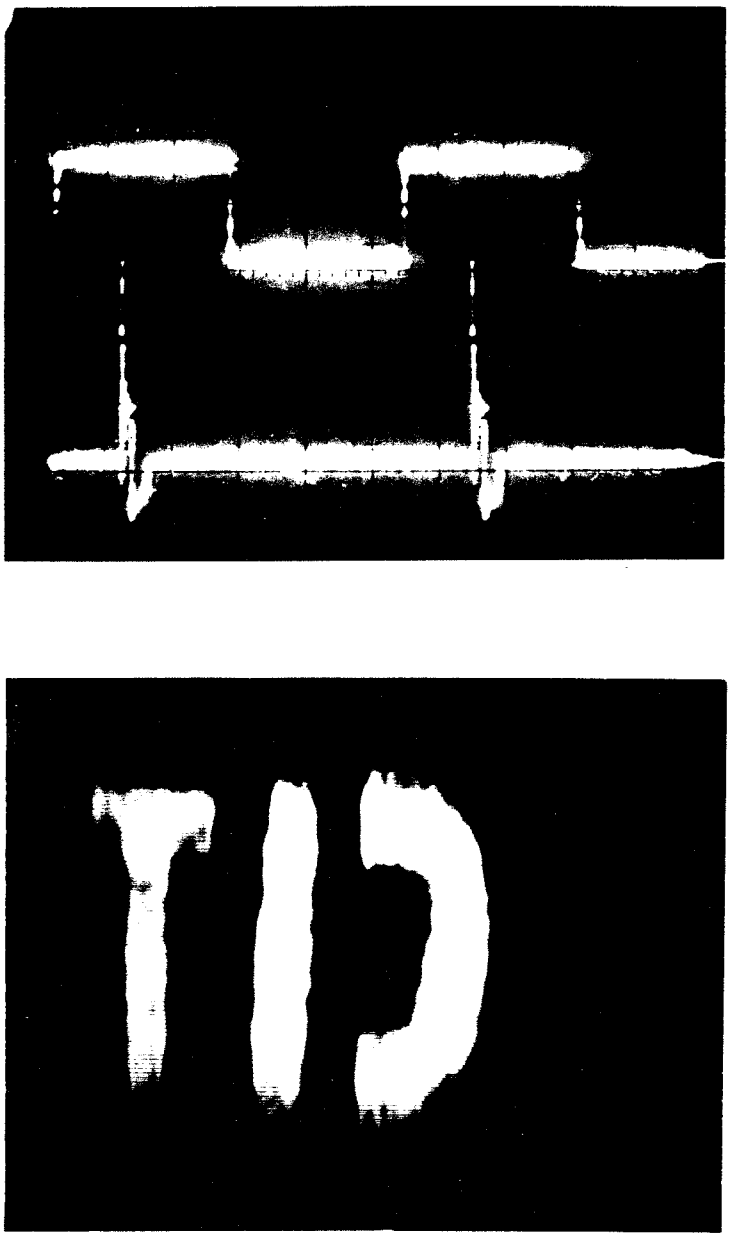
In order to demonstrate the AOC, we first constructed the system shown in Figure 4.8 to evaluate the AO lens. The system shown in the figure is a simple imaging system incorporating the AO lens. A chirp rate of $b = 23MHz/5\mu s$ was



(a)

Figure 4.9 : AO moving lens impulse response.

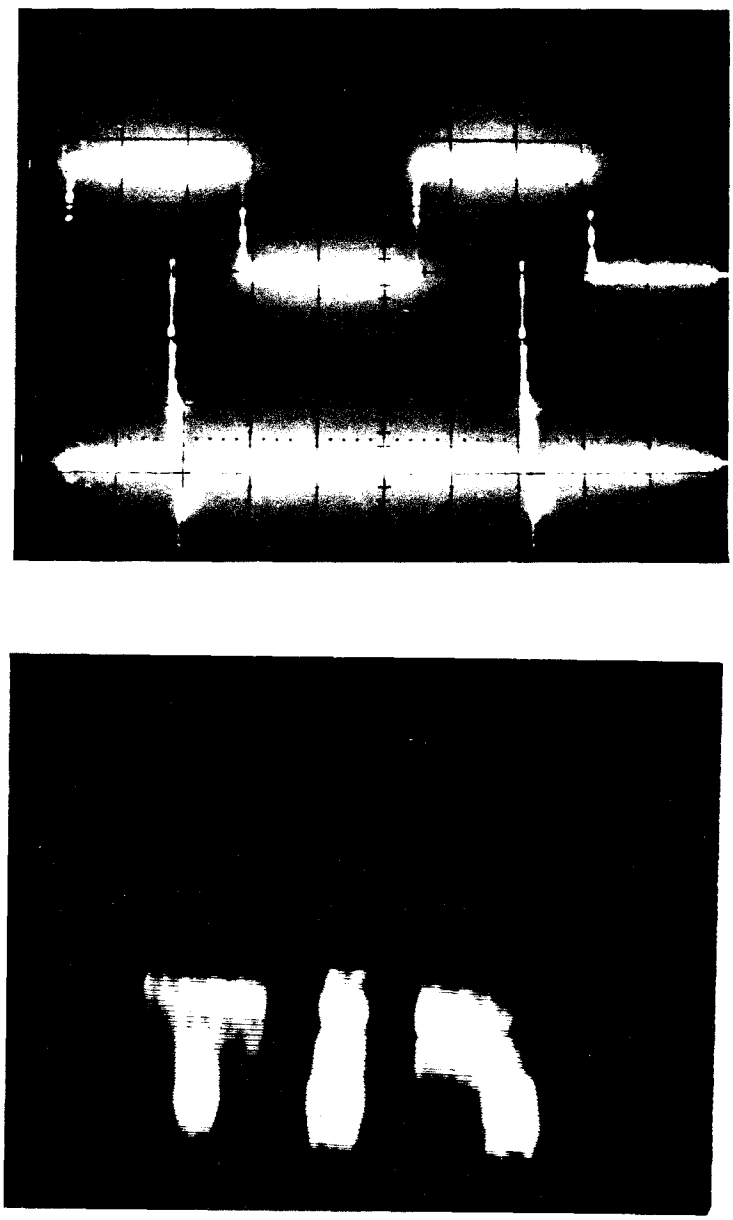
(a) Delay time = 2.5μs [top = chirp gate and LD pulse, bottom = CCD image].



(b)

Figure 4.9 (cont)

(b) Delay time = 5.0μs [top = chirp gate and LD pulse, bottom = CCD image].



(c)

Figure 4.9 (cont)

(c) Delay time = 8.0μs [top = chirp gate and LD pulse, bottom = CCD image].

used together with a TeO_2 Bragg cell with $v_a = 1mm/\mu s$ yielding an AO lens focal length of $f_c = 40cm$. The impulse response of the AO lens scanner is shown in Figure 4.9. The image of an input transparency was formed on a CCD using one cylindrical lens with power in the horizontal dimension and the AO lens for vertical imaging. The input illumination was pulsed so that the AO lens might be frozen in various vertical positions. The delay between the onset of the RF chirp and the laser diode pulse determines the position of the image on the CCD. Figures 4.9 (b-d) show the CCD output for various delay values. As can be seen from the figures, the AO lens imaging characteristics are quite good.

The AO chirp scanner acts as a traveling lens that scans the diffracted image at a rate equal to the acoustic velocity. The correlation rate in the AOC is still constrained by the radial scan speed; however, since this scanning is generated by virtue of the propagation of an RF chirp in the Bragg cell, the resulting correlation rate is much higher than before. The RF chirp parameters are chosen so as to utilize as much AO space-bandwidth product as the input image requires while minimizing scan time. Specifically,

$$t_c = t_{AO}(SBP_{IN}/SBP_{AO}), \tag{4.12}$$

where t_c is the required RF chirp duration, t_{AO} is the AO aperture and SBP_{IN} and SBP_{AO} are the input image and AO device space-bandwidth products respectively. We have built this correlator using a TeO_2 AO cell with a $70\mu s$ aperture and an RF chirp centered at 40MHz with a chirp rate of ≈ 4 MHz/μs. Using the above equation with $SBP_{IN} = 100$ and $SBP_{AO} = 1000$ the required chirp duration was calculated to be $7\mu s$. A SAW device was used to generate the

desired chirp signal. The resulting radial scan rate of $1/7\mu sec \approx 140$kHz, yields a correlation rate of 1400 correlations/sec for 100 $\times$ 100 pixel images.

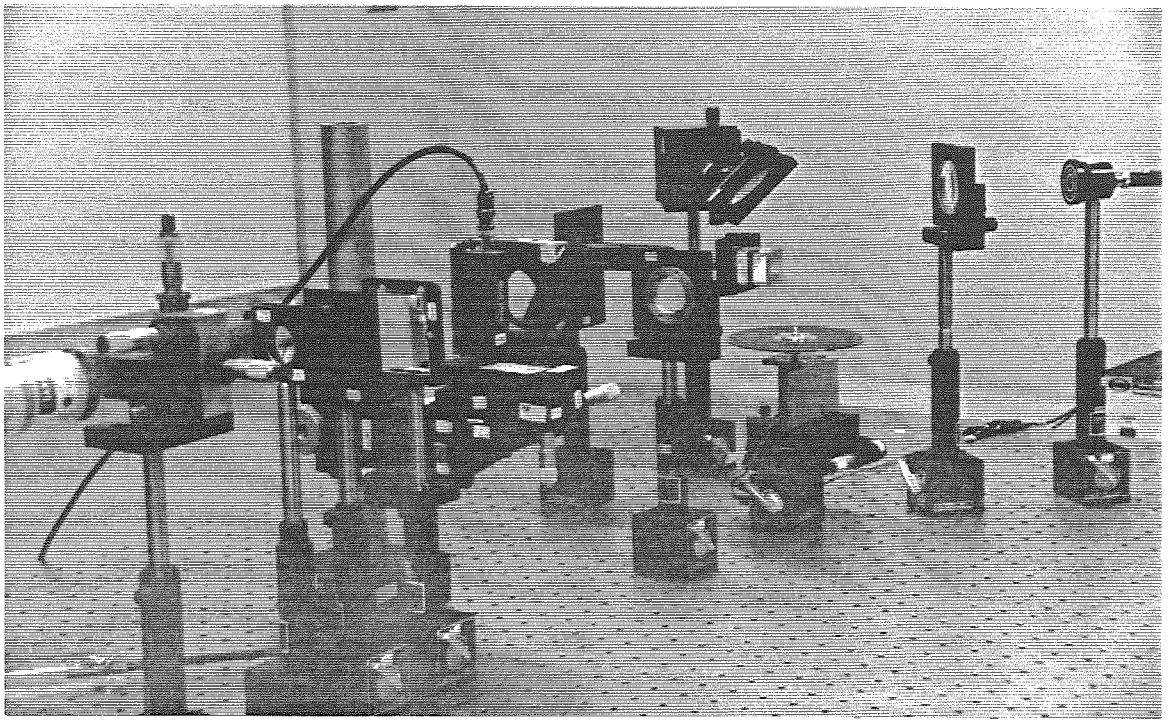
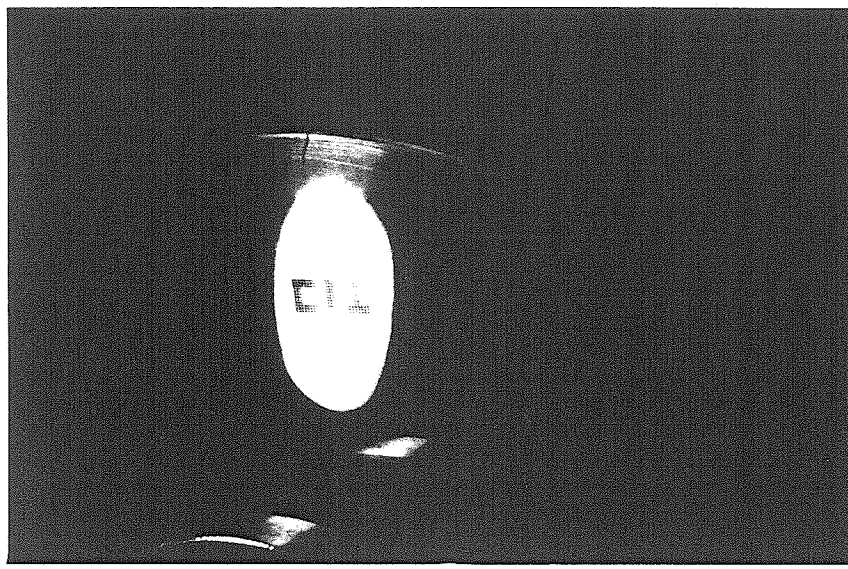
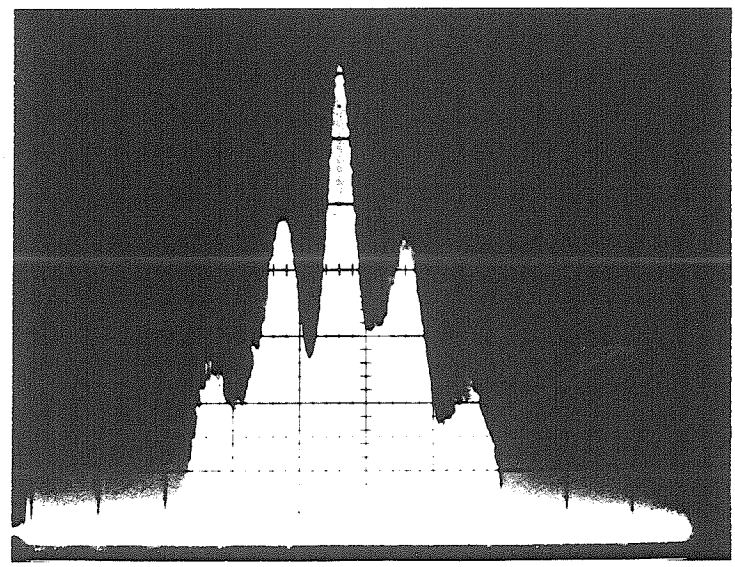


Figure 4.10 : Photograph of AOC experimental apparatus.

A photograph of the AOC is shown in Figure 4.10 and an example of experimental results obtained using this system is shown in Figure 4.11. The input to this system was, once again, a transparency of the acronym CIT, and the reference was a duplicate CIT written on an Optotech WORM disk using the simple recording system described in Chapter 2 because the Sony system was not yet available at the time of these experiments. The reference image shown in Figure 4.11a is relatively large so that a radial scanning distance of $2cm$ was required in order to generate an accurate correlation signal. Figure 4.11b is an oscilloscope trace of the detector signal produced by the optical system of Figure 4.10, and Figure 4.11c shows the same magnified to reveal the individual correlation



(a)

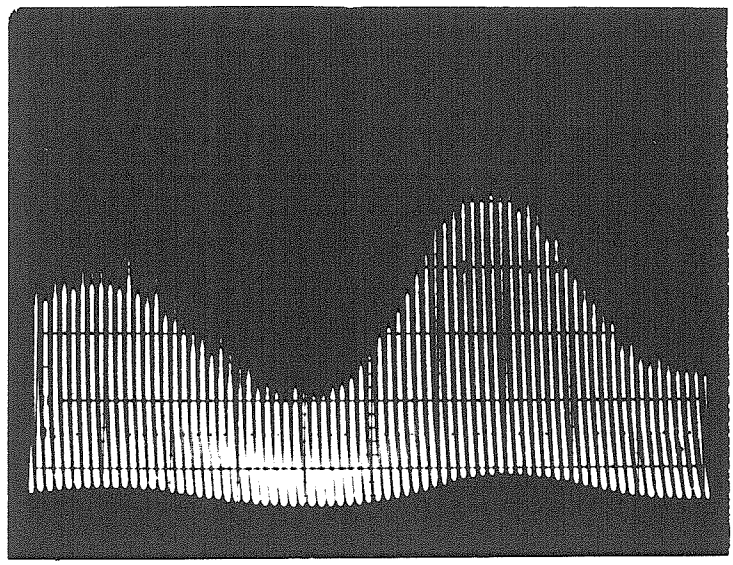


(b)

Figure 4.11 : AOC results.

(a) Reference image written on Optotech WORM disk.

(b) Correlation signal obtained from AOC.



(c)

Figure 4.11 (c) Enlarged view of (b) showing individual radial scan lines.

lines produced by the acousto-optic scanner. The format of the detector signal is similar to a video signal of the 2-D correlation and it can be displayed in 2-D by raster scanning the detector output on a 2-D monitor. As can be seen from Figures 4.11b and 4.11c, the optical system output agrees well with the predicted autocorrelation signal of Figure 4.2c showing that the optically calculated correlation is in good agreement with the expected autocorrelation function of the CIT pattern. Correlations can be produced with our experimental apparatus at rates up to 1400, 100X100 pixel reference images per second. Once again this is an incoherent architecture; however, bipolar input and/or reference images can also be represented by adding a bias at the input stage and subtracting it from the output as we described in Chapter 2. It should also be noted here that the image plane tilt inherent in the previous architecture is not present here thereby

eliminating the resolution limit arising from image blurring. However, this system suffers from several resolution limiting tradeoffs, which we will now discuss together with the efficiency of this image plane correlator.

There are essentially three effects which in our simple model, limit the resolution of the AO lens-based imaging system used in the AOC. First, large diffraction efficiency η_{AO} requires a thick AO crystal (i.e., large L) which in turn results in Bragg mismatch for large chirp bandwidth Δf. So we have that high η_{AO} requires small Δf which results in a long focal length and correspondingly poor resolution. This effect points out a clear trade off between resolution and efficiency, which requires that we sacrifice correlation peak detectability for disk plane resolution. The second resolution limitation arises from our desire to maximize the radial scan rate in order to achieve a high correlation rate. Large scan rate corresponds to a small AO chirp time t_c which results in a small lens aperture and once again, correspondingly poor resolution. Again, we may trade off correlation rate for disk plane resolution. These two limitations are manifestations of a more fundamental limit related to the AOD bandwidth. This limitation may be expressed as

$$\delta_D = 1.22\lambda f_c / A_c, \tag{4.13}$$

where δ_D is the diffraction limited resolution and A_c is the chirp lens aperture given by $A_c = t_c v_a$. We can express the above equation in terms of the AO bandwidth by substituting the expression for f_c derived earlier (see Equation 4.9). We obtain

$$\delta_D = 1.22 v_a / b t_c, \tag{4.14}$$

where bt_c is upper bounded by the AO bandwidth (BW). We can calculate the

important AOC system parameters for two commercially available AO devices. For the TeO_2 AOD used in our experiments, we have $SBP_{AO} = 1000$ and an aperture $t_a = 70\mu s$ yielding a minimum chirp time of $t_c = 7\mu s$ for 100×100 images. The corresponding radial scan rate is $0.14MHz$ leading to a correlation rate of 1400 images/sec. With $v_a = 1mm/\mu s$, and $BW = 30MHz$ we obtain a diffraction limited resolution of $\delta_D = 40\mu m$ which once again is *not* an attractive result from the perspective of disk SBP utilization. If we consider using a $LiNbO_3$ Bragg cell with $SBP_{AO} = 100$, and an aperture of $t_a = 1\mu s$, then we can achieve a radial scan rate of $1MHz$ corresponding to a correlation rate of 10,000 images/sec. This same device has an acoustic velocity of $v_a = 6.6mm/\mu s$ and a bandwidth of $BW = 1GHz$ yielding a resolution of $\delta_D = 8\mu m$ which approaches the resolution limit induced by image plane tilt in the rotating mirror correlator of the previous section while improving on the correlation rate of that system by almost three orders of magnitude. Using the same procedures and image statistics as were assumed in the previous section, we can calculate the expected autocorrelation peak energy for this correlator. Assuming an AO diffraction efficiency $\eta_{AO} = 10\%$, $I_0 = 10mW$ incident illumination, and using the $LiNbO_3$ peak dwell time of $t_p = 10ns$ we obtain a peak energy of $E_p = 5fJ$ or a peak detectability $n_p = 1 \times 10^4$ photons.

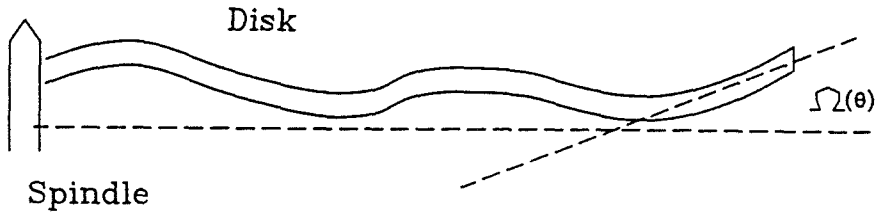
4.3 Effect of Disk Imperfections on Image Plane Correlators

The optical disk systems that are commercially available have been designed to specifications consistent with a serial readout application. Error correction coding (ECC) is extensively used to deal with media defects and feedback is incorporated to offset mechanical imperfections such as wobble, center offset, and coating thickness variations in real time. These same error sources and imperfections can affect the performance of the image plane correlation architectures we have just discussed as well as the Fourier plane architectures we will be discussing next. Although revised design specifications for commercial systems would eliminate the need for much of the following discussion, the SONY system used in this work *does* suffer from the above mentioned limitations and therefore, a discussion of the effect of such imperfections is included.

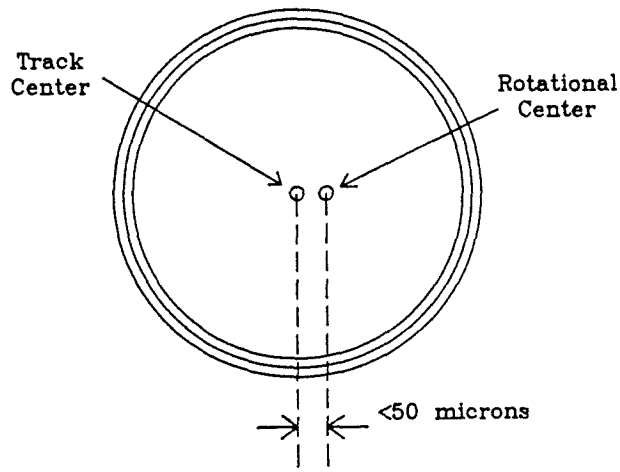
In Chapter 3 we characterized the optical disk in terms of the uniformity of its coating. We found that for glass covered Sony disks, thickness variations were negligible over areas of several mm^2. Since areas of this size may contain up to a million pixels, we concluded that coating thickness variations will not play a role in determining the behavior of the systems discussed here; therefore, thickness variations will not be considered in this section as sources of error in correlation architectures. It should be noted that the thickness variations measured in Chapter 3 could be reduced even further by more careful processing during manufacturing, thereby improving the phase characteristics of the optical disk SLM for other applications. In contrast to the lack of care taken to insure good coating thickness uniformity in commercial optical disk systems, media defects *are* minimized in conventional disks. Since the storage industry is concerned with

long term, high density memory with high integrity, media defects are minimized in commercial systems and sophisticated ECC is employed to reduce the system error rate even further. A typical level of raw media imperfections in a commercial system is 10^{-4} BER and can be compensated using a variety of error correcting codes to achieve a bit error rate of $< 10^{-12}$ at the system level.[95] For the parallel readout applications discussed here, it is difficult to conceive of *all parallel* ECC schemes to combat the relatively high raw media defect rate quoted above. The use of holographic encoding has been suggested and is discussed elsewhere. For the correlation architectures discussed here, media defects may be treated as a noise source in the reference plane and as such are already compensated to some extent by virtue of the recall mechanism itself. It is well known that the correlation filter, or matched filter, is the optimal decoder for a channel that is corrupted by stationary noise. Given that we know the characteristics of the disk, specifically the media noise spectrum, then we may form the optimal reference filters to be recorded on the disk, thereby achieving the highest expected performance. This is true for the image plane correlators as well as for the Fourier plane correlators of the next two sections. The inherent readout mechanism employed when using an optical disk in an image correlation system therefore is optimal for the tolerance of media defects when such defects may be characterized as arising from a stationary noise source.

The previous discussion regarding the effects of disk coating imperfections and media defects has allowed us to reduce the number of nonnegligible disk imperfections to just two ; disk wobble and disk center offset. These two imperfections play distinctly different roles in each style of correlator, therefore we will



(a)

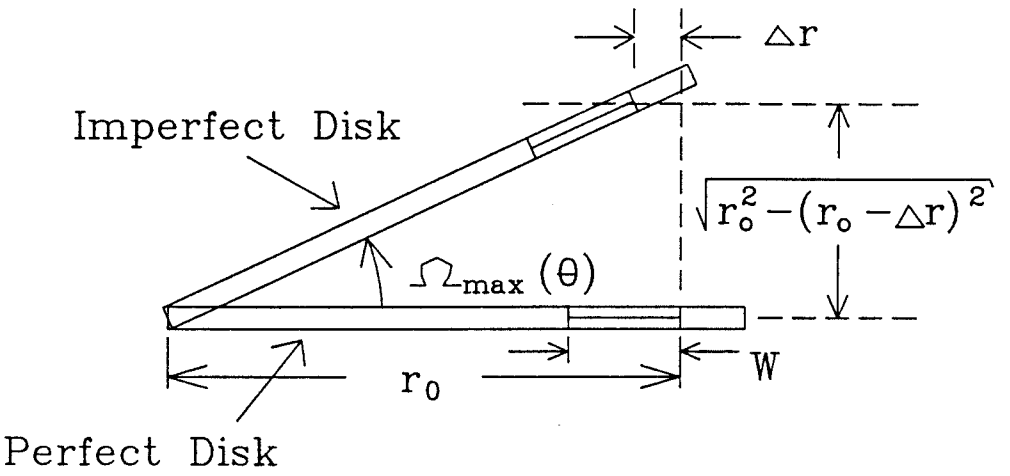


(b)

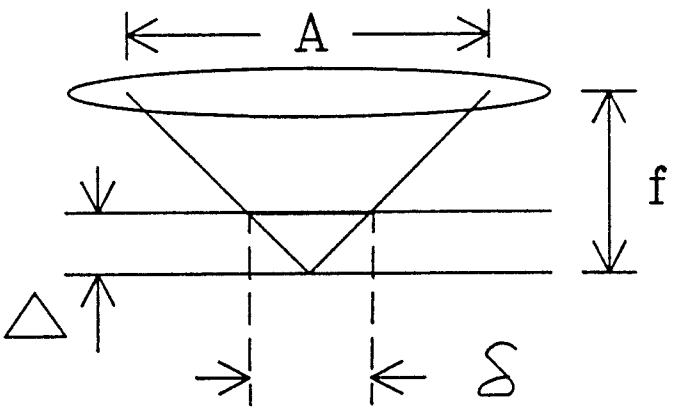
Figure 4.12 : Schematic depicting disk imperfections.

(a) Disk wobble.

(b) Disk center offset.



(a)



(b)

Figure 4.13 : Schematic depicting the effects of disk wobble.

(a) Reference image shift.

(b) Reference image blurr.

discuss them here in the context of image plane correlators and again later when we present the Fourier type. Disk wobble and disk center offset are mechanical imperfections in the disk planarity and track centering respectively. Although the effects are fairly self explanatory, Figure 4.12 makes the terminology explicit. In Figure 4.12a we show disk wobble as resulting from a warping of the disk surface. It can also result from mechanical imperfections in the rotation mechanism; however, we will combine these effects and discuss them together. Wobble can be characterized by the maximum magnitude and maximum frequency of $\Omega(\theta)$ where $\Omega(\theta)$ is the disk surface inclination angle expressed as a function of disk rotation angle or equivalently timing clock count. For the SONY system, $\Omega(\theta)$ is characterized by a maximum angle of 1° and a maximum frequency of 50 cycles/rotation. There are two potential sources of image plane correlation error arising from wobble. These will be referred to as reference image *shift* and reference image *blurr* and are depicted in Figure 4.13. Figure 4.13a shows the reference image shift induced by disk wobble. This effect will not play an important role in correlation error as the correlation function is shift invariant. If the reference image shift is too large however, the radial scanner may not scan the entire reference image. This is easily compensated by insuring that the radial scanner have a radial scan overhead of

$$\Delta r = r_0(1 - cos(\Omega_{max}(theta))). \tag{4.15}$$

where $\Omega_{max}(\theta)$ is the maximum of $\Omega(\theta)$. That is, instead of scanning a distance equal to the radial image extent W, the radial scanner will now be required to scan a distance of $W + \Delta r$. Figure 4.13b shows the maximum reference image

blurr arising from disk wobble. This effect is similar to the reference image blurr that results from reference image tilt in the rotating mirror correlator and may be quantified the same way. For a given $\Omega_{max}(\theta)$, we calculate the maximum image plane offset Δ as before

$$\Delta = r_0 tan(\Omega_{max}(\theta)), \tag{4.16}$$

and the resulting blurr spot size as

$$\delta = A\Delta/f, \tag{4.17}$$

which defines the maximum available disk resolution as limited by disk wobble. For the disk system used here we can calculate the maximum radial scan overhead as $\Delta r = 9.1\mu m$ or 10% of the 100 pixel total scan distance, and the effective minimum spot size as $0.6mm$ as limited by disk wobble. This result states that for a commercially available disk system, the tolerance on disk wobble is such that an uncompensated image plane correlator must sacrifice a factor of 10^5 in available disk SBP in order to insure that accurate correlation signals are produced! We will discuss a compensation mechanism that will allow us to retrieve this SBP in the next section. The other source of correlation error is disk center offset. Disk center offset is a manifestation of the fact that the rotational center of the optical disk differs from the common track center. Since it is the tracks on which we have written reference image data, we would like the rotational center to coincide with the track center; however, in the disk system we employed, the disk center offset can be as large as $\Delta r_c = 50\mu m$. As in the case of disk wobble induced reference image shift, this means an increase in radial scan overhead of an additional $50\mu m$.

Note that increasing the radial scanning distance does not actually *correct* the error induced by reference image shift. The resulting correlation signal will differ slightly from the true 2-D correlation pattern; however, the *peak* of the 2-D correlation signal will be unchanged. Since the pattern recognition applications discussed here will make use of the correlation peak exclusively, the details of the correlation signal distortion may be neglected.

We have seen that the primary imperfection induced limitation of the image plane correlator is reference image blurr. The question arises whether disk wobble may be compensated, thereby improving reference image resolution. A potential compensation scheme is shown in Figure 4.14. The proposed system is very simple and makes use of feedback from a probe beam to monitor the instantaneous disk tilt. The feedback signal is applied to a piezoelectric mirror which will tilt the input image so as to compensate disk wobble. As long as there is no appreciable departure from planarity over the reference image itself, this scheme will insure that wobble introduces no error. Since the maximum frequency of $\Omega(\theta)$ is limited to 50 cycles/rotation or $\approx 1kHz$ for a $20Hz$ disk rotation rate, the time response of the feedback circuit of Figure 4.14 is not an issue.

4.4 VANDER LUGT CGH CORRELATOR

The next style of correlator we will discuss is based on Equation 4.2. These correlators are referred to as Fourier-based correlators. This type of system is well suited to an optical implementation owing to the ease with which 2-D Fourier transforms may be computed using a simple lens. The basic idea of the Fourier-based correlator is depicted in Figure 4.15 where we have shown both the

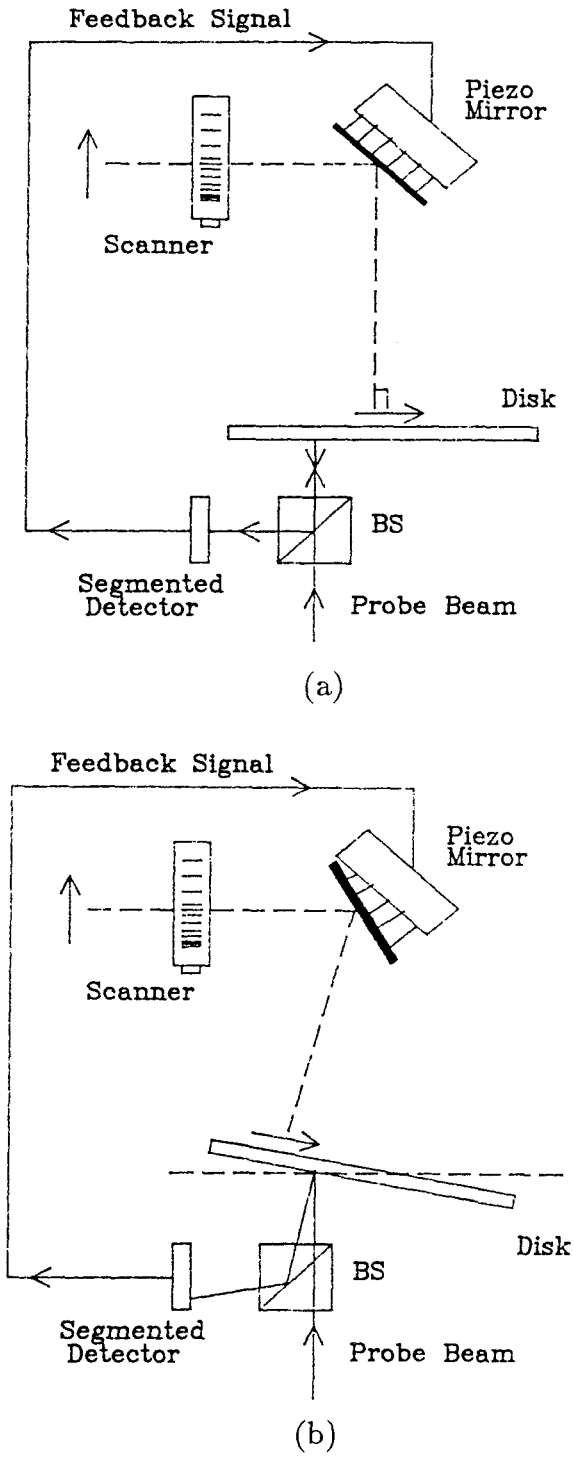


Figure 4.14 : Candidate wobble compensation scheme for image plane correlators.

(a) No wobble yields equal outputs from segmented detector and no piezo mirror tilt.

(b) Wobble is detected and piezo mirror corrects reference image tilt.

recording and the readout processes for a simple Vander Lugt correlator. Figure 4.15a shows the recording process in which the interference pattern between the object beam $E_o(x,y)$ and a plane wave reference beam $E_r = Ae^{ikz}e^{i\pi u_0 x}$ is formed on a holographic plate. As usual, we have assumed harmonic time dependence of the electric field and will not be including this dependence explicitly. The transmittance of the holographic plate after developing is given by

$$t(x,y) = \|E_0(x,y) + E_r(x,y)\|^2. \tag{4.18}$$

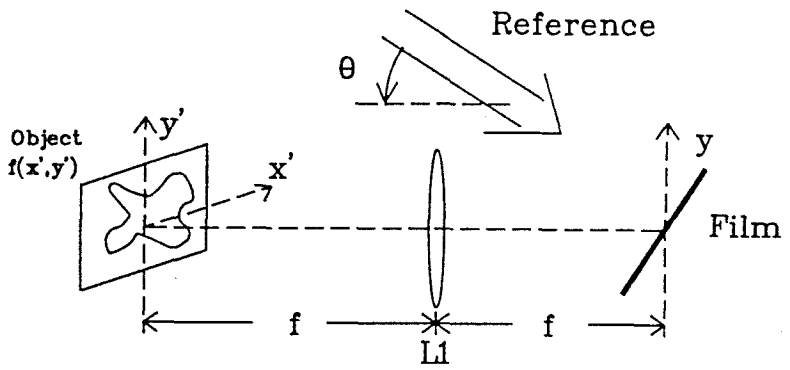
Since E_0 is obtained in the back focal plane of the Fourier transform lens L1, the field E_0 is proportional to the Fourier transform of the input object $f(x',y')$. That is

$$E_0(x,y) = BF(u,v), \tag{4.19}$$

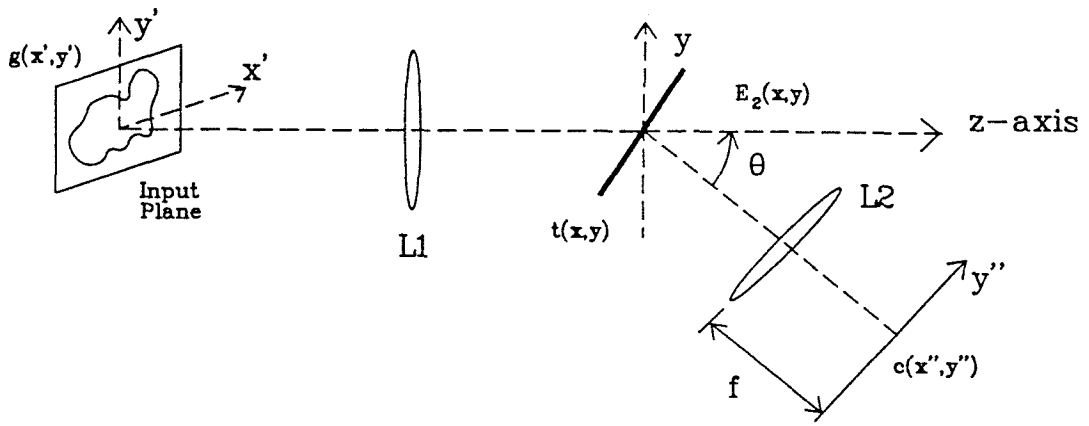
where u and v are spatial frequency coordinates given by $u = x/\lambda f$ and $v = y/\lambda f$, f is the lens focal length and B is an arbitrary constant. We have that the developed plate transmittance is given by

$$t(x,y) = A^2 + B^2\|F(u,v)\|^2 + ABF(u,v)e^{-i2\pi u_0 x} + ABF^*(u,v)e^{+i2\pi u_0 x}. \tag{4.20}$$

Figure 4.15b shows the reconstruction process in which the developed plate is placed in the back focal plane of lens L1 and illuminated using a new object $g(x',y')$ as shown. Lens L2 is used to compute the Fourier transform of the diffracted light, thereby obtaining the desired signal in the (x'',y'') output plane. The new object beam is proportional to the Fourier transform of the new object



(a)



(b)

Figure 4.15 : Vander Lugt correlator.

(a) Recording of Fourier plane filter.

(b) Readout or correlation process.

$g(x', y')$, so that the field immediately behind the plate is given by

$$E_2(x, y) = G(u, v)t(x, y), \tag{4.21}$$

which will consist of three terms. The DC term will propagate along the z-axis and the two diffracted orders will propagate at angles of $\pm\theta = sin^{-1}(\lambda u_0)$ as shown. We are interested in the diffracted field arising from the last term in $t(x, y)$. Specifically, in the back focal plane of the inverse transform lens L2 we have a term like

$$c(x'', y'') \propto \mathcal{F}\{F^*(u, v)G(u, v)\}, \tag{4.22}$$

which is just the 2-D correlation of the two images $f(x', y')$ and $g(x', y')$ as defined in Equation 4.2. So we have seen that an appropriately constructed Fourier plane transparency $t(x, y)$, may serve as a matched filter. We will now see two ways in which the optical disk may be used to implement this type of correlation architecture.

The first Fourier transform-based image correlator to be described is the simple Vander Lugt correlator shown in Figure 4.16. As can be seen from the figure, a Fourier transform computer generated hologram recorded on the optical disk is used as a Fourier plane filter for the input image. As we discussed above, the product of the transforms of the input and reference images is formed at the disk and an inverse transform yields the desired 2-D correlation in the output plane. As the disk rotates, a new correlation pattern is generated every time a different CGH aligns with the input image FT. Therefore, whenever there is a match between the input FT and the CGH, a peak occurs in the output plane

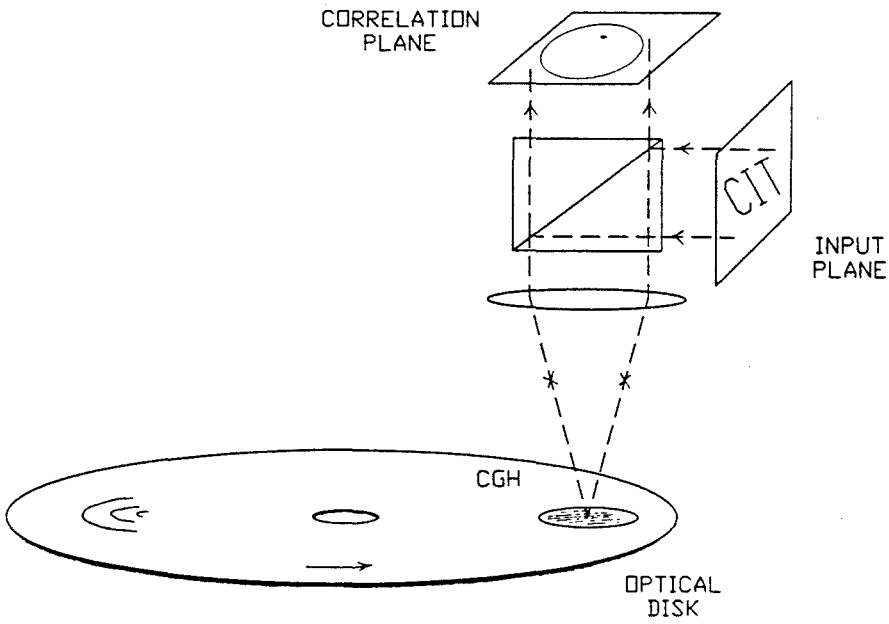


Figure 4.16 : Disk-based Vander Lugt correlator.

of the system. The location of this peak, which corresponds to the location of the object of interest in the input plane, may be anywhere in the correlation plane; therefore, a 2-D detector array is required to acquire the correlation data. Furthermore, since the correct correlation only occurs during the brief periods of alignment between the input FT and the CGH, the detector array need only be queried at these times. A pulsed laser or an electronically gated detector array could be used to achieve the appropriate sampling. The successful operation of this system depends on the optical quality of disk coating materials since these are capable of introducing random phase distortions across the system filter plane. As we discussed in Chapter 3, the optical quality of several commercially available disks is sufficient to make them suitable for these applications.

As with other FT-based architectures, one advantage of this correlator is

its potentially high speed. The correlation rate in this system is limited by disk rotation rate and detectability of the correlation peak. Taking a typical disk rotation rate of 40Hz, we calculate a correlation rate $\rho_c = 400{,}000$ correlations/sec for 100×100 pixel images. This correlation rate implies that in order to detect a correlation peak, each element of the 2-D detector array in the output plane must have a bandwidth of $\geq$ 40MHz. We can once again calculate the peak detectability n_p, given by the number of photons detected at the correlation peak, by first calculating the peak dwell time t_D as before, multiplying this by the expected power in the autocorrelation peak P_c, and dividing by the photon energy. That is

$$n_p = \lambda t_D P_c / hc, \tag{4.23}$$

where once again, λ is the wavelength of the optical field and h is Plank's constant. Peak dwell time is given simply by $t_D = 1/N_\theta \rho_c$ where N_θ is again the number of pixels in the reference image in the along track direction.

In the case of Fourier-based correlation, the autocorrelation peak power P_c, is a bit more difficult to calculate than in the previous image plane systems. In order to compute P_c we will first compute the expected fraction of the *diffracted* power which will arrive at the *peak* of the autocorrelation plane. We will assume that we have random images and will consider two cases. First we will treat the case of bipolar images. Let $a_{ij} \in \{\pm 1\}$ represent the value of the (i,j) pixel in the random image, and let a_{ij} be characterized by the simple statistics

$$p(a_{ij} = 1) = p(a_{ij} = -1) = 1/2 \quad \forall i,j \tag{4.24}$$

$$p(a_{ij} = k, a_{mn} = k') = p(a_{ij} = k)p(a_{mn} = k') \quad \forall i \neq m, j \neq n, \tag{4.25}$$

where $p(a_{ij} = k)$ is the probability that the (i, j) pixel takes on the value k. The second condition denotes statistical independence among pixels. The above statistics are equivalent to stating that the value of each pixel is determined by the outcome of a single toss of a fair coin. Given these image characteristics, we can write an expression for the value of the electric field at each point in the correlation plane of our Vander Lugt correlator as

$$c_{mn} = \sum_{i=m+1}^{N} \sum_{j=n+1}^{N} a_{ij} a_{i-m,j-n} \tag{4.26}$$

$$c_{00} = \sum_{i=1}^{N} \sum_{j=1}^{N} (a_{ij})^2, \tag{4.27}$$

where we have ignored unimportant proportionality constants and have just written c_{mn} as the value of the 2-D correlation pattern $c(x'', y'')$, at the (m, n) output pixel. For bipolar images, $c_{00} = N^2$, therefore the optical *power* diffracted into the autocorrelation peak goes like $c_{00}^2 = N^4$. In order to find the expected power *outside* the peak, we must first calculate the expected power measured at (m, n) which is

$$\begin{aligned} E\{c_{mn}^2\} =& E\{ \sum_{i=m+1}^{N} \sum_{i'=m+1}^{N} \sum_{j=n+1}^{N} \sum_{j'=n+1}^{N} a_{i,j} a_{i-m,j-n} a_{i',j'} a_{i'-m,j'-n} \}. \\ =& (N-m)(N-n) \end{aligned} \tag{4.28}$$

The *total* power diffracted outside the peak therefore is given by

$$P_{\Sigma} = 4 \sum_{m=1}^{N} \sum_{n=1}^{N} E\{c_{mn}^2\}, \tag{4.29}$$

where the factor of 4 is used to take into account all four quadrants of the correlation plane. Finally, we find that the *fraction* of *diffracted* power reaching

the autocorrelation peak is given by

$$\eta_P = c_{00}^2/(P_\Sigma + c_{00}^2) \tag{4.30}$$

$$= N^4/(2N^4 - 2N^3 + N^2), \tag{4.31}$$

which for large N reduces to

$$\lim_{N\to\infty} \eta_P = 1/2. \tag{4.32}$$

This means that one half of the diffracted power from the Vander Lugt correlator reaches the peak when the input and reference images are matched bipolar images. We can perform a similar analysis for the case of unipolar input and references. We now consider image statistics given by

$$p(a_{ij} = 0) = p(a_{ij} = 1) = 1/2 \quad \forall i, j \tag{4.33}$$

$$p(a_{ij} = k, a_{mn} = k') = p(a_{ij} = k)p(a_{mn} = k') \quad \forall i \neq m, j \neq n, \tag{4.34}$$

where once again we assume statistical independence among pixels. Without including the details of this computation, we find that in the limit of large N, we have

$$\lim_{N\to\infty} \eta_P = 1/N^2 \tag{4.35}$$

for unipolar images. Given the correlation peak efficiency η_P, the disk diffraction efficiency, η_D, and the source power, P_s, the correlation peak power is

$$P_c = \eta_P \eta_D P_s. \tag{4.36}$$

Substituting P_s=10mW, $N = 100$ and η_D=0.1% we find an expected peak power P_c= 5μW for bipolar images and $P_c = 1nW$ for the unipolar case. With a peak

dwell $t_d = 25ns$ we have a detectability $n_p = 4 \times 10^5$ photons for the bipolar case and a nearly undetectable $n_p = 165$ photons with unipolar images. This result reveals the fact that in order to realize the optical disk based Vander Lugt correlator described here operating at 400,000 correlations/sec, a phase-only input SLM and disk will be required. It has been shown that a polarization modulation scheme may be used to realize binary phase modulation through the use of an output polarizer oriented perpendicular to the mean transmitted polarization, making this constraint less restrictive. Of course, correlation rate can be traded off for detectability and may yield an acceptable n_p for binary images, given the actual ρ_c required.

Aside from the poor light efficiency shown for the case of unipolar images above, the three most significant drawbacks of this system are computational overhead, detector bandwidth, and alignment criticality. For each reference filter, a 2-D FT CGH must be computed and written on the disk. Although this computation will take place only once, for a large reference library the time required for this procedure can be very long. More importantly, the shift invariant nature of the correlation function that is so attractive from the perspective of pattern recognition, makes the correlation plane detector array requirement quite stringent. The output plane must comprise of a 2-D array of 200×200 detectors each with a bandwidth of 40MHz. Since serial interrogation of this output plane would require a 1600 GHz channel, we instead must have correlation plane processing that is incorporated into the detector array itself. This represents a major limitation of the proposed system since dedicated, high bandwidth VLSI processing is required in the output plane. Perhaps the most important limitation of

this system is the alignment criticality. The alignment of the input FT and the CGH is critical to within the resolution of the CGH ($\approx 1\mu m$). As discussed in the previous section, nonuniformities resulting from wobble and disk center offset lead to nonuniformities in reference CGH location with respect to the optical system. These nonuniformities must be compensated for the output correlation to be accurate. Wobble and disk center offset affect the disk-based Vander Lugt correlator in the same way as they did the image plane correlators discussed earlier. Wobble results in reference image shift and blurr while disk center offset will result in additional reference image shift. From $t(x,y)$ we can estimate the resolution tolerance of the disk-based CGH and can compare this result with the imperfection induced resolution limitations calculated before. The resolution required in the plane of the disk-based CGH is given by the larger of $\delta_D = 1\mu m$ the optical disk pixel size, and δ_{FT} the resolution of the transparency $t(x,y)$. We recall that

$$t(x,y) = A(x,y) + B(x,y)cos(2\pi u_0 x), \tag{4.37}$$

and we know that the bandwidth of $A(x,y)$ is equal to twice the bandwidth of $B(x,y)$. Now in order to avoid aliasing we require that the bandwidth of $A(x,y)$ be less than half the carrier frequency u_0 so that the highest frequency present in $t(x,y)$ is $u_{max} = 3u_0/2$. This implies that the resolution required is

$$\begin{aligned} \delta_{FT} &= 2/3u_0 \\ &= 2\lambda/3sin(\theta) \end{aligned} \tag{4.38}$$

For $\theta = 30°$, this yields a CGH resolution of $\delta_{FT} = 0.8\mu m$ which is very close to the disk resolution and far below the best uncompensated resolution achievable

in the presence of wobble as derived in the previous section. Fortunately however, the compensation scheme introduced earlier will be sufficient to compensate the Vander Lugt correlator since the source of the resolution error is identical to that of the image plane architectures.

The above system requires storage of the reference images in the form of computer generated Fourier transform holograms. In addition to the computational overhead incurred, another disadvantage is an increase by a factor of 100 or more in the space bandwidth product required to record the hologram as compared with the space bandwidth product of the image itself. Also, this increase in the area required to record each FT CGH results in an increase in optical power requirements and more stringent phase uniformity specifications. For these reasons we would prefer that it be only necessary to record the reference images as binary patterns on the disk, in which case each pixel of the image can be directly recorded as a separate spot on the disk. Gray scale images can also be recorded using some form of area modulation as we discussed in Chapter 3. The next Fourier-based system we will present is of this type. In addition to eliminating the computational and storage overheads associated with the CGH of the previous system, the next architecture improves the light efficiency and reduces output detector array requirements without sacrificing correlation rate.

4.5 Photorefractive Correlator

Since wobble and offset problems introduce *slowly varying* nonuniformities (< 50 wobble cycles/rotation), the problem of alignment sensitivity in Fourier-based correlators can be effectively dealt with using real time compensation with

feedback of the sort used in commercial disk drives; however, the computational overhead associated with generating the desired reference library in the previous system remains a problem. A schematic of the photorefractive (PR) correlator is shown in Figure 4.17. Since this architecture allows the reference *images* to be recorded on the disk instead of Fourier transform CGHs, the PR correlator significantly reduces both types of overhead associated with the Vander Lugt system. These savings can be measured in terms of both disk SBP usage for which the PR correlator can save a factor of several hundred to several thousand (e.g., these numbers are typical SBP overheads for high SNR CGH reconstructions) which arises from CGH modulation encoding, and precomputation time for which the PR correlator saves a factor of $2N^2log_2(N)$ in computation time for an N×N pixel image. In addition to these savings in overhead, we will see that the PR correlator is superior to the previous architecture in terms of light efficiency and output detector simplicity as well.

In the system shown in Figure 4.17, a photorefractive crystal is used to record a hologram of the input Fourier transform. In place of the PR crystal, any other realtime holographic storage medium may be used; however, the PR approach was chosen here to take advantage of the high diffraction efficiency obtainable from volume holograms while retaining the realtime nature of the recording process. During the recording phase, the disk illumination is blocked and the input transparency is illuminated from the right. For realtime applications, this input transparency may be an SLM interfaced to a video camera for example. A hologram is formed in the crystal between the input Fourier transform and the reference beam as shown. This is the same process as we depicted in Figure 4.15a

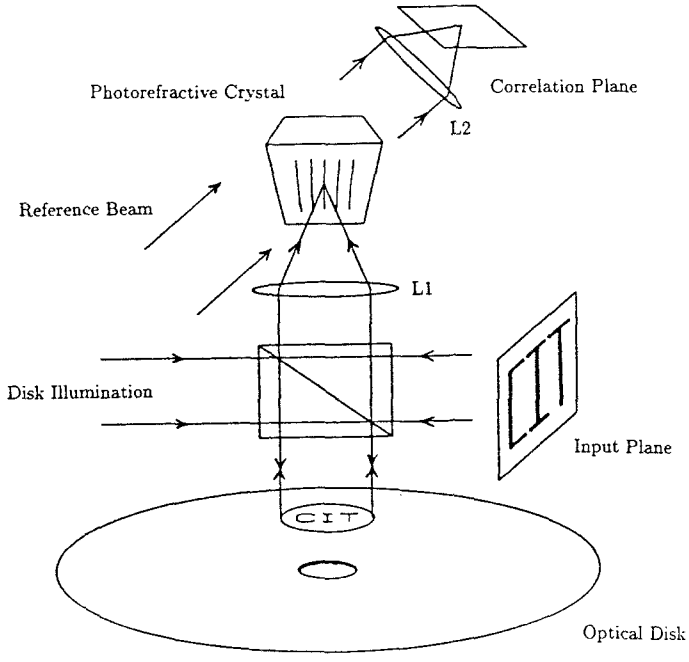
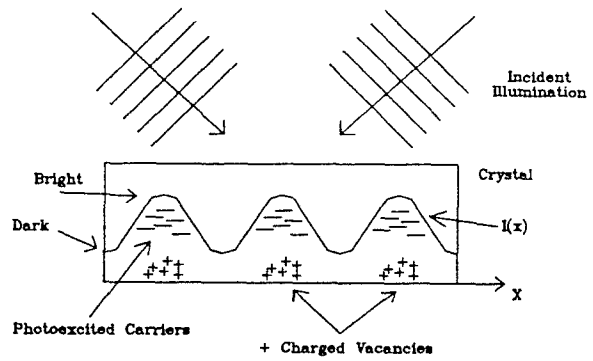


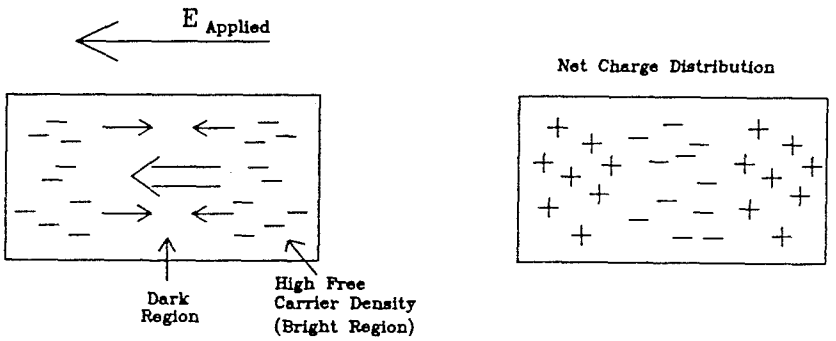
Figure 4.17 : Disk-based photorefractive correlator.

where we have simply replaced the holographic plate of that system with a PR crystal. The hologram thus recorded in the PR crystal will be read out using the reference library from the disk. On readout, the input is blocked and the disk is illuminated. The product of the input and reference Fourier transforms is formed in the crystal and the diffracted light is inverse transformed to obtain the desired correlation pattern in the output plane. Due to the volume nature of the signal stored in the PR crystal, the above explanation is somewhat incomplete. In order to better describe the operation of the PR correlator, we will briefly review the photorefractive effect and the formation of volume holograms in PR media.[110–112]

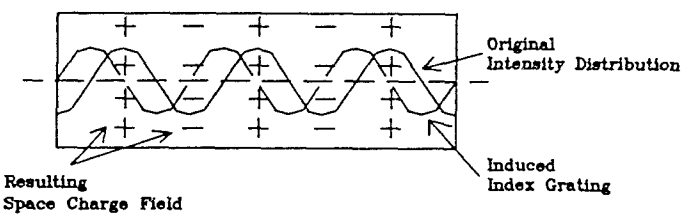
The PR effect can be understood in terms of the three steps depicted in Figure 4.18. First, an incident intensity distribution excites carriers to the con-



(a)



(b)



(c)

Figure 4.18 : Diagram depicting three stages of photorefractive recording process.

(a) Photogeneration of carriers (electrons).

(b) Redistribution of carriers and resulting charge distribution.

(c) Space charge field and induced index grating.

duction band of the PR media where they become free to move throughout the crystal. The energy of the incident photons must be sufficient to photoexcite these carriers from the crystal trapping sites. Spatially, the local density of these free carriers is determined by the incident intensity. This means that in bright regions there will be a high density of photogenerated carriers whereas in dark regions there will be few or none as shown in Figure 4.18a. In the figure, the carriers are assumed to be negatively charged. By virtue of the free carrier density gradient *or* under the influence of an applied DC electric field, the free carriers will diffuse/drift from bright to dark regions of the crystal and get trapped there, leaving behind positively charged vacancies. In this way, the incident intensity distribution is transformed into a volume charge distribution as shown in Figure 4.18b. As a result of this photoinduced charge distribution, a pattern of electric fields exists within the crystal. This space charge field induces a refractive index modulation through the linear electro-optic effect. In this way the incident intensity distribution is transformed into a volume index transparency or volume phase hologram. When the incident intensity is removed, the induced index modulation remains. This semi-permanent index pattern will persist in the dark until erased by the action of thermally induced carriers. The hologram may also be erased by further illumination. For example, if a crystal in which we have recorded such a hologram is uniformly illuminated, then the resulting photogenerated charge distribution will be uniform, thereby erasing the previously induced nonuniform distribution of charge.

The above explanation has shown how in PR media, an incident intensity distribution may give rise to a semi-permanent volume phase transparency which

mimics the original intensity pattern. In other words, PR crystals may be viewed as realtime volume holographic storage media with realtime recording/erasing capabilities and resolutions on the order of the average crystal trap separation ($\approx$ 100Å). In reality the recording and readout processes are more complicated than we have described above; however, this simplified view will be sufficient for our purposes. Returning to the PR correlator of Figure 4.17, each point in the input plane may be thought to give rise to a plane wave in the crystal which interferes with the reference beam and results in the recording of a single sinusoidal grating. What about interference between plane waves arising from pairs of points in the input plane? In the absence of an applied external field, gratings of such low spatial frequency are suppressed due to the long diffusion lengths involved. For this reason, we will assume that these intermodulation terms are not present in the recorded hologram. That is, after recording using the arrangement shown, the crystal transmittance can be approximated by that of a thin phase transparency with the low frequency terms suppressed. In order to take into account the crystal thickness L, we require that on readout the Bragg condition must be met. This condition will play an important role in the behavior of the PR correlator as we will discuss below.

If the photorefractive crystal is replaced by a thin medium such as a holographic plate, then the output pattern is exactly the desired 2-D correlation. When a thick hologram is used in the filter plane of such a system however, the resulting output is a 1-D slice of the 2-D correlation pattern.[113–115] This can be simply understood by considering the recording arrangement shown in Figure 4.19. On recording, each plane wave corresponding to each of the points in the

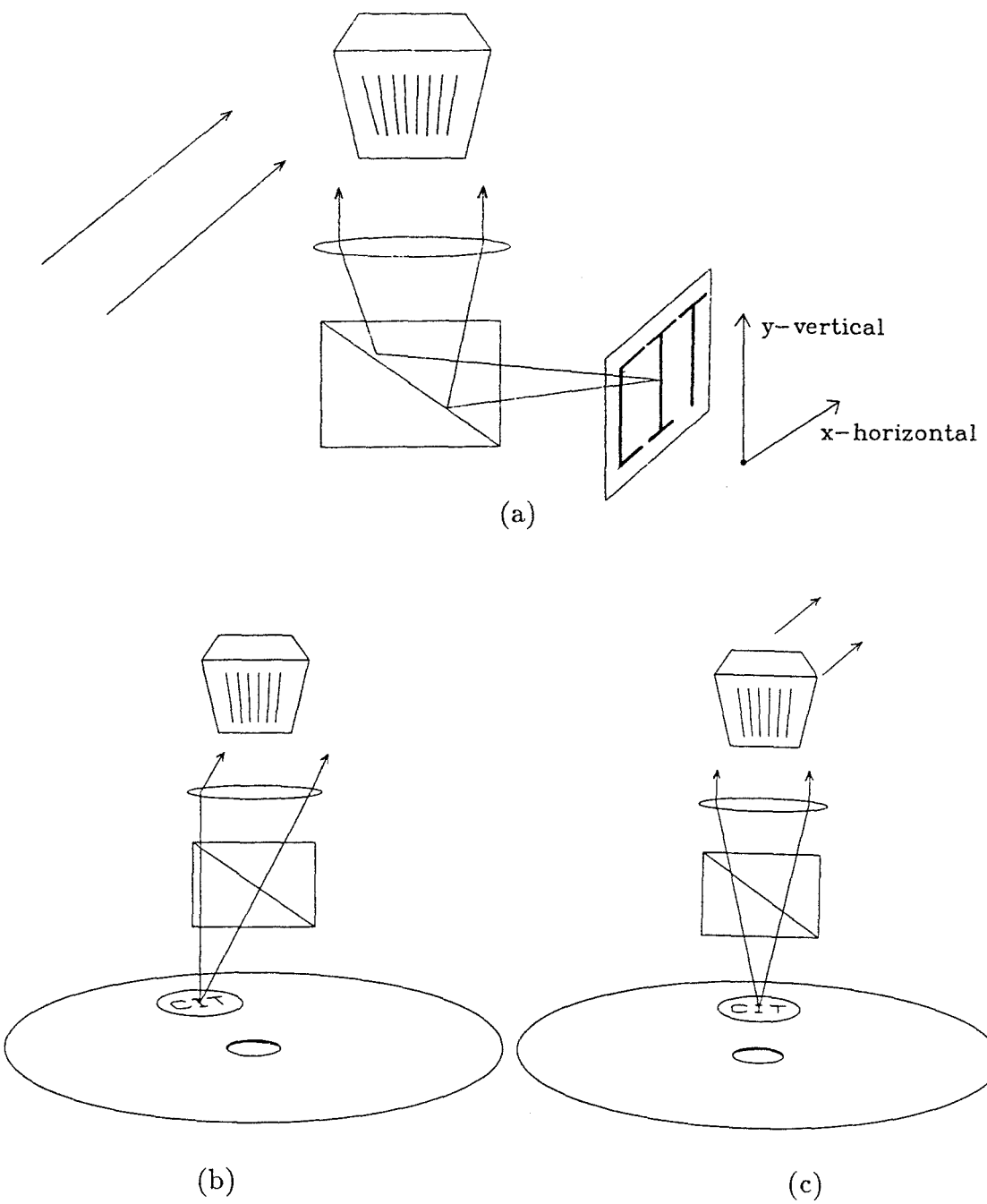


Figure 4.19 : Diagram depicting loss of horizontal shift invariance.

(a) Recording geometry.

(b) Shifted reference reading out grating recorded in (a).

(c) Centered reference reading out grating recorded in (a).

input image forms a grating with the reference beam. This process is depicted in Figure 4.19a. The resulting hologram exhibits Bragg selectivity in the horizontal direction. On readout, a point along a given radial line on the disk can only read out those gratings formed by points along one vertical line in the input as indicated in Figures 4.19b and 4.19c. Each such line on the disk reads out a corresponding array of holograms and generates a vertical array of spots in the correlation plane at the horizontal location corresponding to the reference beam FT. The coherent sum of all such reconstructions comprises the output of the correlator. This output pattern is the desired 2-D correlation multiplied, in the horizontal direction, by a *sinc* function whose width is inversely proportional to the hologram thickness. In the system of Figure 4.19, this property does not represent a limitation since all 1-D slices are obtained sequentially via disk rotation as shown in Figure 4.19c. Furthermore, instead of requiring a full 2-D detector array at the output, a 1-D array is sufficient to sequentially detect each slice of correlation output. This is an advantage over the Vander Lugt system in that only N detectors are required to operate at 40MHz as opposed to N^2, thereby reducing the serial readout channel bandwidth to a more reasonable 4 GHz. Since the crystal thickness results in Bragg selectivity in the vertical direction also, making L too large will limit shift invariance vertically as well. We can derive a condition such that both of these constraints may be satisfied.[116] Specifically, we would like to derive the PR crystal thickness required to achieve one line of correlation in the horizontal dimension while retaining N pixels of shift invariance in the vertical.

The condition for no shift invariance in the horizontal or x-direction is easily

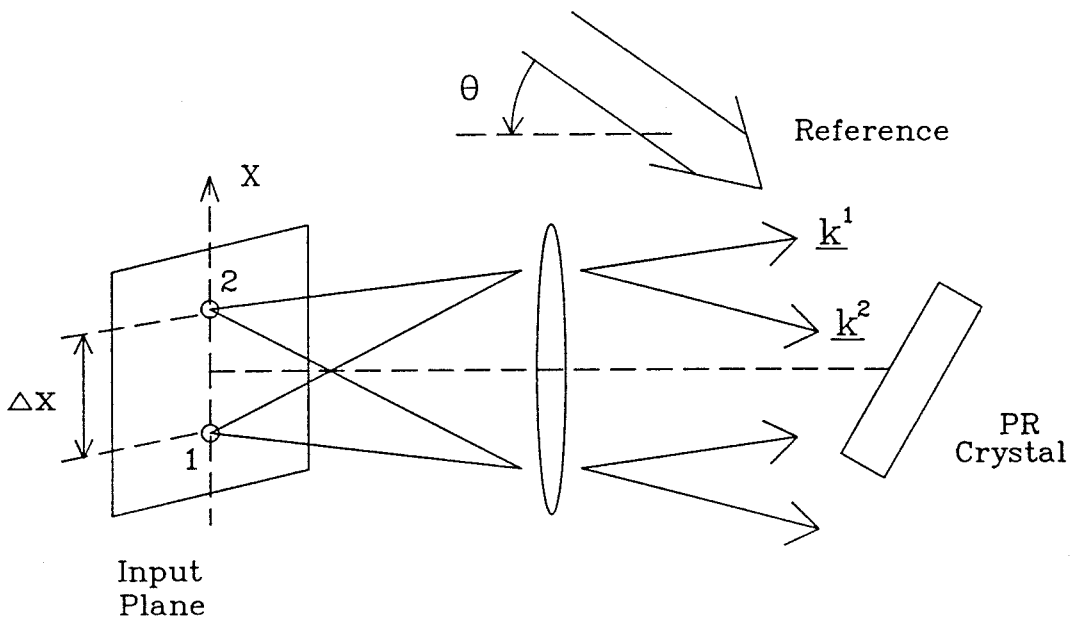


Figure 4.20 : Recording arrangement for two input points separated in x.

understood by considering Figure 4.20 where we have shown two input points located at y=0 and separated by Δx. This figure represents two adjacent points in the input plane of the PR correlator. In the Fourier plane, each input point gives rise to a plane wave which may be characterized by the x-component of its propagation vector. Specifically,

$$k_x^1 = \pi \Delta x / \lambda F \tag{4.39}$$

$$k_x^2 = -\pi \Delta x / \lambda F, \tag{4.40}$$

where F is the lens focal length and the superscripts indicate the input point with which the plane wave is associated. We have that in order for the plane wave arising from point 1 to readout the grating formed by the interference between point 2 and the reference, from the Bragg condition we require that $\Delta k_x < 2/L$ where L is the grating extent in the z-direction and $\Delta k_x = (k_x^1 -$

k_x^2). Conversely, in order to guarantee that points in adjacent columns do *not* readout each other's gratings, we require that $\Delta k_x > 2/L$. Substituting k_x^1 and k_x^2 from above we have that in order for exactly one line of correlation to appear in the output plane of the PR correlator, the crystal thickness must satisfy $L > \lambda F/\pi\Delta x$. This equation states that smaller input plane resolution requires larger crystal thickness to eliminate crosstalk. For reasonable system parameters such as $F = 10cm$, $\lambda = 633nm$, $\Delta x = 10\mu m$, we have $L > 2mm$. This is a reasonable requirement since in general, large L is desirable owing to the corresponding increase in diffraction efficiency resulting from the long interaction length. There is a maximum desirable crystal thickness however since for very long interaction length, shift invariance may be lost in the y-direction as well. In order to better understand the following discussion, consider Figure 4.21. The recording geometry is depicted in Figure 4.21a and resembles the geometry used above while the input plane now consists of three points along the y-axis at $y = 0$ and $y = \pm N\Delta y/2$. As before, each point gives rise to a plane wave that interferes with the reference beam. The four interfering fields of interest are given as

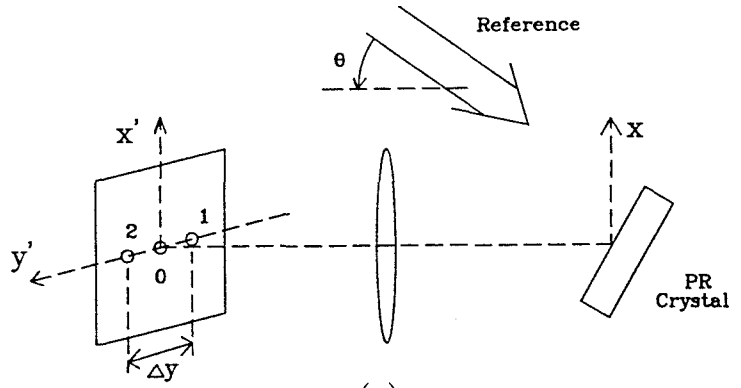
$$E_{REF} \propto e^{i(kxsin\theta + kzcos\theta)} \approx e^{ikxsin\theta} e^{ikz} \tag{4.41}$$

$$E_0 \propto e^{ikz} \tag{4.42}$$

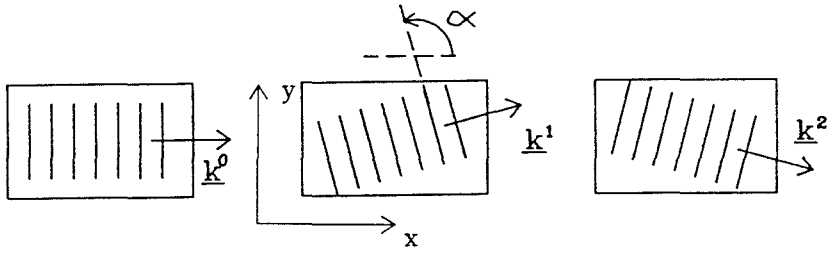
$$E_1 \propto e^{ikyN\Delta y/2F} e^{ikz} \tag{4.43}$$

$$E_2 \propto e^{-ikyN\Delta y/2F} e^{ikz}, \tag{4.44}$$

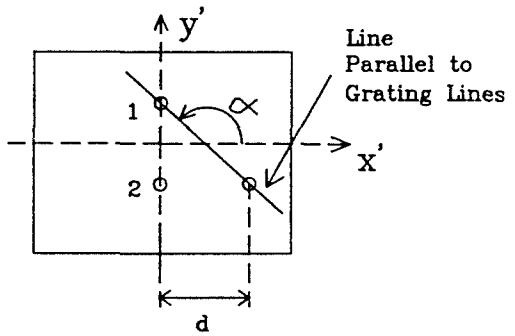
where Δy is the input plane resolution in the y-direction and N is the number of pixels in a single column of the input. The three gratings arising from these



(a)



(b)



(c)

Figure 4.21 : Loss of vertical shift invariance.

(a) Recording arrangement for three input points separated in y.

(b) Gratings resulting from recording shown in (a).

(c) Line in input plane along which point 1 is Bragg matched.

four fields are shown in Figure 4.21b. As an example, we consider the second of Figure 4.21b and write an expression for one of the grating lines or planes of constant phase. The grating itself is defined by

$$g(x,y) \propto cos(kNy\Delta y/2F + kxsin\theta + \phi). \tag{4.45}$$

Since the grating lines are simply lines of constant phase we can write an expression for *one* of these as

$$\frac{N\Delta y}{2F}y + xsin\theta = 0. \tag{4.46}$$

From this equation we can determine the angle α shown in the figure as

$$\alpha = tan^{-1}(2Fsin\theta/N\Delta y). \tag{4.47}$$

If we now go back to the input plane, it is clear that moving away from point 1 along the line shown will not result in Bragg mismatch since we are moving along a line parallel to the grating lines. Specifically, if we consider a point at the bottom of the image plane lying along this line, it *will* be able to read out the grating from point 1. We would like that point 2 be able to read this grating. This requires that $d < \Delta x_{MAX}$, where d is the distance shown and Δx_{MAX} is the maximum shift in the x-direction that can be tolerated as determined by the crystal thickness L. Δx_{MAX} can be determined from our previous result as

$$\Delta x_{MAX} = \lambda F/\pi L. \tag{4.48}$$

Putting these conditions together with our expression for α we have

$$d < \Delta x_{MAX} \tag{4.49}$$

$$d = N\Delta y / Tan\alpha, \tag{4.50}$$

so that

$$(N\Delta y)^2 / 2F sin\theta < \lambda F / \pi L, \tag{4.51}$$

which determines an upper bound on the crystal thickness as

$$L < 2\lambda F^2 sin\theta / \pi (N\Delta y)^2. \tag{4.52}$$

For the typical parameters we used before with $\Delta x = \Delta y$, $N = 100$, and $\theta \approx 60°$ we obtain $L < 3.5mm$. Once again we see that this thickness is in a reasonable range of available crystal thicknesses and is also compatible with the previous result. Summarizing these two results we have that in order for our PR correlator to generate a *single* line of correlation in the output plane while not sacrificing shift invariance in the orthogonal dimension, we require that

$$\lambda F / \pi \Delta x < L < 2\lambda F^2 sin\theta / \pi Y^2, \tag{4.53}$$

where Y is the vertical extent of the input image. It is interesting to see when such a crystal thickness exists. We find that

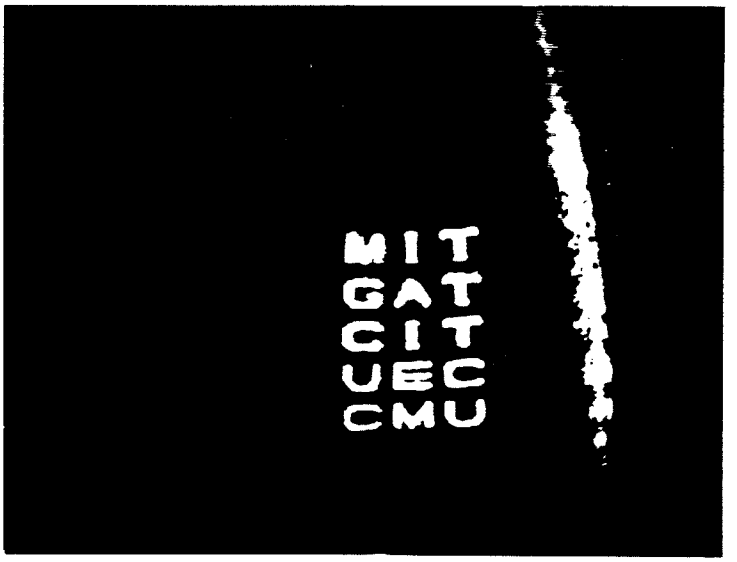
$$1/sin\theta < 2F\Delta x / Y^2 \tag{4.54}$$

$$1/sin\theta < 2F / SBP_{IN}\Delta x, \tag{4.55}$$

where in the second equation above, we have set $\Delta x = \Delta y$ and labeled $N^2 = SBP_{IN}$ as the space bandwidth product of the input plane equal to the total number of input pixels.

The PR correlator shown in Figure 4.17 has been experimentally demonstrated in two forms. First, we verified the operational principle of the PR correlator using a holographic plate in place of the PR crystal. This approach was necessitated by our desire to use the SONY WORM disk as an *input* SLM as well as the reference image library in the system. The low SONY disk diffraction efficiency made the required hologram writing time unrealistic for the PR media due to its relatively low sensitivity; therefore, we chose to use a Kodak type 131 high speed holographic plate as the holographic storage device. In order to compare this approach to that using the PR crystal, we may consider source illumination derived from a 5mW He-Ne laser. Given the plate sensitivity of $0.5\mu J/cm^2$ and assuming a factor of 10 loss due to collimating optics, we obtain a required exposure time of $t_E = 4s$, where we have also included losses arising from disk fill factor and diffraction efficiency. For the case of Strontium(0.6)Barium(0.4)Niobate (SBN), which in our case has a sensitivity of $\approx 200mJ/cm^2$, the required exposure time would increase to an unreasonable $1.6 \times 10^6 s$.[117,118]

The input to the correlator is shown in Figure 4.22a and was recorded in sector 1 of disk 1 where it measured 2mm×4mm. The photo in Figure 4.22a was taken from the disk using the diffractive retrieval scheme discussed in Chapter 2. The Kodak plate was exposed for $\approx 4s$ with a $0.125\mu W$ Fourier transform of the input image and was developed and bleached. Following bleaching, the plate was returned to the system and the resulting autocorrelation peak is shown in Figure 4.22b. This peak was obtained by illuminating sector 1 of disk 1 and allowing the Fourier transform thus obtained to read out the stored hologram. The somewhat broader spot in Figure 4.22b arises as a result of reflection of



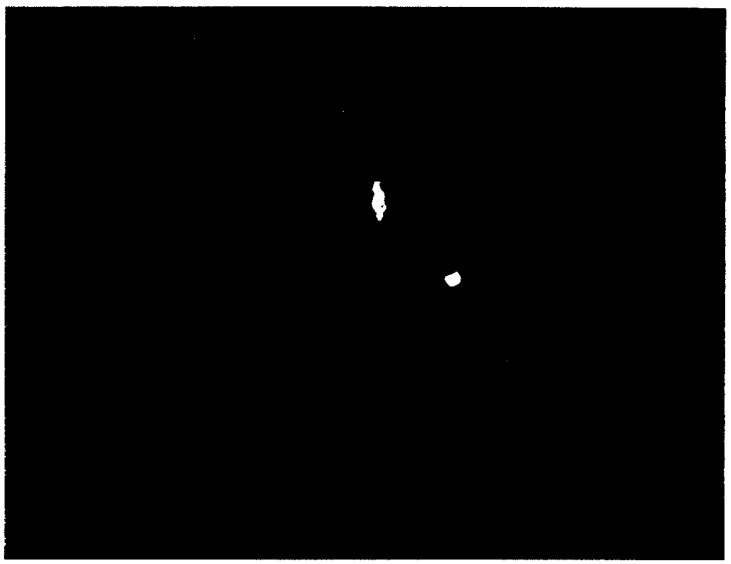
(a)

(b)

Figure 4.22 : Results of PR correlator using holographic plate.

(a) Reference image recorded on Sony disk.

(b) Correlation peak obtained using original image as reference.

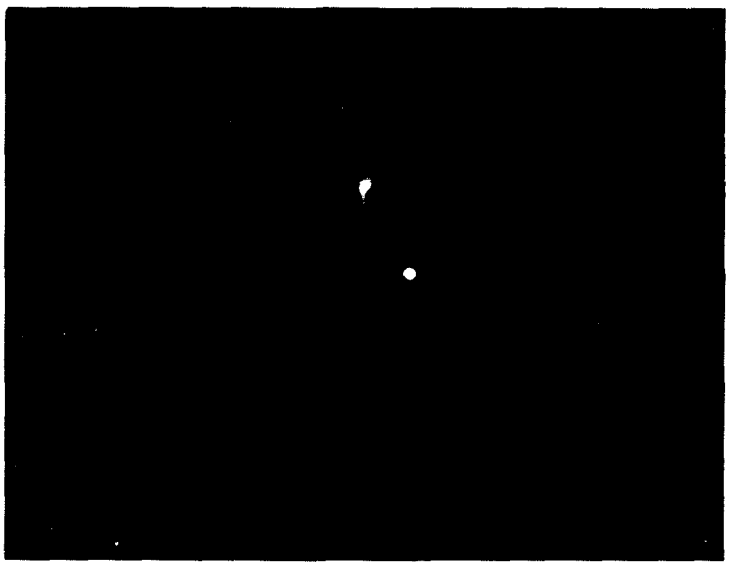


(c)

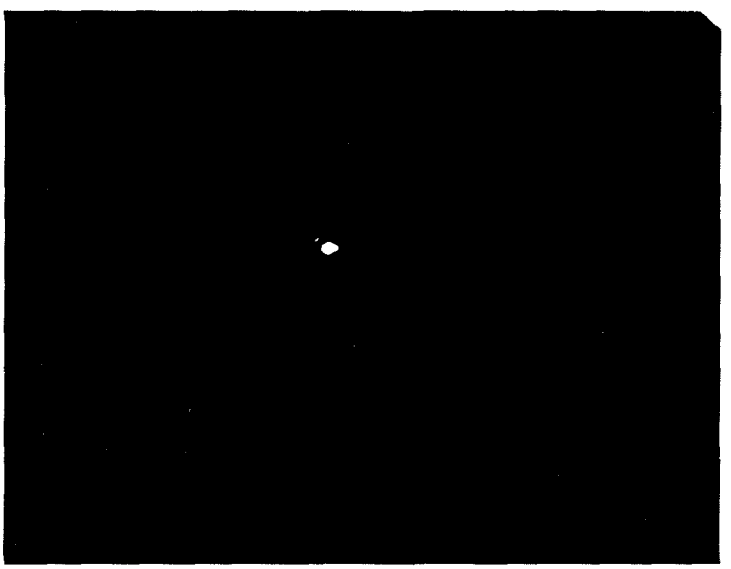
Figure 4.22(cont)

(c) Correlation peak obtained using shifted version of original image as reference.

the incident disk illumination from the air/disk interface and serves as a spatial reference position in the correlation plane. Figure 4.22c shows the autocorrelation peak obtained when the reference is shifted in the plane of the disk which results in a corresponding shift in the correlation peak. This system is obviously shift invariant in both x and y. The results of Figure 4.22 were obtained for an input image which is identical to the reference image. The image of Figure 4.22a was also recorded in sector 3 of disk 1 as well as in sector 1 of disk 2. Figures 4.23a and 4.23b show results obtained using these other reference images. The alignment spot is gone in Figure 4.23b owing to the difference between disk 1 and disk 2 surface tilts.



(a)



(b)

Figure 4.23 : Results of PR correlator using holographic plate.

(a) Reference image in different sector.

(b) Reference image on different disk.

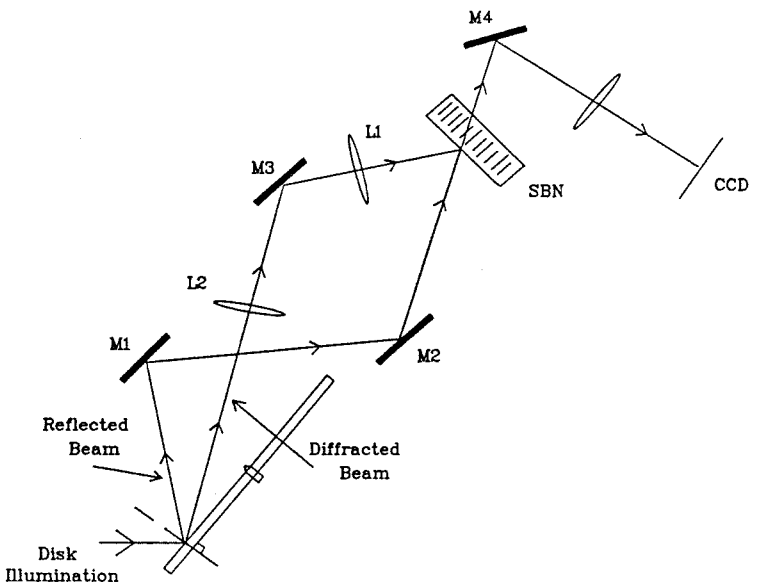
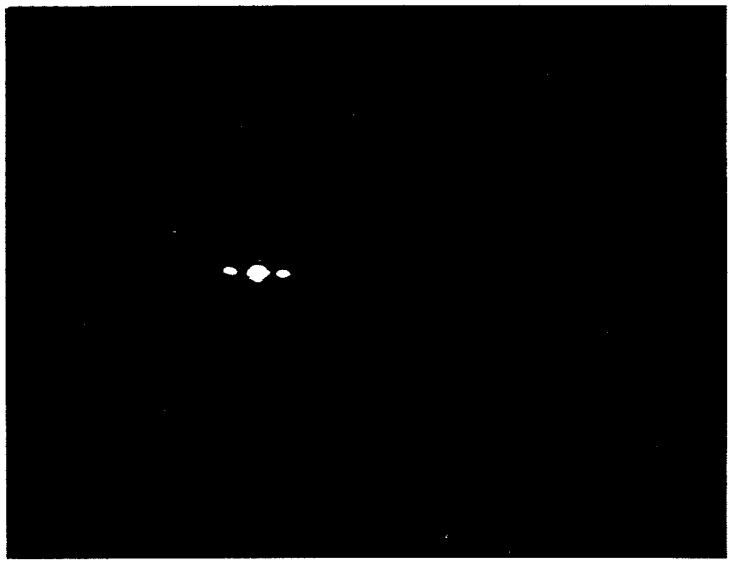
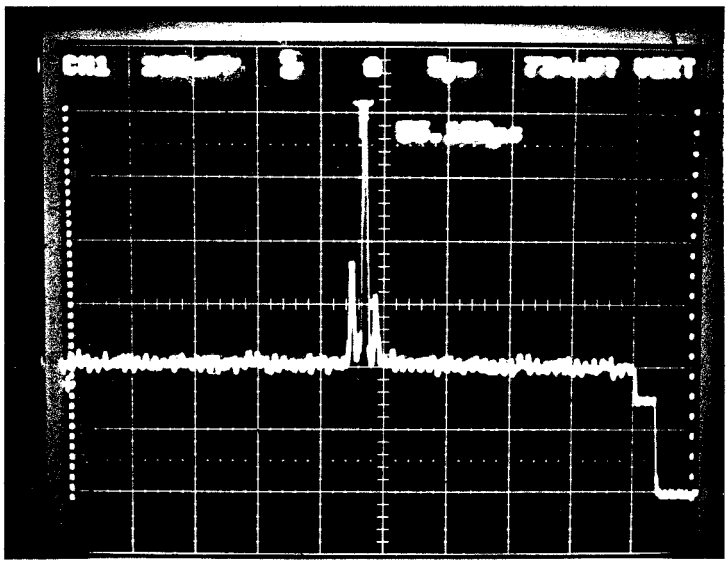


Figure 4.24 : Optical system used to realize PR correlator with SBN.

We were quite satisfied with the above results; however, the use of a holographic plate eliminates the possibility of a realtime implementation. In order to realize the PR correlator with PR media, we chose to use a 5W Ar^+ laser as a source, reduce the effective illumination area in the crystal by a factor of 100, and use the optical setup shown in Figure 4.24 to gain a factor of 2 through the use of the air/disk interface as a beamsplitter. We should note here that these power issues will not be a serious concern in practice since an input SLM with an efficiency greater than $\eta_D = 0.1\%$ will most likely be utilized. The power limitations we encountered here arose as a direct result of our desire to use the optical disk as an input SLM. Returning to the system shown in Figure 4.24, the modifications outlined above resulted in a saturation exposure time of $\approx 16s$. We assume negligible birefringence in the disk coating layer and orient both the reading and writing polarizations parallel with the crystal c-axis to take advan-



(a)



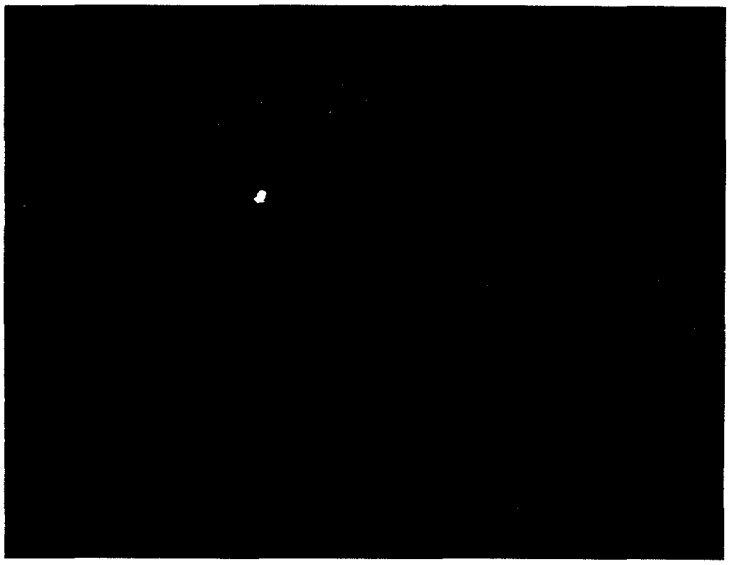
(b)

Figure 4.25 : Results of SBN PR correlator using CIT as input and reference images.

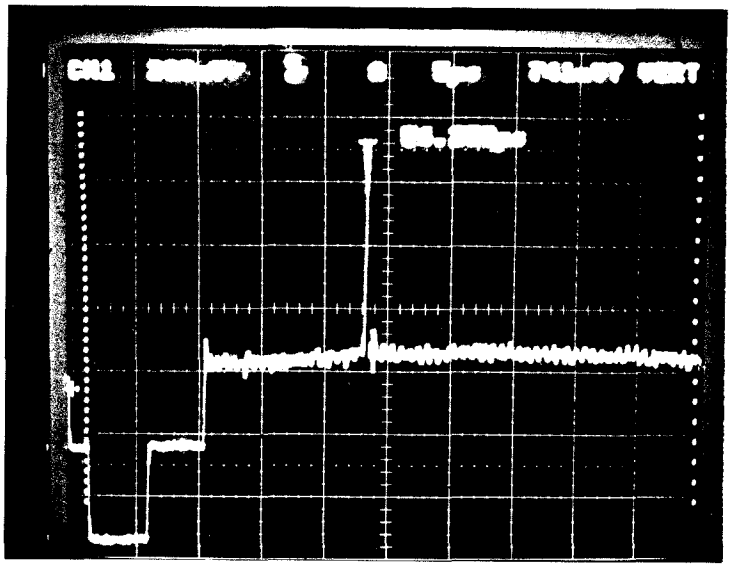
(a-b) Configuration 1 in which horizontal shift invariance is retained (c-axis vertical).

(a) 2D correlation output.

(b) Signal representing slice through (a).



(c)



(d)

Figure 4.25(cont)

(c-d) Configuration 2 in which vertical shift invariance is retained (c-axis horizontal).

(c) 2D correlation output.

(d) Signal representing slice through (c).

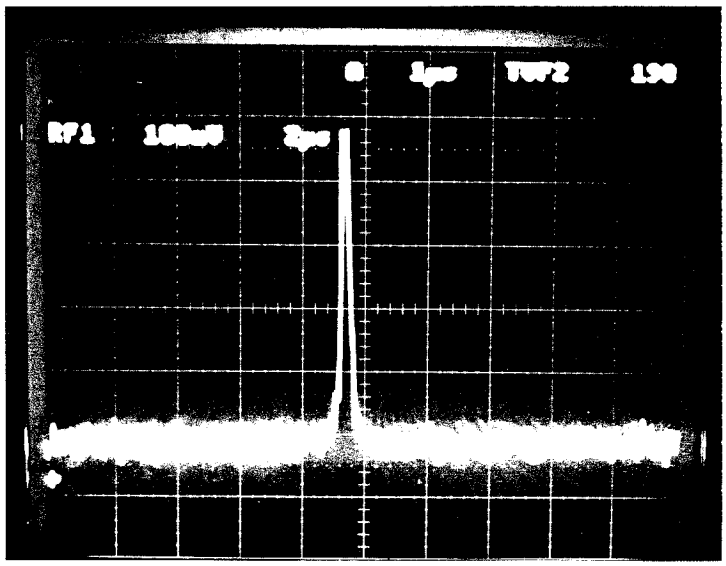
tage of the large r_{33} coefficient in SBN ($r_{33} = 1340 \times 10^{-12} m/V$). The correlation results using this setup are shown in Figure 4.25. Once again the acronym CIT was recorded on two different sectors of the same disk for use as input and reference images. There are two distinct configurations for the PR correlator; the configuration in which the c-axis and the Bragg selectivity are in the horizontal direction and the complementary configuration in which the system is shift invariant in the horizontal direction. In actual practice we would implement the correlator in the former configuration allowing disk rotation to recover horizontal shift invariance; however, for the purposes of demonstration, both configurations were tested. Results from these two configurations are shown in Figures 4.25a to 4.25d. From previous results we know that the autocorrelation of the acronym CIT has a five lobed structure in the x-direction and essentially no structure in the y-direction. In Figures 4.25a and 4.25b where we have shift invariance in x, we do indeed see this structure whereas in Figures 4.25c and 4.25d where we have Bragg selectivity in x, we see none. These results were obtained using identical recording and readout images (i.e., from the same sector). Referring to Figures 4.25a and 4.25b, we notice that the outermost two sidelobes are missing from the autocorrelation pattern. This is in agreement with our previous result which predicted a loss in shift invariance in the direction perpendicular to the Bragg direction owing to excessive crystal thickness. Using values from our experimental system we have $F = 12cm$, $\lambda = 488nm$, $L = 5mm$, and $\theta = 60°$ which yields an allowable image extent of $Y < \sqrt{2\lambda F^2 sin\theta / L} = 1.5mm$. Our actual image size was about $4mm$ so that we would expect to lose the sidelobes of the correlation pattern. Shown in Figures 4.26a and 4.26b are the correlation peaks obtained for

readout using the same sector image (4.26a) and different sector image (4.26b). Although phase nonuniformities due to disk coating materials, as well as hologram erasure due to readout, play minor roles in the signal reduction observed, the primary factor is disk alignment. As the disk is rotated to illuminate sector 2, disk wobble/tilt results in a slight shift of the Fourier transform and a corresponding reduction in correlation peak intensity. Shown in Figure 4.26c is the output obtained for the original input image after the disk has completed one full rotation. Again, the decrease in signal is primarily due to the sensitivity of the system to mechanical imperfections although it is likely that hologram erasure plays a role in the signal degradation of Figure 4.26c as well. A more detailed discussion of the sensitivity of the PR correlator to disk imperfections is included at the end of this section following the upcoming discussion of light efficiency in the PR system.

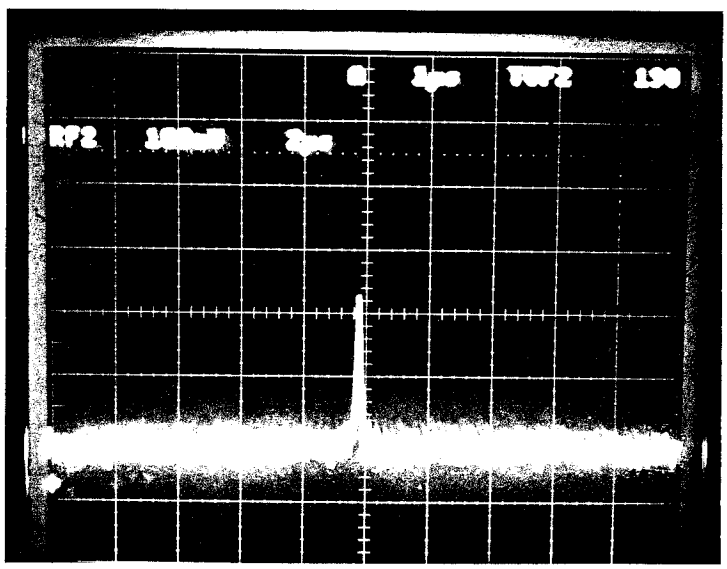
The expected correlation rate for this system is again limited primarily by disk speed and peak detectability. In contrast to the Vander Lugt system however, the PR correlator utilizes a high efficiency ($\eta > 50\%$) volume hologram and computes adjacent lines in the correlation pattern sequentially, thereby improving light efficiency drastically. In order to calculate the efficiency, we return to the image statistics given in the previous section. Once again we can express the autocorrelation peak power as

$$P_0 = c_0^2 \tag{4.56}$$

$$c_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}^2. \tag{4.57}$$



(a)

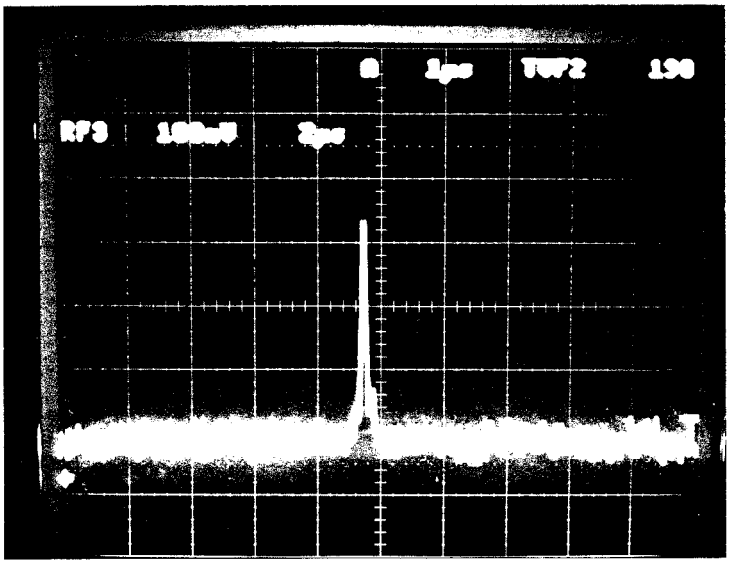


(b)

Figure 4.26 : Degradation of correlation peak from mechanical sensitivity.

(a) Original correlation peak signal strength.

(b) Peak signal obtained for readout using reference recorded in different sector.



(c)

Figure 4.26(cont)

(c) Peak signal obtained for readout using original input after 1 disk rotation.

In the present case however, the total diffracted power that appears away from the peak is modified as compared with the Vander Lugt system. Since only one line of correlation may be detected at any given time, we can write an expression for the field away from the peak as

$$c_k = \sum_{i=1}^{N} \sum_{j=k+1}^{N} a_{ij} a_{i,j-k}, \tag{4.58}$$

where a single index, k, indicates the position along the single line of correlation output. Once again we find the expected value of the power measured at any pixel off peak as

$$P_k = E\{c_k^2\}, \tag{4.59}$$

and the total diffracted power off peak has an expected value of

$$P_{\Sigma} = 2\sum_{k=1}^{N} E\{c_k^2\}. \tag{4.60}$$

Using bipolar image statistics we find that $E\{c_k^2\} = N(N-k)$ yielding $P_{\Sigma} = N^3 - N^2$ and $P_0 = N^4$ so that the fraction of diffracted power reaching the peak is given by

$$\eta = c_0^2/(c_0^2 + P_{\Sigma}) \tag{4.61}$$

$$= N^4/(N^4 + N^3 - N^2). \tag{4.62}$$

Once again we can obtain a useful estimate for the efficiency of the PR correlator by letting $N \to \infty$.

$$\lim_{N\to\infty} \eta = 1. \tag{4.63}$$

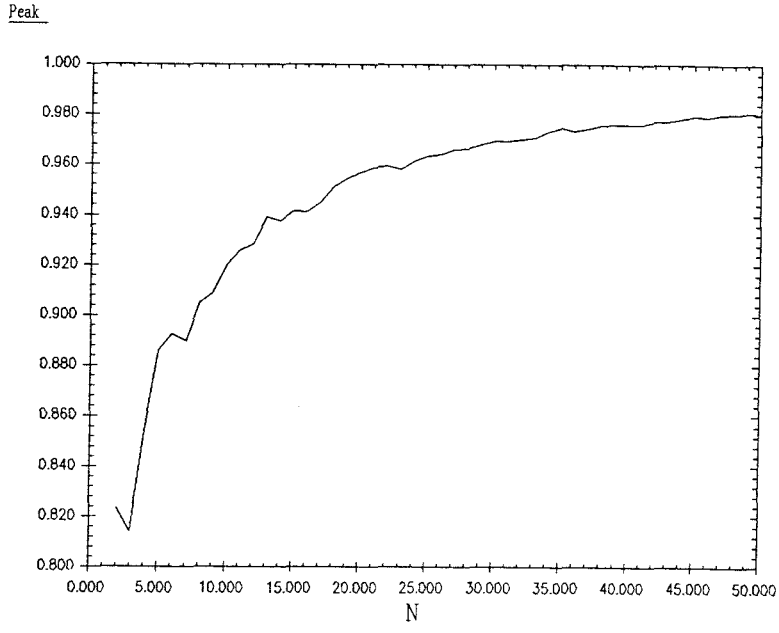
This result reveals that for large enough random bipolar input images, we can expect essentially all of the *diffracted* light to reach the correlation peak when there is a match between input and reference. In order to verify this result, computer simulations of the PR correlator were performed. The simulation procedure consisted of first choosing a random bipolar image of a particular size. The 2D autocorrelation of this image was computed and the resulting pattern was squared to represent correlation plane intensity. The ratio of the peak of this intensity pattern to the sum of all pixel values in the central column is the desired efficiency η. Using image sizes ranging from 2×2 to 50×50 and 100 statistics for each image size, the plot in Figure 4.27a was generated. As we see from the

figure for N=50, 98% of the light energy diffracted by the PR hologram reaches the peak. A similar analysis can be performed for unipolar images and we obtain

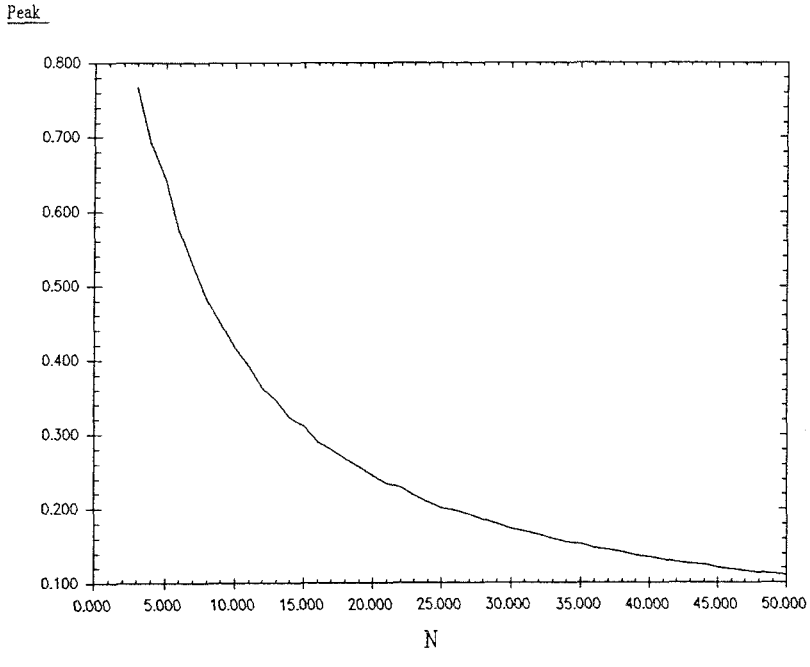
$$\lim_{N\to\infty} \eta = 6/N. \tag{4.64}$$

Notice that the efficiency of the unipolar PR correlator is higher by a factor of N than the unipolar Vander Lugt system. Again, this result has been verified by computer simulation and the result is shown in Figures 4.27b and 4.27c. Fitting a line to the asymptotic behavior of Figure 4.27c yields a slope of -0.91 and an intercept of 1.37 resulting in a best fit expression of $y = 4/N$ in good agreement with the analytic expression obtained above.

Given the efficiencies derived above, we can estimate the peak detectability in the PR correlator. With reference images of 100×100 pixels, $10mW$ readout illumination, and a disk efficiency of $\eta_D = 0.1\%$, we obtain a peak dwell of $t_D = 25ns$ for a 400,000 images/sec correlation rate. This corresponds to an autocorrelation peak power of $5\mu W$ for the bipolar system and $0.3\mu W$ for the unipolar case. In the previous calculation we have assumed a reasonable 50% photorefractive diffraction efficiency. Once again we may compute the number of photons in the autocorrelation peak for these two cases. We will use an illuminating wavelength $\lambda = 633nm$ in order to make a fair comparison with the Vander Lugt system. The result is that for the PR correlator operating on bipolar images, we have a peak detectability of $n_p = 4 \times 10^5$ photons. In the unipolar case for which the Vander Lugt system performed so poorly, we obtain a peak detectability of $n_p = 2.4 \times 10^4$ photons which once again, is an easily measurable quantity.



(a)

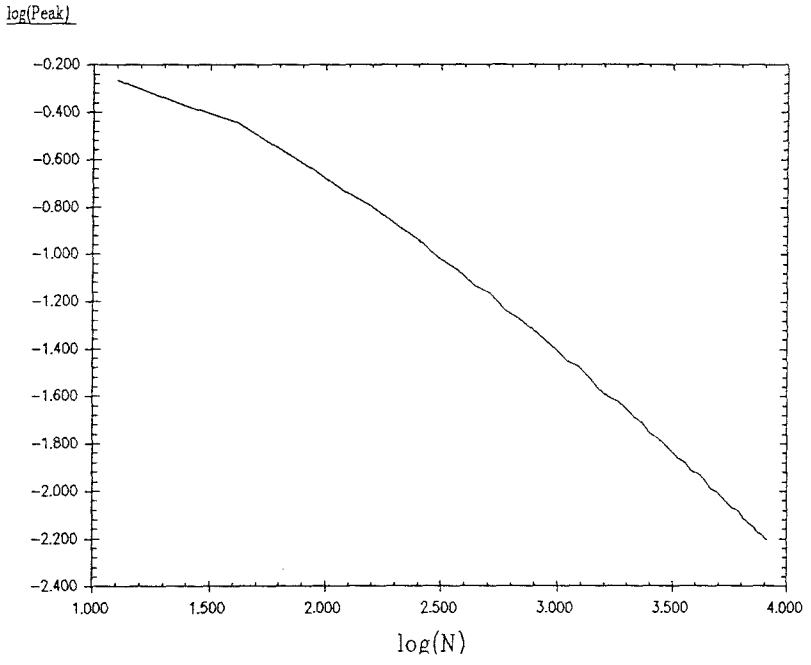


(b)

Figure 4.27 : Autocorrelation peak efficiency vs. image size.

(a) Random bipolar images.

(b) Random unipolar images.



(c)

Figure 4.27(cont)

(c) Log-log plot of (b) showing asymptotic linear behavior.

Despite the advantages gained in terms of computational overhead, detector simplicity, and light efficiency, alignment compensation remains a critical issue with this system. We will now go on to discuss alignment criticality and potential compensation schemes.

Once again we can ignore the effects of coating thickness variations and media defects for the reasons discussed earlier. In contrast to the Vander Lugt correlator, we can also ignore disk center offset and reference image shift due to disk wobble. This is true because we are not storing Fourier transforms on the disk. The critical alignment constraint imposed on the lateral positioning of the Fourier transform is removed when we store the reference images themselves owing to the shift invariance of the system. We do need to insure that the disk illumination is broad enough to tolerate these shifts however, which will result in a small loss in light efficiency. The most significant alignment error will come about by virtue of wobble induced disk plane tilt. Disk tilt will result in a shift of the reference Fourier transform. As an aside we should note here that this Fourier plane shift will not take place if a transmissive disk is used. In the case of a reflective disk-based system however, this shift *will* cause a misalignment between the readout illumination and the stored hologram. The misalignment is easily quantified as

$$\Delta x' = F sin\theta, \tag{4.65}$$

where θ measures the disk tilt angle ($< 1^\circ$ for SONY media) and x' measures the shift of the Fourier transform in the crystal plane. We may also calculate the Fourier plane alignment sensitivity as the effective resolution spot size in the

crystal. This is given by

$$\Delta x'_{MAX} = 1.22\lambda F/A. \tag{4.66}$$

We may determine a limit on allowable disk tilt by setting $\Delta x' < \Delta x'_{MAX}$ and we have that $sin\theta < 1.22\lambda/A$. So we see that once again, wobble tolerance can be traded off for disk SBP usage. Since wobble tolerance requires a small Fourier transform system aperture A, this in turn implies a low system cutoff frequency which results in a minimum disk plane resolution of $\Delta x > 1.22\lambda F/A$. For typical system parameters, we have $\theta = 1°$, yielding $A < 44\mu m$ which is smaller than a 100×100 pixel image on the disk. Another consequence of enlarging the spot size in the Fourier plane to tolerate disk tilt is the concomitant reduction in writing energy density and the resulting longer exposure times.

We show a candidate wobble compensation scheme in Figure 4.28. If we insist that $F/A \approx 1$ to insure good disk SBP usage, then we have that $\Delta x < 1\mu m$. The system of Figure 4.28 can achieve realtime alignment compensation in the Fourier plane by adjusting the disk illumination angle according to the instantaneous disk tilt as measured by the probe beam shown. The probe beam may be incident from the same side as the readout illumination as well; however, it is shown probing the opposite side of the disk for clarity. If the piezo mirror rotates by an amount equal to the disk tilt then the reflected light will always be aligned with the z-axis as shown. Piezo mirrors with the capability to provide up to 1° of angular displacement and with response times fast enough to respond to the probe signal (<200Hz) are readily available. The requirement for $1\mu m$ positioning accuracy in the Fourier plane and 1000 resolvable spots on the disk, translates to

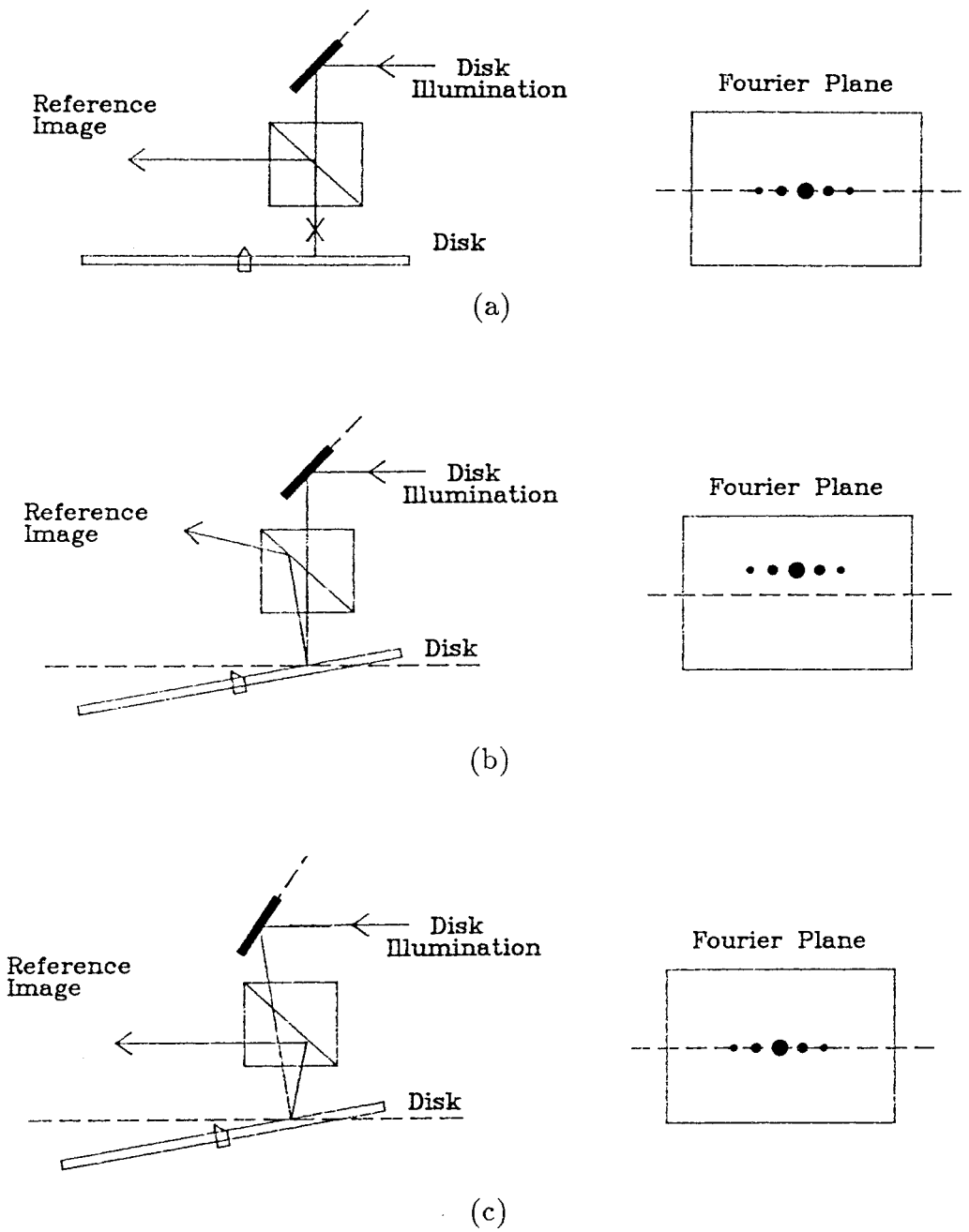


Figure 4.28 : Wobble compensation scheme for PR correlator.

(a) No wobble.

(b) Wobble without compensation.

(c) Wobble with compensation.

a piezo mirror angular accuracy of $\theta_{min} \approx 0.7mrad$. The probe beam therefore, is required to detect a disk tilt of $0.7mrad$ out of 1° which requires a detector sensitivity of at least 1 in 26. This capability is not *beyond* the reach of present technology but does represent an engineering challenge.

More than 5 billion bits can be stored in the type of disk that we use for most of our work. The number of 100×100 pixel images that can be stored in such a disk is more than 5,000, assuming a generous factor of 100 for loss of space bandwidth product due to representation (e.g., area modulation for gray scale representation). The rate at which all these images can be interrogated for a possible match with the input is limited by one or more of the following factors: the scanning speed of the disk (40Hz in our case), the speed of the radial scanning mechanism, and the sensitivity and the bandwidth of the output detectors and the electronics following them. We have seen that this rate can realistically approach 400,000 2D image correlations per second. It would be extremely difficult to duplicate this capability electronically and it can be achieved with *existing* optical technology. Moreover it is precisely such capability that is required for practical pattern recognition problems.

In this chapter we have seen how optical disk-based image correlators can yield very high correlation rates with good light efficiency. We have explored potential alignment compensation schemes which may be used in realtime implementations of these correlators to provide immunity from such disk imperfections as wobble and disk center offset as well as providing environmental immunity. We have successfully demonstrated several of these systems, verifying the accuracy of the image correlations obtained. In the next chapter, we will investigate the use

of optical disk-based correlators in pattern recognition architectures. The systems presented in this chapter will serve as building blocks in more sophisticated optical neural networks.

5. Optical Disk Based Pattern Recognition

In Chapter 2 we claimed that the correlation function is attractive for pattern recognition owing to its behavior in the presence of noise. Also in Chapter 2 we argued that in order to achieve some measure of success on any *interesting* pattern recognition problem, we would need more than noise tolerance in our recognition scheme, we would need distortion invariance. Distortion here can refer to common invariances such as shift, rotation or scale or it may refer to invariances that do not have a concise mathematical description such as author invariance in handwritten text recognition. The requirement for such invariances led to the notion of the correlation filter library which provides the desired distortion invariance by storing many correlation filters in an attempt to span the space of allowed distortions. Using this approach, a distorted input may be recognized by comparing it with many distorted versions of known objects, again taking the maximum correlation or correlations as an indication of class assignment. In Chapter 4 we demonstrated several optical disk-based correlation architectures which implement both reference library storage and the correlation-based search mechanism. In this chapter we turn to the description of optical systems for pattern recognition that benefit from the approach described above.

There are two sets of issues associated with the use of optical disk technology in the solution of a pattern recognition problem : architectural issues and algorithmic issues. In the realm of algorithmic issues, we are primarily concerned with filter generation; that is, determining what data to record on the optical disk in order that the correlation library approach may yield good recognition results. This problem is closely associated with the problem of learning in neu-

ral networks. In order to generate *good* correlation filters in the sense defined above, we are first given a set of labeled samples representing distorted versions of the objects to be recognized (i.e., the training set) and we would like to design a correlation library that classifies future samples correctly (i.e., we would like the system to generalize). In order to test the success of our filter generation scheme or learning algorithm, we allow the trained pattern recognition system to classify some known patterns that were not in the training set, and use the recognition results on this testing set as an indication of future system performance. A question that naturally arises is " how large a training set is required in order to achieve good generalization?". Unfortunately, this question has not yet been answered; however, a recent paper has described the conditions under which consistency can be expected.[52,60] That is, we can specify the size of the *testing* set that will insure with high probability that the behavior we observe during testing will continue for future samples. Although we cannot say exactly how many training samples are necessary to achieve good generalization, our desire during the training phase is to *minimize* this number through the incorporation of *a priori* knowledge about the problem we are trying to learn. It is our hope that *a priori* problem information can be used to constrain the system so that with relatively few training samples good generalization is achieved. Since the problem of filter generation in the correlation library approach is quite similar to the problem of learning and generalization in neural networks, we have developed a new learning algorithm for multilayer neural networks which yields good generalization for certain classes of problems. We will describe results pertaining to this algorithm in the following section. The filter generation problem is also quite

similar to various classical pattern recognition and interpolation algorithms such as k-nearest neighbors and radial basis functions. In the next three sections we will discuss these algorithms and their application to optical disk-based pattern recognition architectures.

After describing those algorithms to be used in the disk-based pattern recognition systems, we will turn our attention to architectural issues. There are three basic architectures we will describe : single layer shift *invariant*, single layer shift *variant*, and *multilayer* shift variant. In the last half of this chapter we will present experimental results from optical implementations of these architectures. These systems will be demonstrated in conjunction with the algorithms mentioned above as applied to the handwritten character recognition problem.

5.1 Filter Generation : The CLUSTER Algorithm

As we saw in Chapter 2, since a single correlator can be viewed as a shift invariant neuron, a correlation library may be viewed as the shift invariant first layer of a multilayer neural network. It is this analogy that allows us to discuss filter generation in correlation based pattern recognition schemes within the framework of learning and generalization in neural networks. The problem of learning in multilayer networks has received a great deal of attention recently owing to the relative success enjoyed by the Backward Error Propagation algorithm (BEP).[8] In BEP the network mapping is treated as a function with free parameters (i.e., the interconnection strengths or synaptic weights) and the learning process is interpreted as an error minimization task over these weights. In this scheme, the error is defined as the squared distance between the network

output in response to some input pattern and the desired output as defined by the correct classification of that pattern. Since the network output is a function of the weights, a gradient descent procedure can be used to search through the space of weights for a minimum of the error during training. Although BEP has been quite successful in providing a technique by which an arbitrary training set may be loaded onto a network, generalization from the resulting networks has been poor for most interesting problems. One reason for this poor generalization is the lack of incorporation of *a priori* problem information into either the network itself or the training algorithm.

When training neural networks, generally there are many interconnection solution vectors consistent with a given set of training samples. Under the assumption that the network architecture can indeed represent the problem from which the training set was drawn, we know that at least one of the possible solution vectors represents the actual *problem* solution. Since it is typically the performance of the network in the presence of new data in which we are primarily interested, it becomes important to examine techniques for constraining the weights to insure the emerging network will perform favorably on samples drawn from the same mapping as was the training set. Obviously, if we can use *a priori* information about the problem to sufficiently constrain the network so that only *one* solution is consistent with the training data, then the network will generalize the training set to the problem represented by that one solution. Such constraints represent prior knowledge of the problem space, and as a result we should expect specific constraints to be well suited to specific classes of problems.

In Figure 5.1 we depict the problem space, the space of all functions. A subset

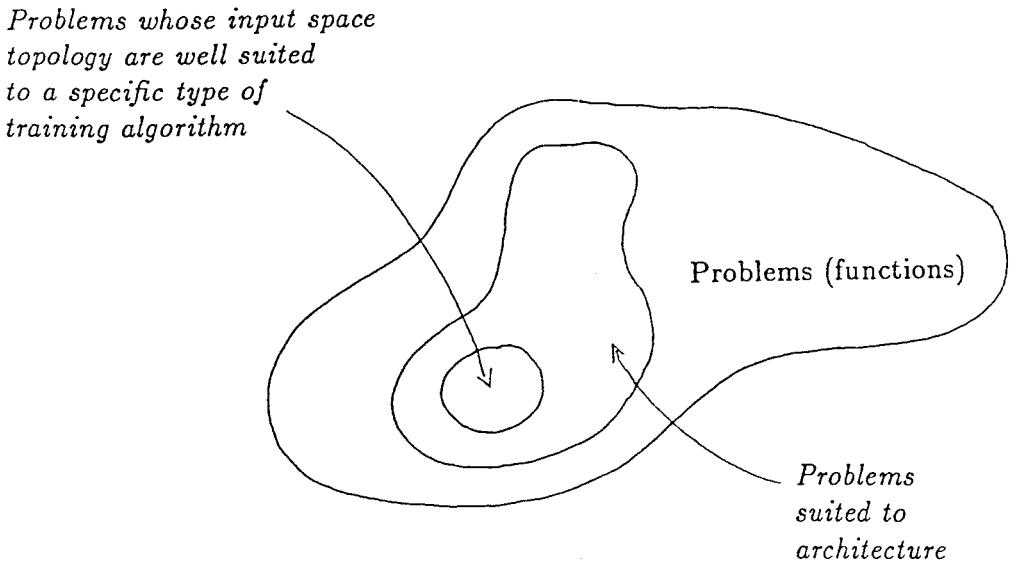


Figure 5.1 : Problem space.

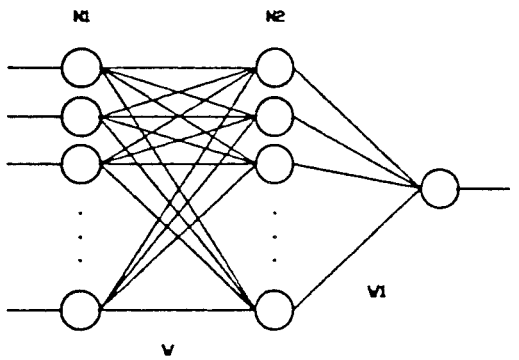


Figure 5.2 : Definition of a 2-layer feedforward neural network.

N1:N2:1

of the problem space is defined by the set of functions that may be realized using a specific architecture. In this section we will be primarily concerned with those functions that can be realized on a 2-layer network architecture (see Figure 5.2). Within the set of such functions, there is a further subset whose characteristics are matched to the aforementioned network constraints. Since the constraints have been chosen to limit the number of solution weight vectors to one, the function characterized by that one solution must lie in this last subset of all functions. A detailed understanding of this subset for a specific constraint then is essentially equivalent to an understanding of that set of problems for which our network / training algorithm pair will be likely to generalize.

In this section we will examine the generalization that emerges by imposing the constraint that the number of distinct internal representations allowed during training is minimized. The section is organized into four subsections. The first subsection provides the motivation for the above constraint from a topological point of view and also introduces the energy function used to effect the desired learning algorithm. The second subsection describes the performance of the algorithm on several problems. In the third subsection, we describe the specific characteristics of several problems with respect to their input space topology and the relationship between these topologies and successful generalization under our constraint. Finally, in the last subsection we discuss training multilayer networks in the context of image recognition.

5.1.1 Generating Intermediate Representations

Any network consists of three sets of elements or neurons; input, output, and hidden neurons. With respect to a specific input vector , we may regard the resulting pattern of activation over a given set of hidden units as the internal representation of the input vector over that set of neurons. In a layered feed-forward network, we generally regard the vector describing the activation in a particular layer of units as the representation of the input vector in that layer. The first layer representation of an input vector is defined by a linear transformation followed by a thresholding function as we have described in Chapter 2. The thresholding may be hard or soft. For simplicity we will assume here that it is a hard threshold function. In a 2-layer machine it is the separability of these first representations that determines the success of the overall dichotomy. The choice of this first mapping is crucial in order to correctly classify the training samples. Further, using the same reasoning it is clear that the efficiency of the first representations with respect to the *problem from which the training set was drawn* determines the performance of the machine in the presence of untrained samples. That is, the choice of first representations determines generalization in 2-layered networks.

The first layer representations label convex regions bounded by hyperplanes in the input space. Using this view of first representations we can say that it is the way in which our first layer partitions the input space into regions that determines generalization with respect to a particular problem. Specifically, in order that our network generalize from a given training set to the problem from which it was drawn, we require that the partition of the input space correspond

not only to the training set but to the problem as well.

The question now becomes, how do we choose a specific partition from the set of admissible partitions (i.e., the set of partitions that classify the training set correctly) so that it will favor some specific class of problem topologies? We propose here that a useful way to constrain the input space topology is to minimize the number of regions into which training points may fall. This is equivalent to minimizing the number of distinct first layer representations generated during training. Before discussing this constraint in detail we should make clear an assumption that was mentioned briefly earlier. The space of problems or functions we are considering are assumed to be *doable* by a 2-layer network. That is, not only can the machine classify the training set correctly, but we also require that there exist a set of weights so that the network's input/output relationship is nearly identical to that of the problem. Given this assumption, the *minimum number of representations* constraint initially assumes that the introduction of each new point in the training set does not require an increase in the number of regions. The algorithm must search for a new configuration of input space hyperplanes consistent with the new set of data, which does not increase the number of labeled regions. If this search is unsuccessful then an additional region must be allowed.

There are many schemes by which one could minimize the number of regions. One of the methods we have investigated is a gradient descent algorithm that minimizes an energy function. The criterion function we minimize is

$$
\begin{aligned}
\mathrm{E} = & \frac{2.}{M_1(M_1-1)} \sum_{p=1}^{M_1} \sum_{q=p}^{M_1} |\underline{x1}^p - \underline{x1}^q|^2 \\
& + \frac{2.}{M_2(M_2-1)} \sum_{p=1}^{M_2} \sum_{q=p}^{M_2} |\underline{y1}^p - \underline{y1}^q|^2 \\
& + \frac{1.}{M_1 M_2} \sum_{p=1}^{M_1} \sum_{q=1}^{M_2} \frac{1.}{|\underline{x1}^p - \underline{y1}^q|^2}
\end{aligned}
\tag{5.1}
$$

where $\underline{x1}^p$ and $\underline{y1}^q$ are the first layer representations of the pth and qth elements of class1 and class2 respectively, and M_1 and M_2 are the number of vectors in classes 1 and 2.

We will refer to this procedure as the CLUSTER learning algorithm since it tends to draw same class representations close together while pushing different class representations apart. Specifically, the first two terms in the above energy function result in minimum interclass scatter for transformed x's and y's , while the third term maximizes the between class scatter of the first representations. This energy function is reminiscent of criterion functions investigated in the 60's with regard to linear pattern classifiers.[119] The primary difference between the present approach and those early investigations lies in the fact that the transformation that defines the first layer representations is nonlinear. The nonlinear (threshold) function defines a set of disjoint regions in the input space, the number of which is minimized by the proper choice of first layer weights. The minimum number of regions is *a priori* unknown.

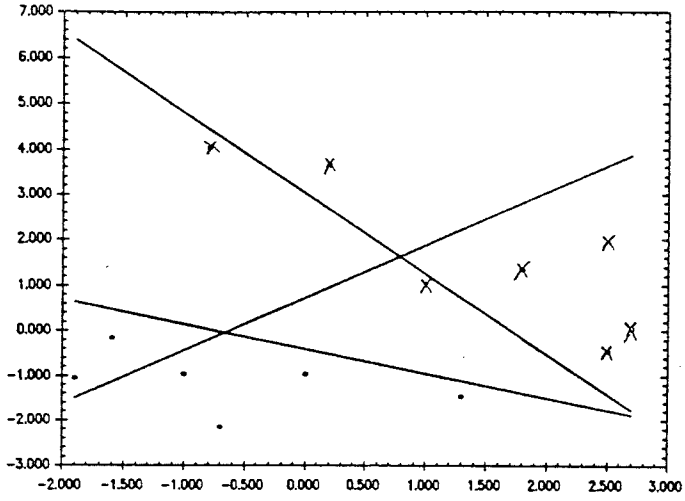
Why is this constraint a good one? It will obviously only be useful when we are considering a certain class of problems for which it is suited. The first observation we can make with respect to this class is that the problems of interest

should be defined by relatively few input regions. At this point the question of learnability also arises with regard to constructing sample sets from such problems. In particular, the sample set must capture the topology of the problem by representing each region sufficiently well to insure it is both represented and distinct. We would also like it if a *random* set of points chosen from each class were very likely to provide these necessary characteristics of region representation. These aspects of problem topology and learnability will be discussed later with regard to some specific problems.

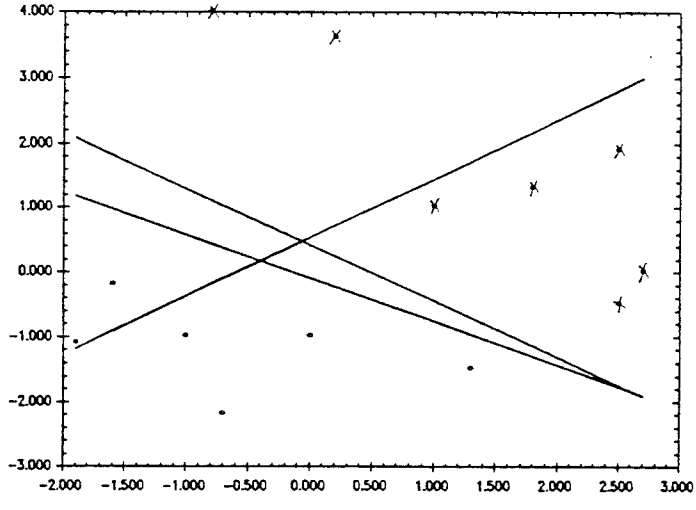
5.1.2 Performance of the CLUSTER Algorithm

In order to realize a gradient descent training procedure on the previously described energy function we must decouple the training of the first layer from that of the following layers. In the case of 2-layer machines, this is in some sense suboptimal since a particular set of first representations may not be linearly separable yet may still represent the minimum of our criterion function. This would preclude the solution of the problem via our method using a 2-layer machine. It is our claim however, that problems well suited to this algorithm will contain sufficiently few regions so as to with very high probability, insure linear separability after the first layer. Following first layer training, the second layer of our example networks were trained using the well known perceptron learning procedure. The results of some of our simulations are shown in Figures 5.3-5.8.

Figure 5.3 shows graphically the weight matrix evolution for a 2:3:1 problem (i.e., 2 inputs, 3 middle layer neurons, and 1 output) as the constraint that minimizes the number of regions is enforced. It is important to notice that the



(a)

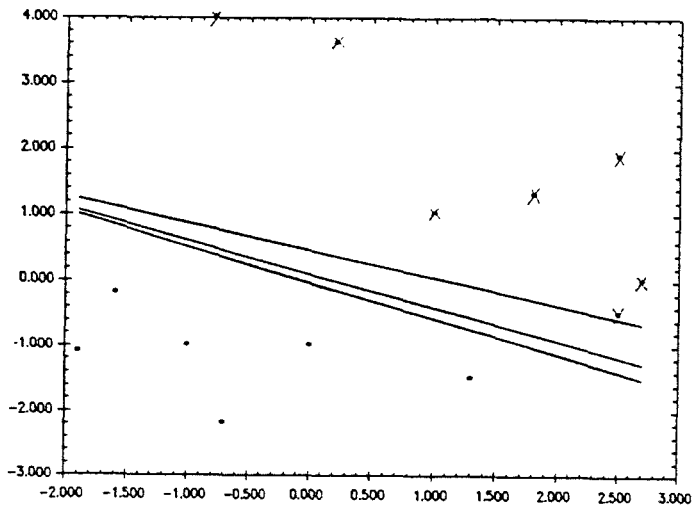


(b)

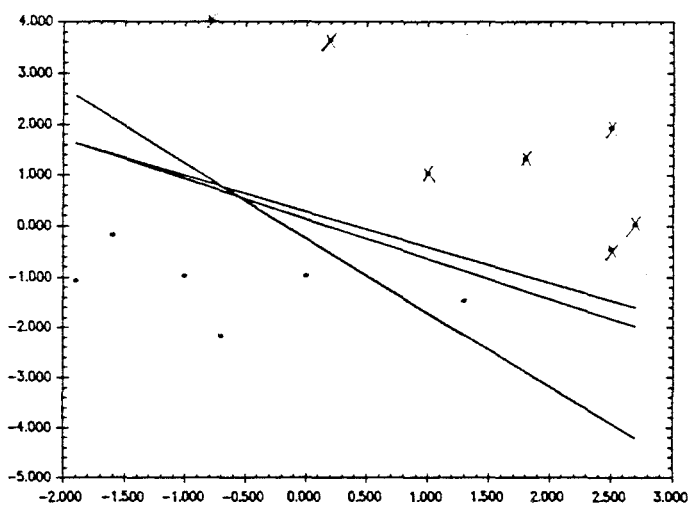
Figure 5.3 : Input space region evolution for a 2:3:1 problem.

(a) Initial input space configuration before training.

(b) Input space configuration using CLUSTER after 10 iterations.



(c)



(d)

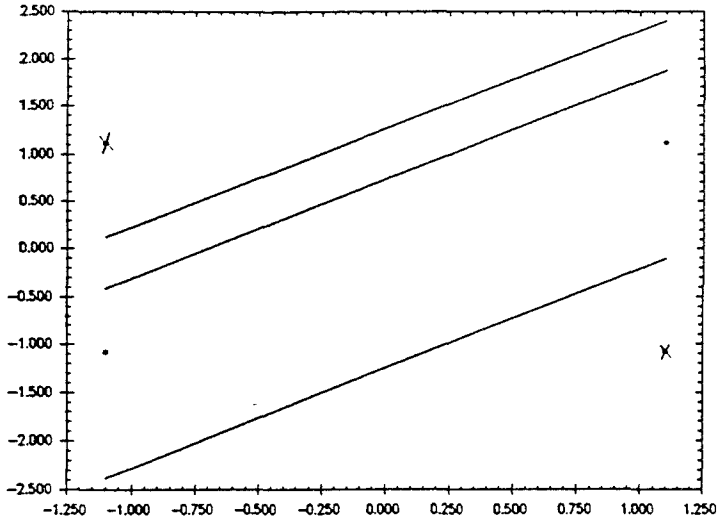
Figure 5.3(cont).

(c) Input space configuration using CLUSTER after 30 iterations.

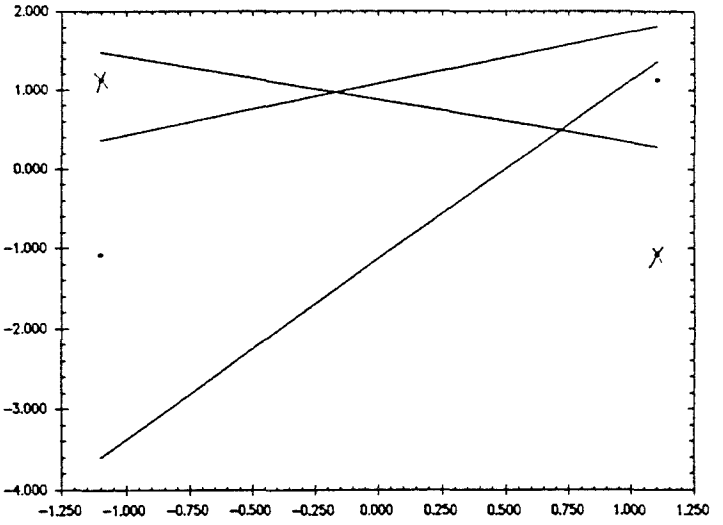
(d) Final input space configuration after training with BEP.

problem could have been solved following iteration number ten. That is, the intermediate representations were linearly separable at this point during training; however, since the input samples only reflect information regarding two regions, the algorithm continues until these two regions emerge (see Figure 5.3c). In contrast to this behavior, Figure 5.3d shows the final solution to this problem obtained using an output error driven gradient descent procedure described earlier (BEP). The existence of unspecified regions (i.e., regions in the input space whose representations did *not* play a role in the training procedure) like the one shown in the upper left of Figure 5.3d, corrupts the performance of the system in the presence of new data. Since we do not know where this region will be mapped in the output, new data falling in this region may be misclassified. In a more complicated example this characteristic will limit the network's ability to generalize.

Since most real world problems are not linearly separable, it is important to examine the behavior of our algorithm in the presence of non-linearly separable training data. Figure 5.4 shows both CLUSTER and BEP solutions to the XOR problem simulated on a 2:3:1 network. Apparently, the region minimization constraint is well suited to capturing the problem topology in this case regardless of the number of degrees of freedom (over the minimum required, of course) we provide to the system. This is an important feature of any algorithm being utilized to solve a problem whose complexity is *a priori* unknown. Providing a network that is too large to BEP will very often result in correct classification of the training set, with poor performance on new data. This is partially the result of undefined or unclassified regions as have been generated in Figure 5.4b.



(a)



(b)

Figure 5.4 : Solution to XOR problem.

(a) Final input space configuration after training with CLUSTER.

(b) Final input space configuration after training with BEP.

Implementing the representation minimization constraint, however, results in the elimination of many unclassified regions thereby only using the network degrees of freedom necessary to solve the problem. The elimination of these undefined regions contributes greatly to the generalization seen in later simulations using the CLUSTER algorithm.

It becomes difficult to discuss regions and problem topologies in spaces of dimension higher than three; however, the problems described in Figures 5.5 and 5.6 do indeed lend themselves to such descriptions. In Figure 5.5 we show the training functions used in this simulation. Input data to the CLUSTER algorithm consisted of 7 dimensional sample vectors taken from 6 strictly increasing functions for class1 and class6 strictly decreasing functions for class2. As in all the simulations described here, the weight matrix for this example was initialized randomly and small. Specifically, the network begins in a fully interconnected configuration. As can be seen from the resulting weight matrix shown in the figure, 4 out of 6 middle layer neurons learned to subtract the endpoints of the input vector in order to determine its nature (i.e., increasing or decreasing). Generalization in this experiment was very high ($> 95\%$) as we would expect from the resulting hyperplane configuration. The only vectors misclassified on testing this network were those whose chord slopes (i.e., slope between first and last points) were particularly small. In Figure 5.6 we modify class2 of the previous problem so that the classes will no longer be linearly separable. Class2 is now given by 7 dimensional sample vectors taken from the 6 *nonincreasing* functions shown. After training on these 12 vectors, the system generalization was tested to be 88%. In these two problems the set of vectors used for testing were chosen

Network 7:6:1

Training Set :

Class 1	Class 2
1. $10. + \exp(x/2)$	1. $-2 - 2x$
2. $\exp(2x) - 4$	2. $(x - 10)^2 - 5$
3. x	3. $1. - 7x$
4. $x + 3$	4. $4\exp(-x/4)$
5. $(x/2) - 7$	5. $\exp(-(x + 10)/10) - 12$
6. $(x + 10)^2 - 25$	6. $10/(x + 11)$
all increasing on [-10,10]	all decreasing on [-10,10]

(a)

$$\underline{\underline{W}}_{cluster} = \begin{pmatrix} 0.2 & 0.0 & 0.0 & -0.1 & -0.1 & -0.2 & -0.2 \\ 1.5 & 0.6 & -0.4 & -0.6 & -0.8 & -1.0 & -1.2 \\ 0.7 & 0.2 & -0.2 & -0.3 & -0.4 & -0.5 & -0.6 \\ 1.4 & 0.5 & -0.4 & -0.5 & -0.7 & -0.9 & -1.1 \\ 1.9 & 0.7 & -0.5 & -0.7 & -0.9 & -1.3 & -1.6 \\ 1.7 & 0.6 & -0.4 & -0.6 & -0.9 & -1.1 & -1.4 \end{pmatrix}$$

(b)

Figure 5.5 : Problem definition and 2-layer solution to the increasing/decreasing function dichotomy using CLUSTER.

(a) Training functions.

(b) First layer weight matrix after training.

Network 7:6:1

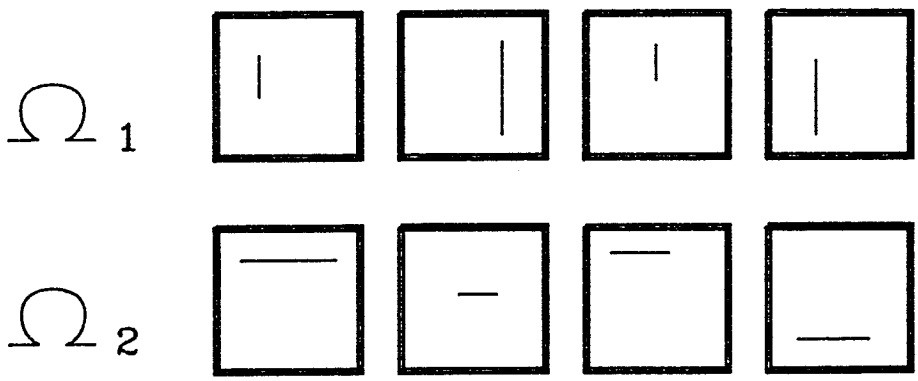
Training Set :

Class 1	Class 2
1. $10. + \exp(x/2)$	1. $-2 - 2x$
2. $\exp(2x) - 4$	2. $3(x/10)^2$
3. x	3. $1. - 7x$
4. $x + 3$	4. $-(x/5)^2$
5. $(x/2) - 7$	5. $4\exp(-x/4)$
6. $(x + 10)^2 - 25$	6. $sin(\pi x/10)$
all increasing on [-10,10]	all nonincreasing on [-10,10]

Figure 5.6 : Input functions used to train increasing/nonincreasing function dichotomy.

randomly.

The last two-class problem we simulated is related to visual processing and consists of training shift and scale invariance. Figure 5.7 shows the training sets used for this example. They consist of various scaled and shifted horizontal and vertical lines. The testing set consisted of 14 vectors (again shifted and scaled horizontal and vertical lines) not presented to the network during the training phase. Since shift and scale invariance do not represent topological features of the problem space to which the algorithm is well suited, we expect in this experiment to be obtaining only a nearest neighbor interpolation characteristic. That is, since we trained the network on many versions of lines we expect the notion of nearness in the input space to be the aspect of the system, which will yield correct



(a)

Network	CLUSTER	BEP
50: 30: 1	28%	14%
50: 10: 1	57%	21%
50: 8: 1	71%	21%
50: 5: 1	36%	14%

(b)

Figure 5.7 : Horizontal vs. vertical line dichotomy.

(a) Examples of training vectors used to obtain shift and scale invariance.

(b) Generalization comparison between CLUSTER and BEP for several network sizes.

classification. In as much as the CLUSTER algorithm will result in input regions that are as large as possible, we might expect this sort of nearest neighbor feature to favor our minimization constraint as compared to BEP. The results shown in Figure 5.7 for networks of different sizes, support this expectation.

The final simulation result included here for completeness, reflects a rather simple modification of the original criterion function so as to facilitate its function in a multiclass environment. The new energy function is given by

$$\begin{aligned} \mathrm{E} = & \frac{1.}{K}\sum_{k=1}^{K}\frac{2.}{M_k(M_k-1)}\sum_{p=1}^{M_k}\sum_{q=p}^{M_k}|\underline{x1}^p(k)-\underline{x1}^q(k)|^2 \\ & +\frac{2.}{K(K-1)}\sum_{k1=1}^{K}\sum_{k2=k1}^{K}\frac{1.}{M_{k1}M_{k2}}\sum_{p=1}^{M_{k1}}\sum_{q=1}^{M_{k2}}\frac{1.}{|\underline{x1}^p(k1)-\underline{x1}^q(k2)|^2} \end{aligned} \tag{5.2}$$

where K is the number of classes or clusters, M_k is the number of vectors in class k, and $\underline{x1}^p(k1)$ is the first representation of the pth vector in class k1. This simple extension of the CLUSTER algorithm to multiclass problems allows it to be used as a first layer training procedure for three layer networks or in problems of distribution estimation. A simple example of dichotomizing 3 classes in 2 dimensions using the modified CLUSTER algorithm is shown in Figure 5.8.

Since, using the CLUSTER algorithm, each input vector must be compared to every other input vector, our procedure will be slower in converging by a factor of $\sum_{k=1}^{K} M_k$ per iteration as compared to BEP. We hope, of course, that in constraining the allowed solution in the described way, the network will require fewer samples to specify a problem completely. This being true would reduce the time factor difference between the two training procedures. Local minima

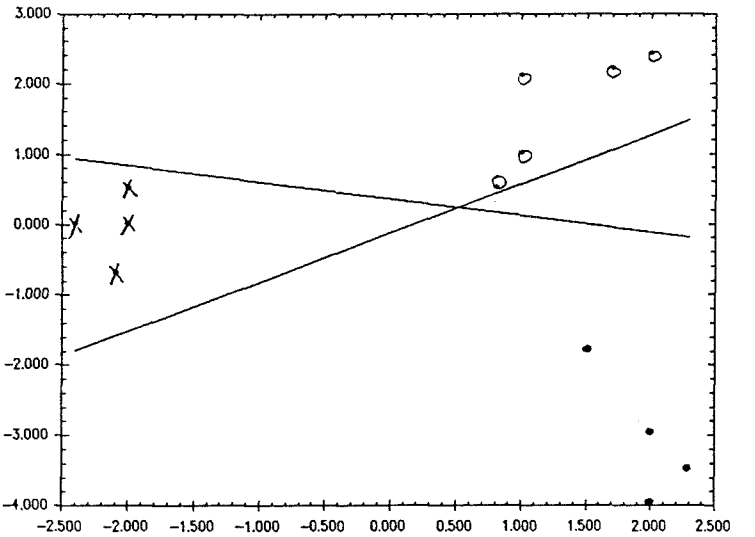


Figure 5.8 : Input space configuration after training 2:2:2 network on a three class problem using CLUSTER.

can become a problem with this as with any other gradient descent procedure. Perhaps the local minimum problem is a bit worse in this case since local minima in the choice of intermediate representations may result in a nonlinearly separable set. This type of behavior was not seen to be critical in the simulations we tried.

5.1.3 Discussion of Problem Topologies

In this subsection we will more closely examine the topologies of several problems. In keeping with our original assumption that the problem be doable on a 2-layer machine, we will address two questions with respect to the problems under consideration here. First, can the problem be adequately described by relatively few regions? We have already seen that this feature of the problem space can be used to advantage when it is well represented in the training data. So the second question naturally arises: Can the input space topology of the

problem be well represented by a relatively small number of training samples? To compile a comprehensive list of problems/functions to which the answers to the above questions are "yes" is a task beyond our understanding at this point. We instead describe several problems with regard to these questions in hopes that similar ways of thinking may be used on future problems with success.

The most attractive problem from the perspective of region minimization is a linearly separable problem. Although treating a linearly separable problem with a multilayered machine might appear artificial, the discussion of the previous section regarding *a priori* problem information together with the results accompanying that discussion, make this a logical place to begin. The increasing/decreasing problem consists of 2 regions in a 7 dimensional input space. Since there are only 2 regions, a particularly small number of points are necessary to specify this problem topology with high accuracy. The continuous nature of the input space is an additional feature of this problem making it attractive from the perspective of a region minimization approach to learning. Specifically, it is quite likely in learning a piecewise continuous function, that the nearest neighbors of the training points will lie in the same class. This is to be contrasted with many problems which operate on a binary input space. In the binary case, the majority of nearest 'allowed' neighbors (i.e., all vectors within unit Hamming distance) of a particular vector are often to be placed in another class. This feature reflects a high degree of criticality associated with the presence of a specific point in the training set. In order to make this point clear, consider the so-called Penzias problem. In this problem we would like to classify binary vectors as class1 if they contain fewer than 2 contiguous groups of 'on' bits while classifying them

as class2 vectors otherwise. In an input space of reasonable size ($N > 4$) any vector in class1 has N nearest neighbors and only 4 of these can belong to class1 also. The remaining N-4 nearest neighbors of course belong to class2. Since the same class nearest neighbor feature is not seen in the Penzias problem, learning this function via a region minimization type of algorithm would most likely be unsuccessful. This was in fact the case in the simulation experiments performed on this problem. We should note here that a "degrees of freedom" minimization constraint should be considerably more successful in this case since the problem is represented by local first layer connections.

The increasing/nonincreasing function problem is a good example of a non-linearly separable problem that is well suited to the CLUSTER algorithm. Class1 is simply defined by one region as $x(N) > x(N-1) > ... > x(2) > x(1)$ where $x(i)$ is the i^{th} bit of the input vector. In terms of learnability, we require only that the class1 region is represented and that the class2 vectors are sufficiently well scattered to have labeled all of the remaining regions. Since class1 is represented as one region, the algorithm will find that one region by virtue of the pulling together force term in our energy function. At that point, the main limitation to generalization will be unlabeled regions belonging to class2. Thus, if the points in class2 are well scattered we expect all the remaining regions to contain a point in class2 . This will insure correct classification of all points falling outside of class1.

A very common testbed problem in neural networks is the parity problem. Parity topology is actually fairly well suited to the region minimization constraint in that there are only N+1 regions. This can be easily seen by considering a 2-

layered network to solve the parity problem. The first layer may essentially count the number of ones in the input vector, each middle neuron having a threshold 1 larger than the neuron below it. Having counted the number of ones in a vector, the resulting representations are now linearly separable for the output layer to dichotomize. Only N+1 representations may be generated this way reflecting the N+1 regions specified by the problem. The difficulty however is again in the training since training vector criticality is even stronger in the parity problem than with Penzias. Every nearest neighbor of a vector in the parity problem is in the other class. The somewhat disappointing conclusion then with respect to the parity problem is that the CLUSTER algorithm *is* well suited to learn the mapping given *all* 2^N input output pairs; however, generalization in the case where the training set is smaller than 2^N elements should not be expected.

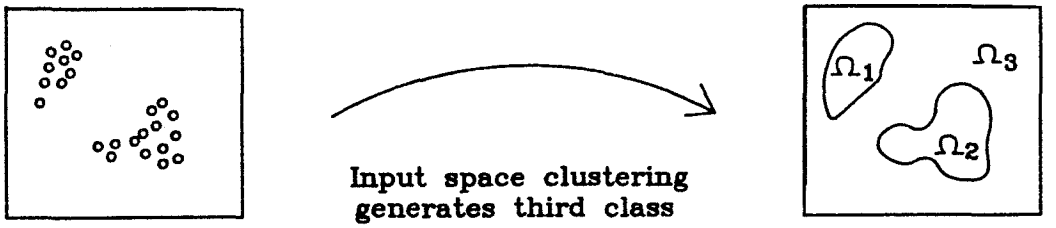
Problems whose topologies consist of clusters in the input space, or whose distributions are modal are also well suited to region minimization techniques. This point may appear obvious since relatively few clusters may be assigned to relatively few regions by the multiclass procedure described earlier. The pushing away force term in the CLUSTER criterion function however, provides a means by which very complicated input space regions might be approximated by hyperplanes. By simply adjusting the number of intermediate neurons, the relative accuracy of this approximation technique could be changed. Using similar reasoning, a three layer training algorithm based on the procedure described by Lippmann may be implemented.[16,27] First, the training set would be divided into K clusters using some classical input space clustering algorithm. Next, these K clusters would be presented to the pushing apart term of the CLUSTER al-

gorithm as K different classes. CLUSTER would then bound each input space cluster by hyperplanes as shown in Figure 5.9 whereupon a neuron in the next layer could be devoted to each input space region containing a point. Classification using such a scheme follows easily and is discussed in reference [27]. This problem has not been approached yet to any significant degree in simulation; however, it appears to be an interesting topic for future work.

5.1.4 Multilayer Networks

Many image oriented pattern recognition problems will be suited to a two layer CLUSTER type solution only after an initial *local* feature extraction stage of processing. As shown in Figure 5.10, local features such as eyes and mouths often define the relevant input space dimensions of the problem at hand. Further, since it is the global configuration of these features that determines the classification of the input vectors, the first layer feature extraction must be unsupervised as shown in Figure 5.11. In this way, the features or principle components of the training vectors may be learned to define the necessary input space dimensions for the next layer. The weights in the next layer may then be trained using a supervised algorithm such as CLUSTER in order to evolve a linearly separable set of representations that the last layer may classify correctly.

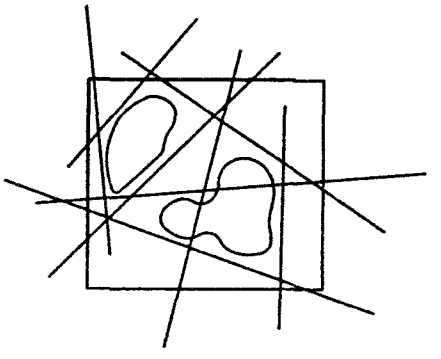
This type of strategy may be extended to the case of networks with large numbers of layers. The gradual progression from local features and local connectivity to more global receptive fields in deeper layers as shown in Figure 5.11, allows for progressively larger features to be built in deeper layers of processing. At the same time there is an increase in the degree of supervision provided



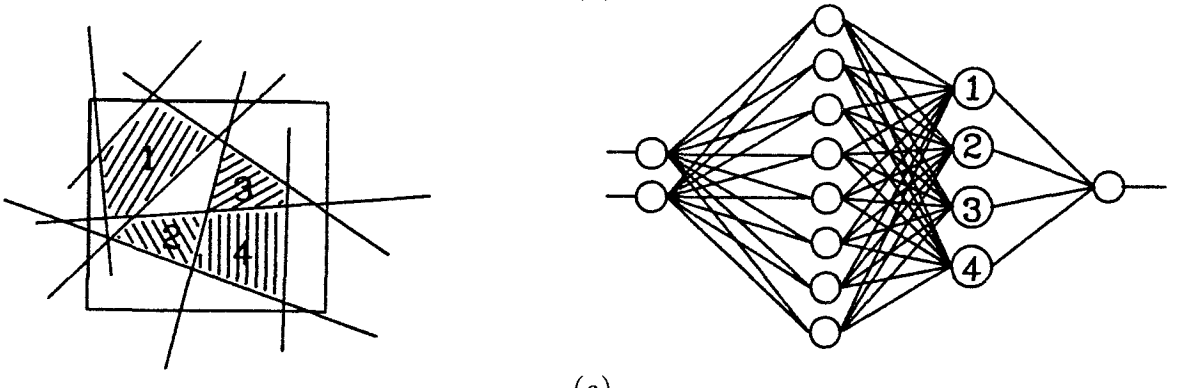
(a)

Pushing apart terms

$$E_{pq} = \frac{1}{|\underline{x1}^p - \underline{y1}^q|^2} + \frac{1}{|\underline{y1}^p - \underline{z1}^q|^2} + \frac{1}{|\underline{x1}^p - \underline{z1}^q|^2}$$



(b)



(c)

Figure 5.9 : Three layer construction using Lippmanns approach with CLUSTER for first layer training.

(a) Define input space clusters with conventional grouping algorithm.

(b) CLUSTER pushing apart terms for learning separating hyperplanes.

(c) Construct second layer 1 neuron per region.

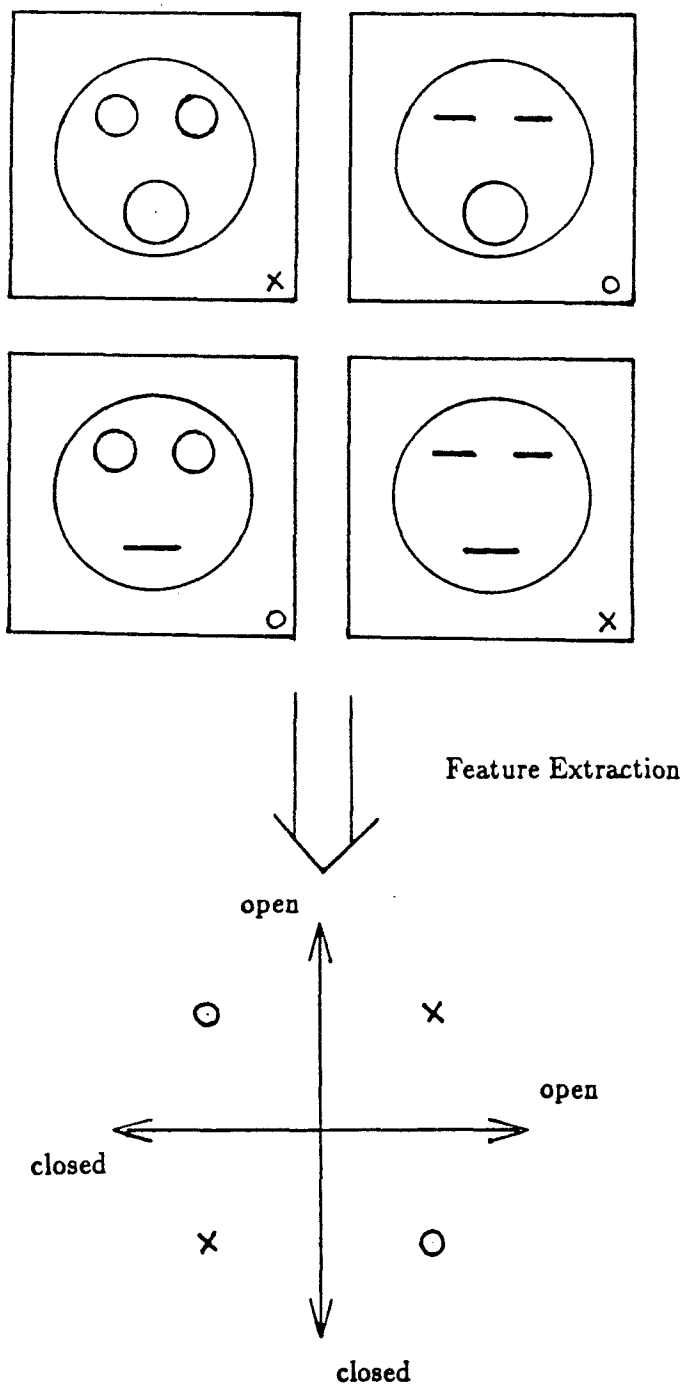


Figure 5.10 : Facial XOR problem.

Local features define relevant input space dimensions with global arrangement determining final class assignment.

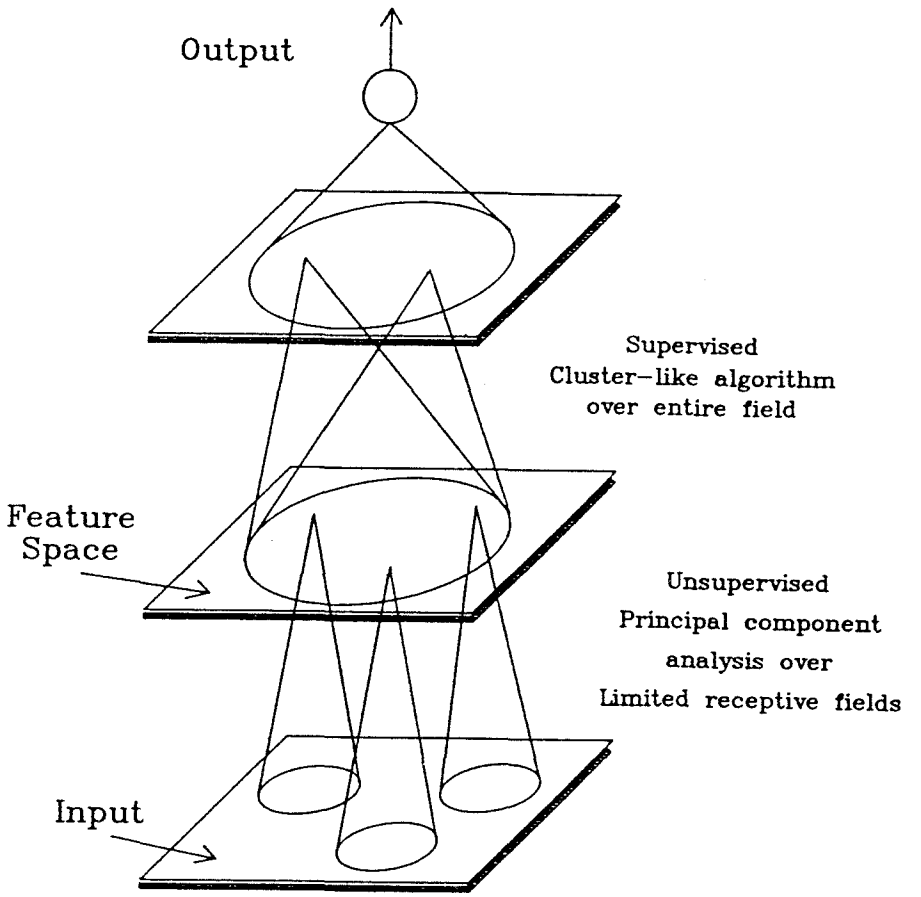


Figure 5.11 : General multilayer network training philosophy. Training becomes more supervised and receptive fields become broader with deeper layers.

during the training of weights in deeper layers. In analogy with human visual processing, such a learning scheme may result in useful adaptive image processing networks. A great deal of future work must be directed toward a deeper understanding of specific network constraints, and the problem characteristics these constraints favor. Such an understanding would mark a large step in the direction of understanding generalization in these networks.

5.2 K-Nearest Neighbor Algorithm

The main limitation of the CLUSTER algorithm, as well as BEP, is learning time. Although in principle relatively few training samples are necessary to train the network with CLUSTER as compared with BEP, actual training times in simulation were still quite long for problems with dimensions greater than $\approx$10. For example, the horizontal vs. vertical line problem takes 1-10 days of computing time on a SUN3/60 workstation. In many cases long training times are unacceptable and a real time algorithm is required so that new training data may be incorporated into the pattern recognition system during operation. One such real time training algorithm is the k-nearest neighbor algorithm (KNN).[120–125]

The procedure defined by the KNN algorithm is simply to accept and store new training samples along with their classifications in a large memory. Training time therefore is determined only by the speed at which we can update the reference library. The bottleneck for the KNN algorithm occurs when we are actually performing the classification of an unknown input. The algorithm specifies that *given an unknown input vector, it should be classified according to the class most abundantly represented by its k nearest neighbors.* The specific procedure then is

to compute the distances between the input and all members of the training set and allow the KNNs to vote. The class assignment of the majority of the KNNs is taken as the system output. In the case of a tie, several tie breaking procedures are plausible among which are letting k=k+1 (i.e., adding a neighbor) or letting k=k-1 (i.e., removing a neighbor). The success of this algorithm is dependent on having a relatively large training set so that the underlying problem structure is well represented over a large portion of the input space. The required training set is therefore potentially exponential in the dimensionality of the input space, resulting in a large storage requirement for this approach. The corresponding calculation time for interrogating such a system is also very long. We will soon see that these requirements are well matched to an optical disk-based implementation. The storage capacity of the disk is large enough to store many vectors, while parallel optical access provides a mechanism through which an unknown input may be compared against the training set. Later in this chapter we will describe two different optical implementations of the KNN algorithm, one in which the correlation peak is used as a distance metric for image classification and another in which true Euclidean distance is computed optically.

5.3 Radial Basis Functions

The radial basis function (RBF) approach to pattern recognition lies between the previous two algorithms in terms of learning time.[56,126] This approach results in a network whose processing units are slightly more complicated than simple linear thresholding devices; however, the overall network computation required to perform a single classification is much less time consuming than in the KNN ap-

proach. The motivation for using a RBF network to perform pattern recognition tasks comes from the relatively well-established mathematical framework associated with regularization theory and hypersurface reconstruction. In hypersurface reconstruction the problem is to construct a *machine* or function $\hat{F}(\underline{W}, \underline{X})$, which takes a vector $\underline{X}$ into a prescribed output $F(\underline{X})$. For simplicity, we will consider only one-dimensional outputs, the generalization to higher dimensions being obvious. In order to construct $\hat{F}(\underline{W}, \underline{X})$ a set of training samples taken from the mapping to be learned (i.e., the underlying hypersurface to be approximated) is provided $\{\underline{x_i} \rightarrow F(\underline{x_i}); i = 1, ..., M\}$ and the problem becomes to choose the *form* of $\hat{F}$ and the appropriate parameters $\underline{W}$, so that $\hat{F}(\underline{W}, \underline{x_i}) = F(\underline{x_i})$ for $i = 1, ..., M$. This problem is very similar to the pattern recognition problem where one is given a set of training patterns and is asked to find a classifier $\hat{F}$ with the appropriate parameters $\underline{W}$, such that the resulting machine classifies the training set correctly. In both cases we desire that future samples be mapped correctly and that the system behaves well in the presence of noise. In order to obtain these desirable characteristics in hypersurface reconstruction, a criterion of smoothness is often placed on the estimator $\hat{F}$. One approach for specifying a smooth $\hat{F}$ is the RBF approach.

If we choose $\hat{F}$ to be of the following form

$$\hat{F}(\underline{W}, \underline{X}) = \sum_{i=1}^{\tilde{M} \leq M} a_i exp(-|\underline{X} - \underline{t_i}|^2 / \sigma_i^2) \tag{5.3}$$

where $\underline{W} = \{\underline{t_i}, \sigma_i, a_i : i = 1, ..., M\}$, then the resulting machine is called a radial basis function interpolator/classifier. We see that the RBF classifier seeks to approximate the underlying function as a weighted sum of gaussian "bumps."

According to the above expression, the approximation $\hat{F}$ comprises $\tilde{M}$ of these bumps each centered at $\underline{t_i}$ with width σ_i and weighted by a_i in the final output. Given a set of training vectors $\{\underline{x_i} : i = 1, ..., M\}$, we may estimate the parameters $\{\underline{t_i}, \sigma_i, a_i : i = 1, ..., M\}$ so that $\hat{F}(\underline{x_i}) \approx F(\underline{x_i})$ using any number of supervised and/or unsupervised algorithms.

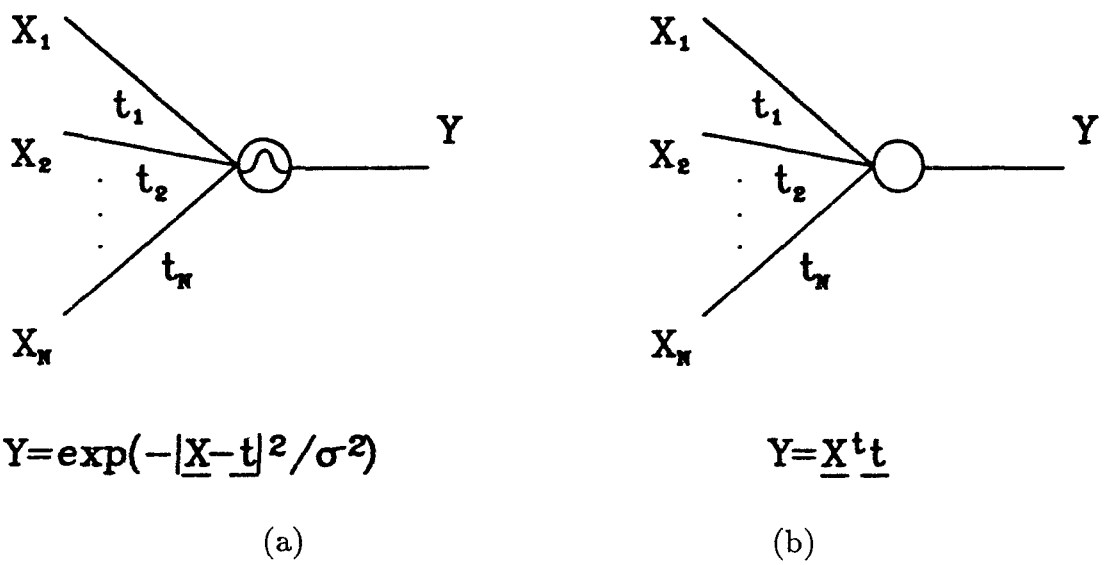


Figure 5.12 : Processing units proposed for RBF network.

(a) RBF unit. (b) Linear unit.

The RBF approach may be mapped quite naturally onto a neural network architecture as shown in Figure 5.12. We define the RBF unit in Figure 5.12a as a neuron with response given by

$$y_i = exp(-|\underline{X} - \underline{t_i}|^2/\sigma_i^2), \tag{5.4}$$

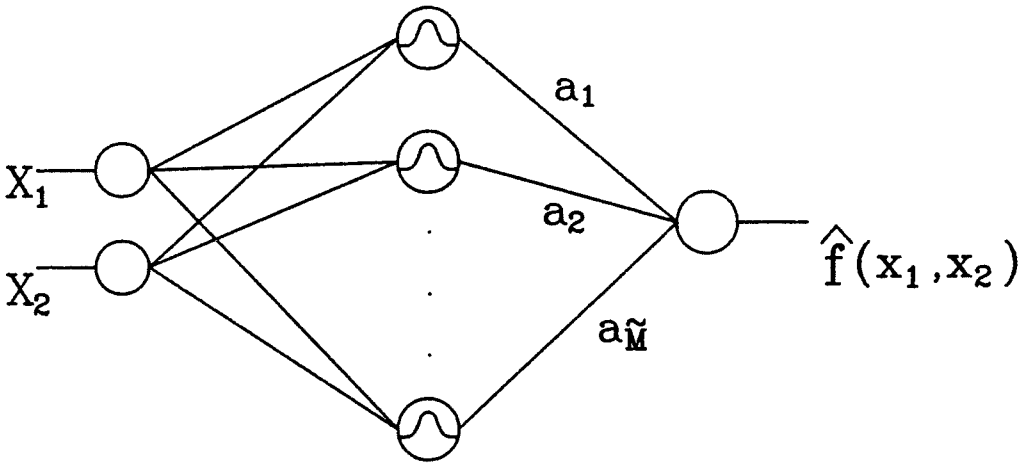
where $\underline{t_i}$ is called the center or template associated with neuron i. The output

layer of the RBF network consists of a single linear unit whose output is simply the weighted sum of its inputs as shown in Figure 5.12b. The overall network mapping then is defined by

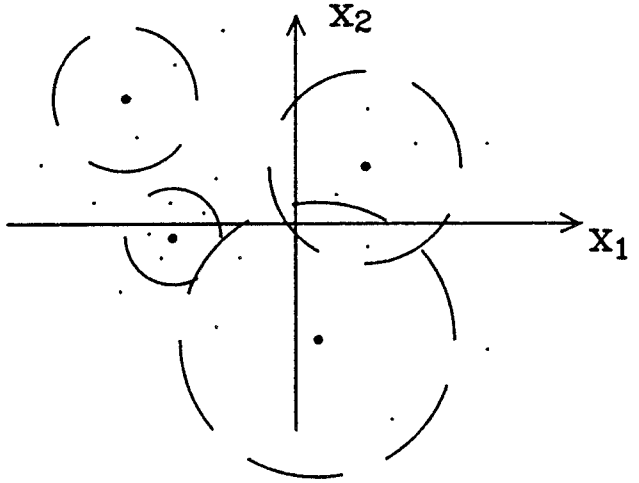
$$\hat{f}(\underline{W},\underline{X}) = \sum_{i=1}^{\tilde{M}\leq M} a_i exp(-|\underline{X}-\underline{t_i}|^2/\sigma_i^2), \tag{5.5}$$

as desired. In Figure 5.13a we show a RBF network for estimating a function of two variables and in Figure 5.13b we depict an example of an input space configuration that may result from the network mapping. The points in Figure 5.13b represent the training data and the broken circles represent the four gaussian basis functions that comprise the RBF network. In order to generate Figure 5.13b we utilized a k-means algorithm with k=4 to determine the centers of the basis functions. This procedure results in determination of the four centers shown. In order to determine the widths associated with each center, a KNN algorithm was used. The five nearest neighbors to each center were chosen and the average of these five distances was used as the width σ, for the associated bump. Note that these procedures result in the determination of the centers $\{\underline{t_i}\}$ and the widths $\{\sigma_i\}$ in a completely unsupervised fashion. Specifically, the first layer of a RBF network may be trained without using an error driven procedure, thereby significantly reducing training time. Training of the output layer can be accomplished through the use of either a mean squared error minimization procedure or the relatively simple perceptron learning algorithm.

In the next three sections we present experimental results for optical disk-based systems designed to implement the two previous networks : the KNN



(a)



(b)

Figure 5.13 : RBF network.

(a) RBF network for estimating a function of two variables.

(b) Input space configuration arising from the network in (a) with four RBF units.

network and the RBF network. Simulation results for these two approaches will appear in the following sections for comparison with their optical counterparts.

5.4 Shift Invariant KNN Network

The first optical pattern recognition system to be described here is the shift invariant k-nearest neighbor classifier shown in Figure 5.14. This system is based on the rotating mirror correlator and the KNN algorithm discussed earlier. The optical system consists of two arms as shown. The bottom arm is the input stage and includes an optical mouse which is used to provide a handwritten character to the system. For our experiments, the character is written on a 16×16 grid and appears on the monitor shown. This monitor is imaged onto the *write* side of a Hughes Liquid Crystal Light Valve (LCLV) which is then read out using a linearly polarized 5mW He-Ne laser.[127–129] The external polarizer P converts the polarization modulation induced by the LCLV to intensity modulation appropriate for input to the rotating mirror correlator which comprises the top arm of the system in Figure 5.14. As the disk rotates, the correlation between the handwritten input character and each member of the correlation library is generated optically as described in section 4.1. An example of the system operation is given in Figure 5.15. In Figure 5.15a, the coherent image of an input pattern is shown immediately after the polarizer P. An exact duplicate of the image shown in Figure 5.15a has been written in sector 0 of the optical disk and the resulting x and y slices of the 2-D autocorrelation pattern are shown in Figures 5.15b and 5.15c respectively. As we can see from Figure 5.15c the y-slice of the correlation pattern is peaked with a width of $\approx 25\mu$s corresponding to a mirror

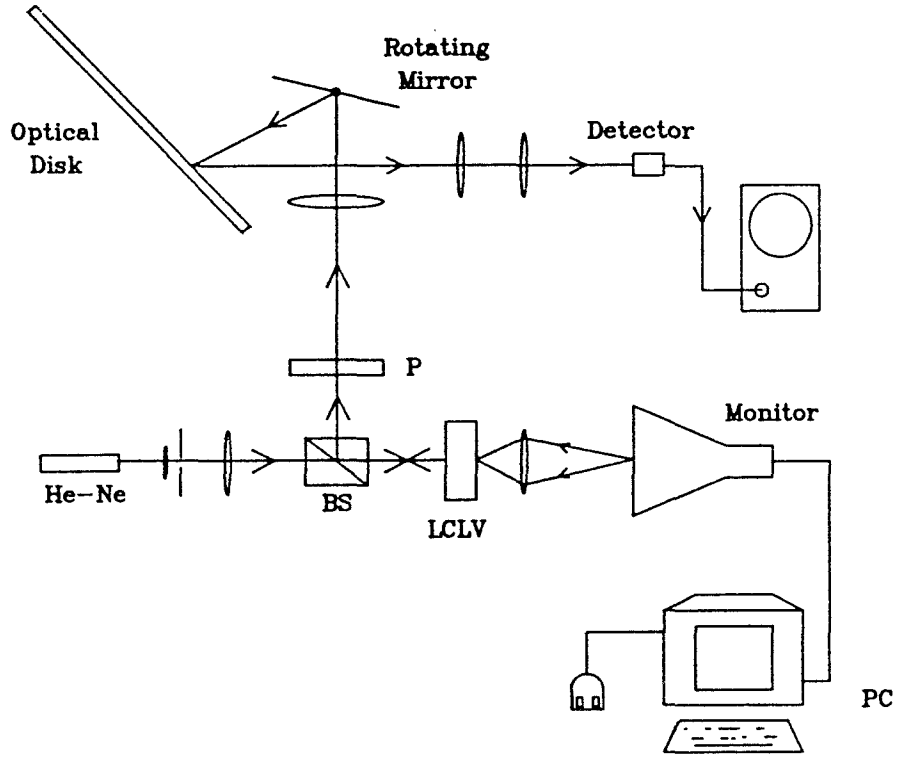
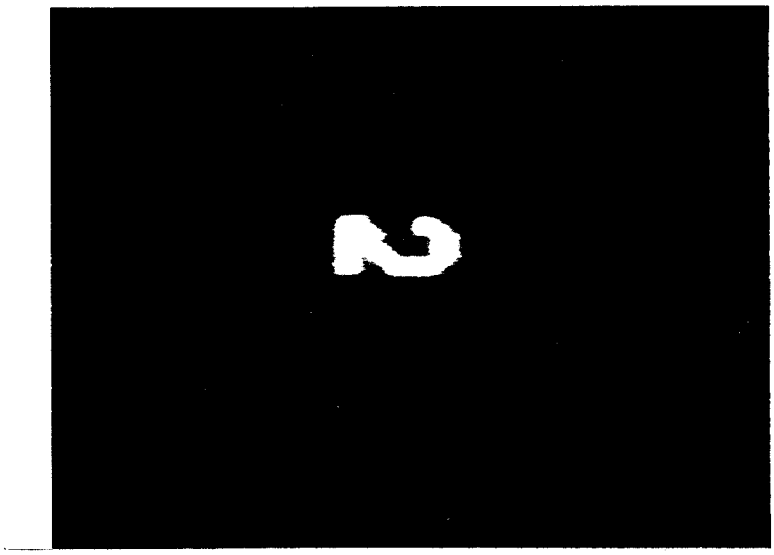
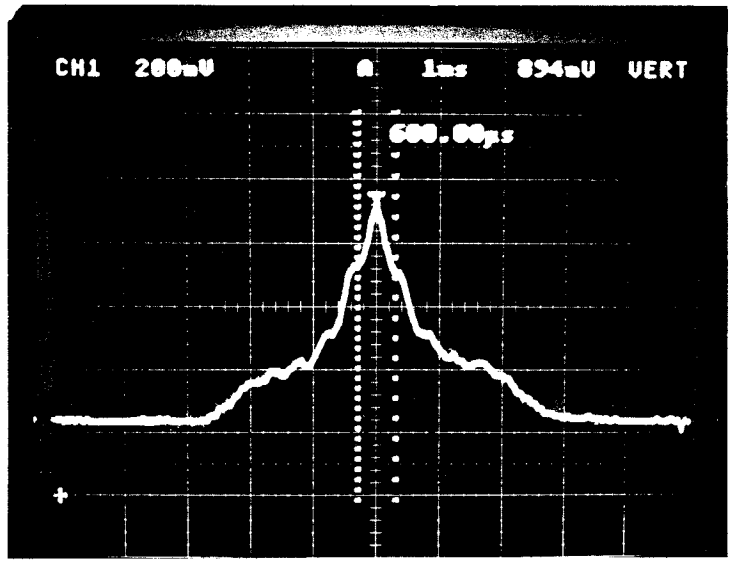


Figure 5.14 : Optical implementation of shift invariant KNN classifier for handwritten character recognition.



(a)

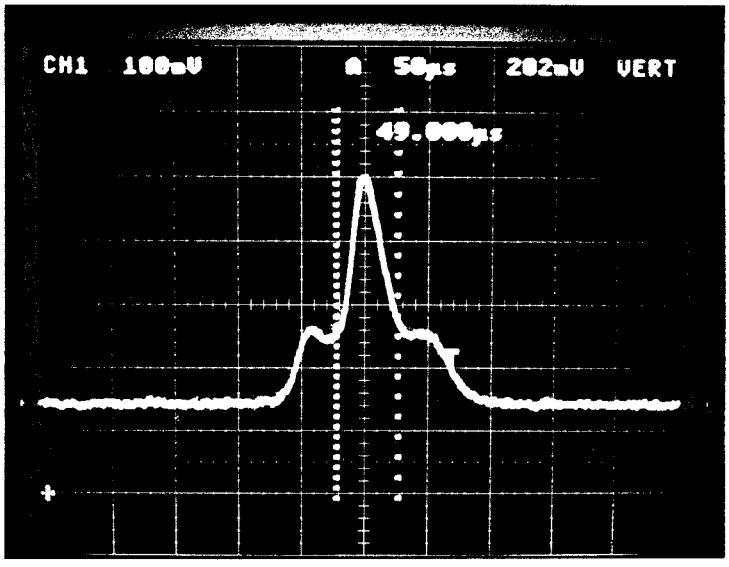


(b)

Figure 5.15 : Demonstration of shift invariant optical KNN classifier.

(a) Coherent image of handwritten numeral 2 after polarizer P.

(b) x-slice through optically generated autocorrelation pattern of (a).



(c)

Figure 5.15(cont).

(c) y-slice through optically generated autocorrelation pattern of (a).

scan rate of 2.5KHz and a radial image extent of 16 pixels. The x-slice of the correlation pattern is much broader and corresponds to a disk rotation rate of ≈2Hz. Using a data acquisition board in an IBM PC, the correlation data from the system is obtained from the detector for postprocessing. The peaks of the 2-D image correlations are detected and used as a similarity measure to determine the KNNs of the unknown input character. In this way, a handwritten input character is classified according to which images on the optical disk it correlates most strongly.

For this experiment, a relatively small library of reference images was stored on the disk to discriminate among three classes of handwritten numerals : twos, threes, and fours. The initial training set comprised 60 images (20 per class) and an optimization was performed in order to determine the 8 best images per class for recording on the disk. Only 24 filters were used here so that one filter could be stored per sector, thereby making the optical system as well as the data acquisition system more straightforward. The following discussion describes the algorithm used to choose the 24 *best* images for disk storage. First we form the 60×60 matrix of cross correlation peaks $\rho_{ij} = max\{\underline{x^i} * \underline{x^j}\}$, where $\underline{x^i}$ and $\underline{x^j}$ are two of the 60 initial reference images and the max function is taken over the 2-D correlation plane. We then choose the 8 templates per class with the largest χ_i where

$$\chi_i = \sum_{j \in \text{correct class}} \rho_{ij} - \sum_{j \in \text{wrong class}} \rho_{ij}. \tag{5.6}$$

The performance of these 24 filters is then tested on the training set using the k=1 nearest neighbor rule. If there is an error during testing so that template i

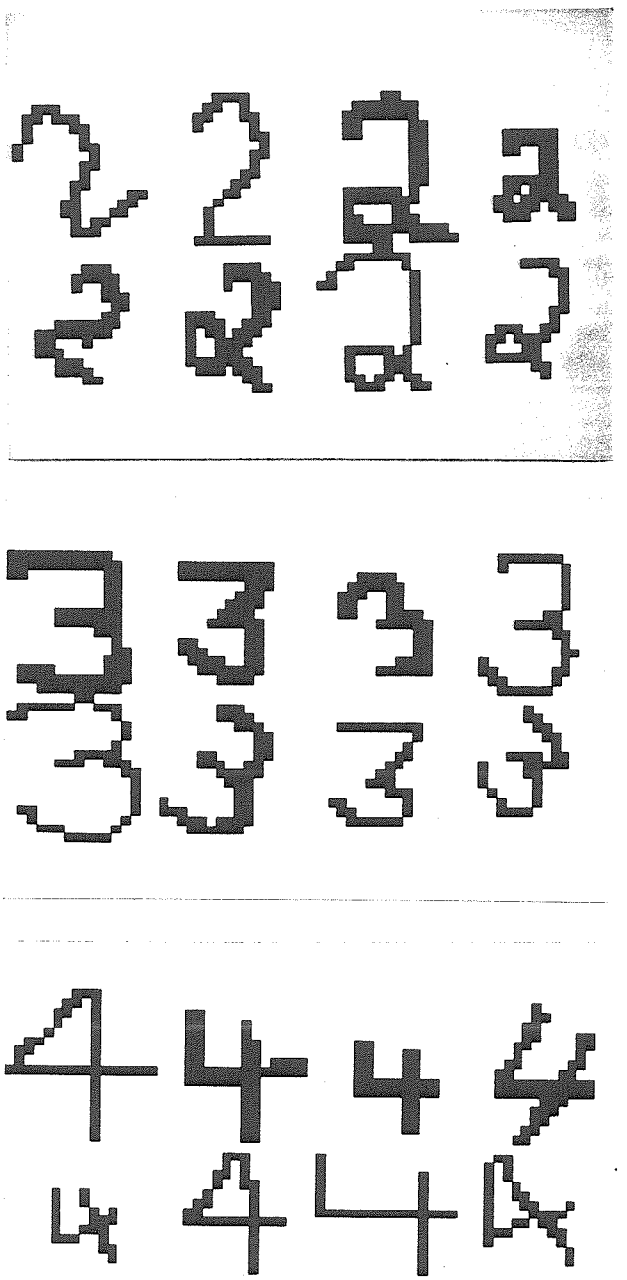


Figure 5.16 : Reference images used in 2-3-4 recognition experiment.

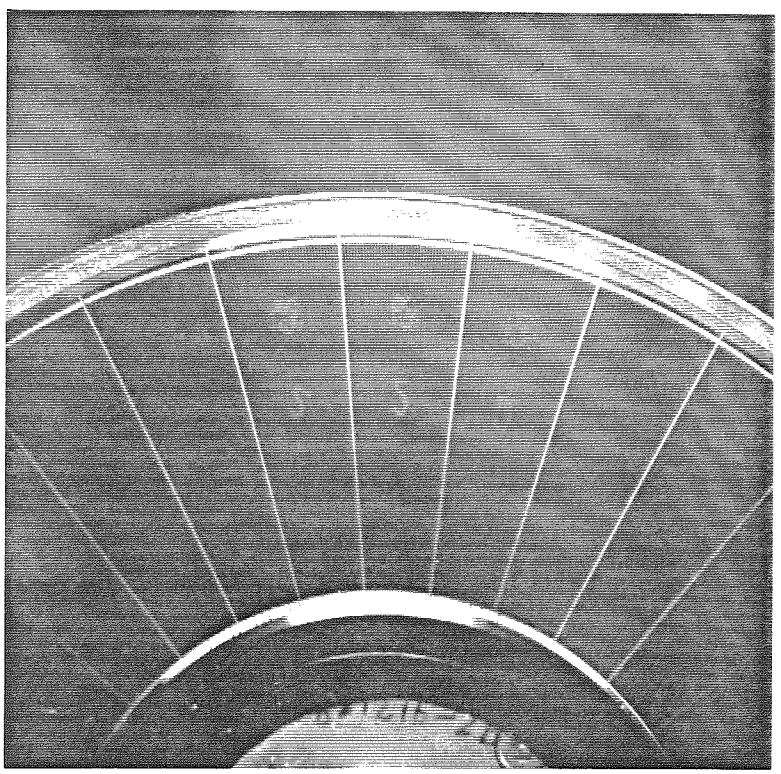


Figure 5.17 : Photograph of Sony disk containing 2-3-4 filters.

misclassifies vector j, then template i is replaced by template i' where $\rho_{i'j} < \rho_{ij}$ and χ_i is maximized. This procedure insures that the new template is less likely to misclassify vector j while still maintaining a good overall classification score χ. This algorithm was found to converge to the set of 24 reference images shown in Figure 5.16 with correct classification of all 60-24=36 remaining training patterns. In Figure 5.17 we show a photograph of the optical disk on which these 24 filters were recorded. Several '2' filters can be seen in the photo, one filter per sector. Also visible in the photo are the sector markers on the disk which appear as bright radial lines between filters. The reference images shown in Figure 5.16 were originally drawn on a 16×16 grid but were enlarged before recording on the disk so that a single reference superpixel actually comprised 128 tracks and 192 pixels along track. In addition to these 24 filters, three master filters were recorded on the disk for alignment purposes. The layout of the reference library is shown in Figure 5.18. During testing, it was found that only 16 of the 24 reference images were responsible for performing the classification; therefore, only these filters were actually recorded on the disk. It is the layout of these 16 filters together with the master filters that we depict in Figure 5.18. In Figure 5.19 we show examples of the raw output from the optical system for each of the three master images as input. The equally spaced peaks of equal height that can be seen in the photos arise as a result of diffraction from the sector markers on the disk. In some sectors there are filters so that the peaks seen between some sector markers represent the correlation between the input pattern and the reference image stored in that sector. The pattern of full and empty sectors in these photos enables us to identify sector zero by comparison with the layout in Figure 5.18,

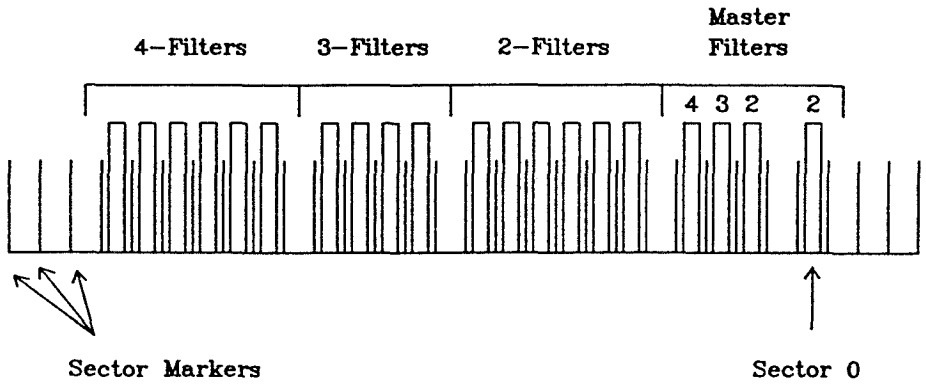
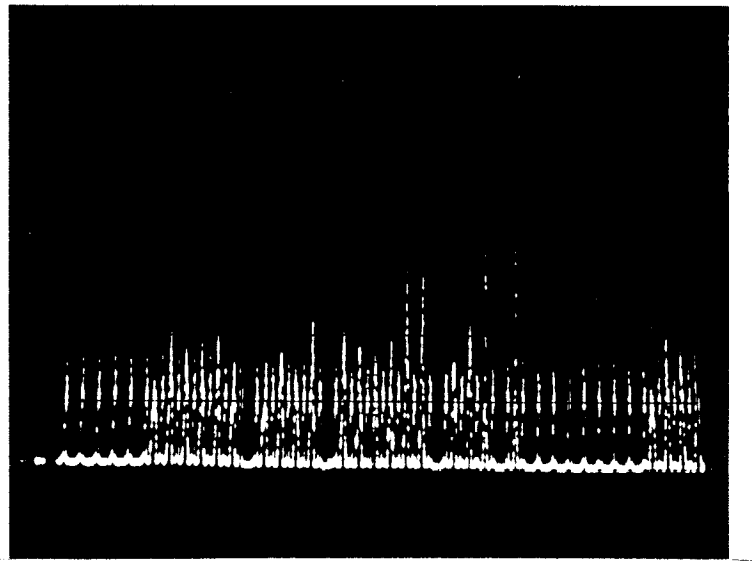


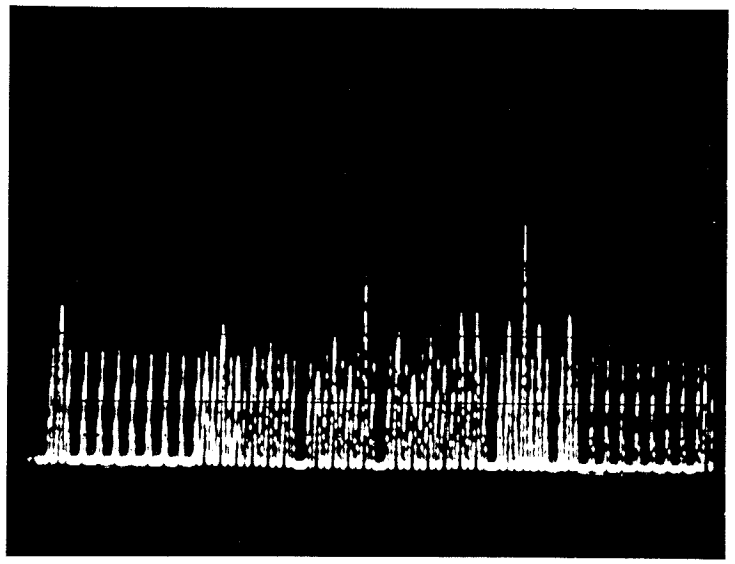
Figure 5.18 : Reference library layout.

thereby allowing us to label each correlation peak with its corresponding class assignment. Returning to Figure 5.19, since the master image of '2' is exactly matched to the master reference of '2' written in sectors 0 and 2, a large peak is obtained from these two sectors as shown in Figure 5.19a. Although the master images were *not* part of the training set, we see large outputs from several of the '2' filters also, indicating that our reference library has classified the input correctly. Similarly, for '3' and '4' master inputs we see that large responses are illicited from the correct batches of reference filters on the disk.

The correlation peaks shown in Figure 5.19 must be normalized before they can be quantitatively compared. Normalization here requires that each correlation peak value be divided by the number of ones in the corresponding filter. There are two ways of extracting the necessary normalization data (i.e., the number of ones in each reference image). One way is to simply store the values in memory while the filters are being written to the disk and synchronously read



(a)

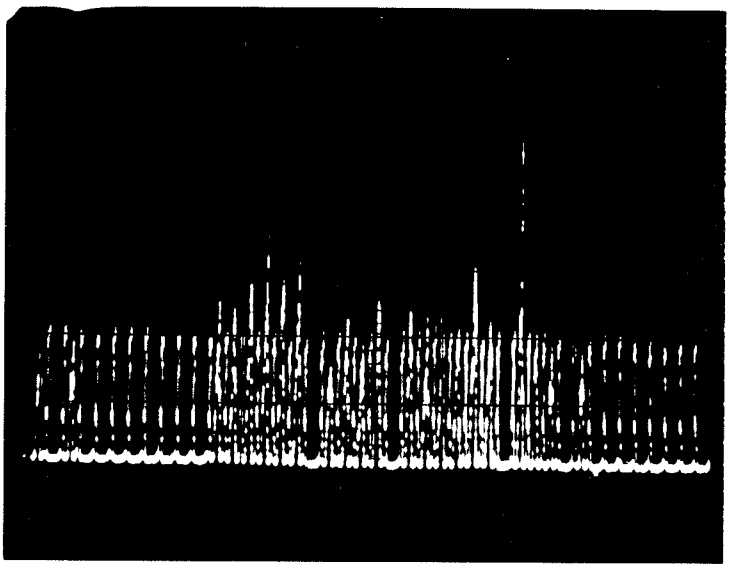


(b)

Figure 5.19 : Raw output from optical KNN system.

(a) Input = master image 2.

(b) Input = master image 3.



(c)

Figure 5.19(cont).

(c) Input = master image 4.

out this memory during classification. The second way is to present the *all one* input image to the optical system and use the resulting correlation peaks as normalization data. For comparison purposes, both of these methods were used and the two sets of data are presented in the histogram of Figure 5.20. The agreement between the optically generated normalization data (open bars) and the actual filter weights (solid bars) is excellent and leads us to conclude that the optical system is performing as we expect. Once this normalization data is stored in normalization memory the system is ready to perform classifications. A summary of the recognition algorithm used in this experiment is shown in Figure 5.21 and a photo of the optical system is shown in Figure 5.22. When this system was tested on a 200 element testing set, a 75% recognition rate was achieved for the k=1-NN rule. This is in good agreement with a simulation result of 73% for the same testing set. When the number of nearest neighbors was increased to k=3, the optical system performance was improved to 85% correct recognition. System performance was observed to decline sharply for k$\geq$7-NN (recognition $\leq$72%) in agreement with theoretical predictions.

Although 85% recognition rate is quite good as compared with the expected *random* rate of 33%, a commercially useful PR system recognition rates above 95%. An increase in the size of the reference library which requires the use of a larger training set could potentially result in this level of performance; however, the principal limitation of this system is the post processing speed and data requirements. The need to search an entire 2-D correlation pattern for the necessary peak is quite time consuming. We saw earlier that the correlation peak width is $\approx$ 15μs requiring a sampling rate $>\approx$200 kHz with this, the slowest

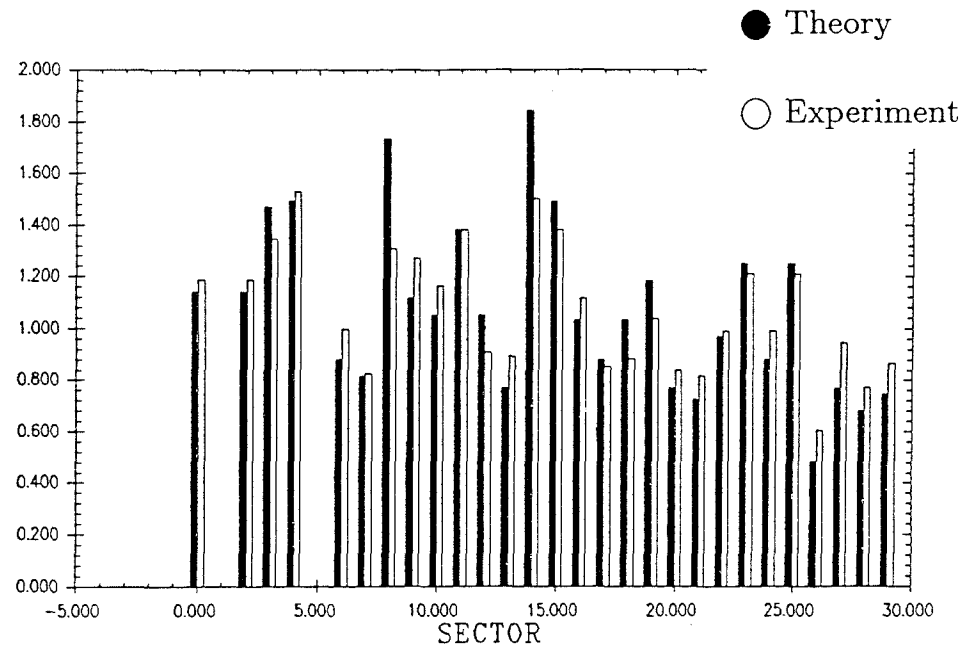


Figure 5.20 : Histogram comparing experimental and theoretical normalization data.

of the disk-based correlators. Further, in order to scan the entire reference library, a full disk rotation of 0.5s must be sampled resulting in a data array of 100,000 samples representing a maximum of 195 filters. The next optical system we describe reduces this high output data rate at the expense of a shift and scale normalizing preprocessor. As a result of this preprocessing, we will have the potential to realize a much larger reference library ($> 10^6$ reference images) in a realtime implementation requiring a minimum of dedicated postprocessing hardware.

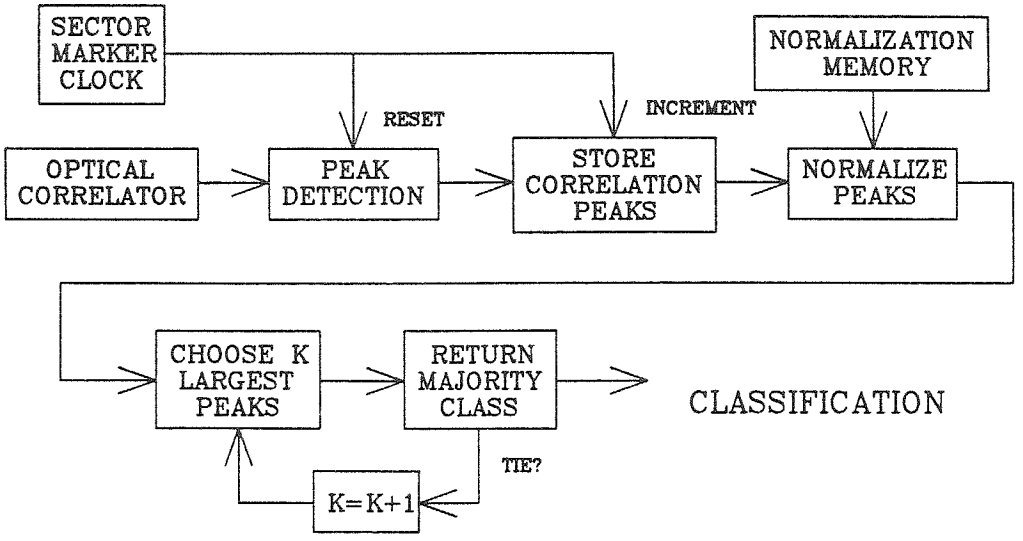


Figure 5.21 : Block diagram of recognition algorithm used in 2-3-4 experiment.

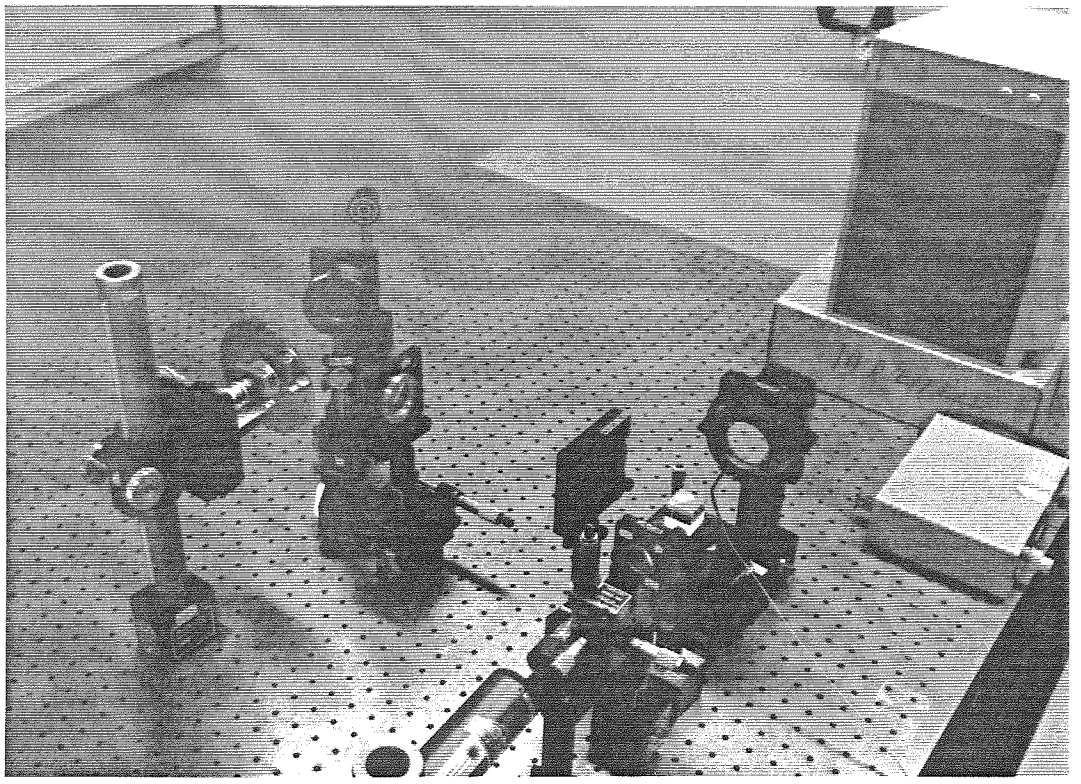


Figure 5.22 : Photograph of optical KNN classifier for 2-3-4 recognition.

5.5 Shift Variant KNN Classifier

The capability of optics to compute inner products is one of the most powerful motivators for the use of optics in neural network implementations. In the previous chapters of this thesis we have exploited this capability in designing image correlators whose operation is defined as the generation of *sequences* of inner products. Although shift invariance is achieved in the previous architectures by virtue of performing many inner products per filter, the complexity of the optical system as well as that of the postprocessing electronics can be reduced by only requiring one inner product per filter to be computed. Reducing the complexity of the previous architecture in this way leads to the shift variant k-nearest neighbor classifier shown in Figure 5.23. In this system the inner product is used as a similarity measure between the input and each reference image. Since this distance metric is not invariant to shifts of the input image, the PR system is no longer shift invariant. This loss of shift invariance can be compensated through the use of a centering preprocessor based on either a moment or image extent computation. We found much better performance with the latter approach for the case of handwritten character input images. In the image extent approach, the maximum length of the input character is found in both the horizontal (x) dimension and the vertical (y) dimension resulting in the orthogonal image extents $xdim \leq 16$ and $ydim \leq 16$. The required centering is performed by shifting a left/top justified version of the input character by $(16 - xdim)/2$ pixels in x and $(16 - ydim)/2$ pixels in y. During this procedure it is straightforward to incorporate a scaling operation so that the preprocessed image is centered *and* scaled to a 10×10 window. The scaling is achieved by applying the scale factor

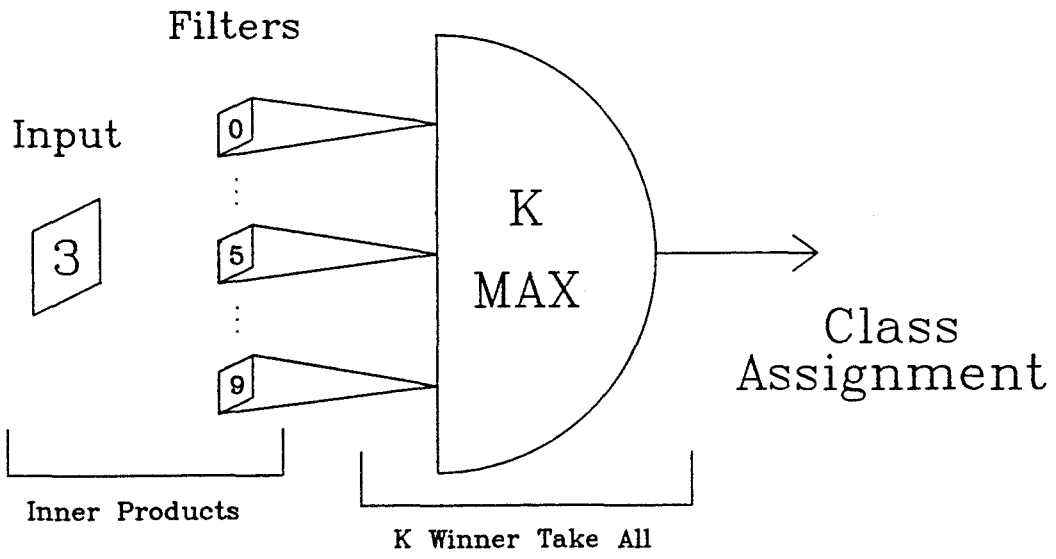
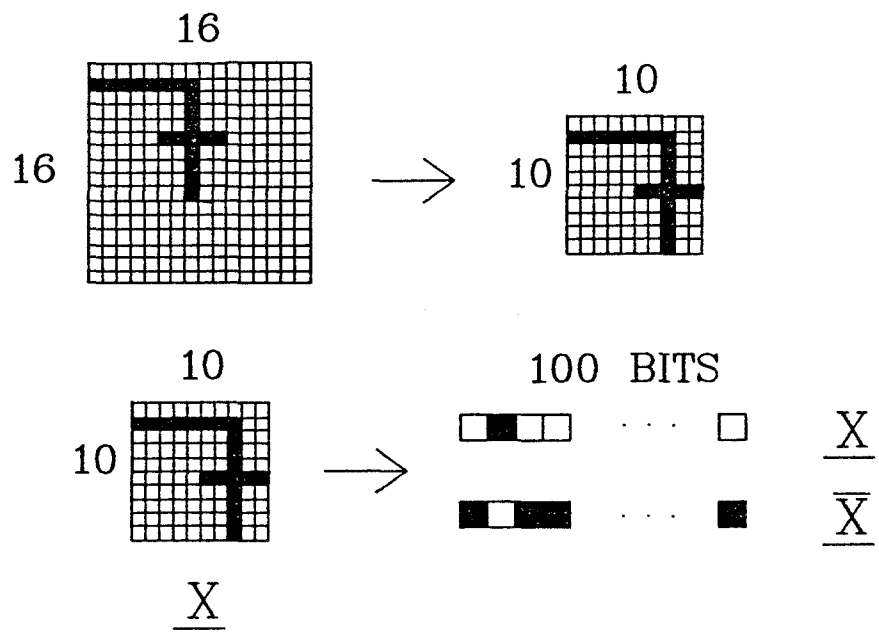


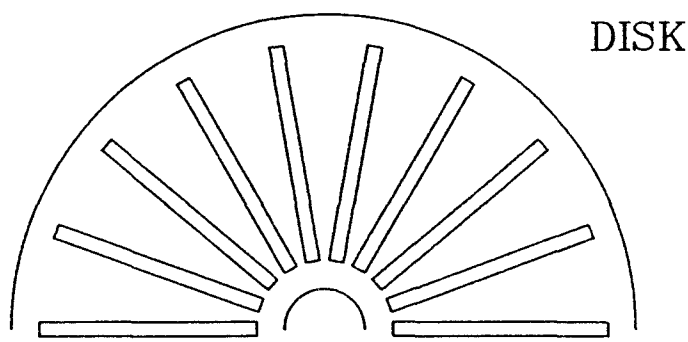
Figure 5.23 : Shift variant optical KNN classifier.

$\max\{xdim, ydim\}/10$ to both the x and y input dimensions. Following preprocessing, the 10×10 pixel input field is *unrastered* to form a 100 bit binary vector as shown in Figure 5.24. Each such vector $\underline{x}$ corresponding to each preprocessed filter or reference image, is stored on the optical disk as a radial line. For each vector $\underline{x}$, we also store its complement $\overline{\underline{x}}$ in an adjacent sub area. This method of *dual rail encoding* allows us to simulate bipolar filters, thus eliminating the need to remove signal-dependent bias from the detector output. The pixel size in this experiment was chosen to be 177 tracks by 116 pixels along track. Storing one vector per sub area allows us to record up to 43 vectors per sector or 1376 templates per disk. The upper bound on template library size is achieved when every pixel along track is associated with a different stored image. This implies that a reference library of $> 10^6$ images is possible using this scheme.

A schematic of the optical system used to perform the comparison between some unknown preprocessed input image and the reference library stored on the



(a)



(b)

Figure 5.24 : Preprocessing required to convert 16×16 image into 100 bit binary vector.

(a) Shift and scale normalization with unrastering.

(b) Vectors stored on disk as radial lines.

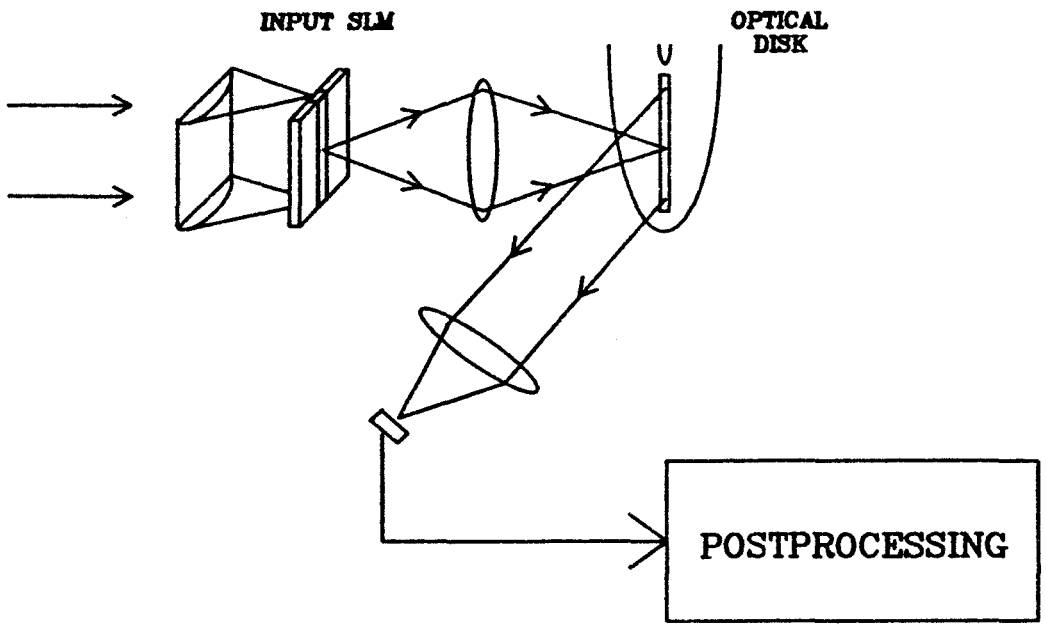
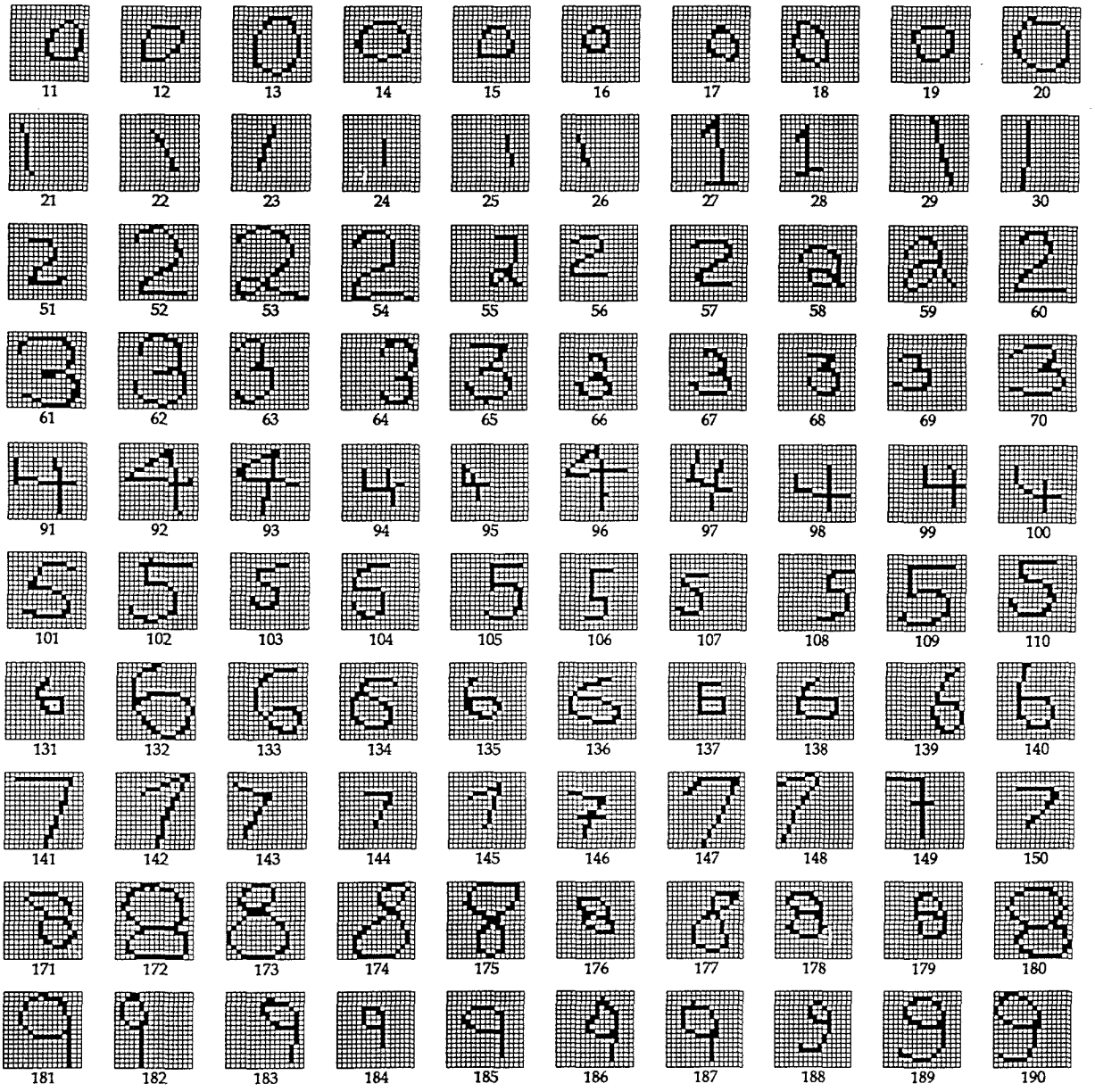


Figure 5.25 : Optical system used to compare input vector provided via 1D SLM with optical disk-based library.

disk in the format described above, is shown in Figure 5.25. In this architecture, a 1-D SLM is used to present the preprocessed input character to the system as a 100 bit binary vector. An image of the input vector is formed as a radial line on the disk as shown, and the total diffracted intensity is collected by the output lens and measured using a Photodyne 1500XP detector. The detector output once again represents the inner product between the input vector and the illuminated reference vector. The postprocessing system for this experiment consists of two parts. First, a sample/hold (S/H) circuit is used to detect the peaks of the raw detector output. These peaks represent the desired inner products. The S/H circuit is clocked by a signal that is phase locked to the sector markers in the data stream. The second stage of postprocessing consists of an A/D converter board in an IBM PC followed by software that implements the desired recognition algorithm.

In this experiment we will implement the KNN algorithm in the solution of the handwritten character recognition problem. We will consider the 10 class problem of identifying handwritten numerals 0-9. Recently, this problem has received much attention and several automated recognition techniques have been quite successful in providing a solution. The U.S. Postal Service is particularly interested in these recognition systems for application to zip code recognition. Using a SUN3/60 workstation, several authors were asked to draw the numerals 0-9 on a 16×16 grid. The resulting database of 950 images (95 digits per class) was separated into a 300 element testing set and a 650 element reference library. Examples of characters from the training and testing sets are shown in Figure 5.26. The 650 reference images were preprocessed as described above and stored on the disk along with their complements, as 100 bit binary vectors. Using a disk rotation rate of 20Hz, these 1300 vectors were processed at a rate of 2600 inner products per second equivalent to 260,000 binary operations per second. The 300 testing images were preprocessed and stored as 100 bit vectors on a set of 10 transparencies, an example of which is shown in Figure 5.27.

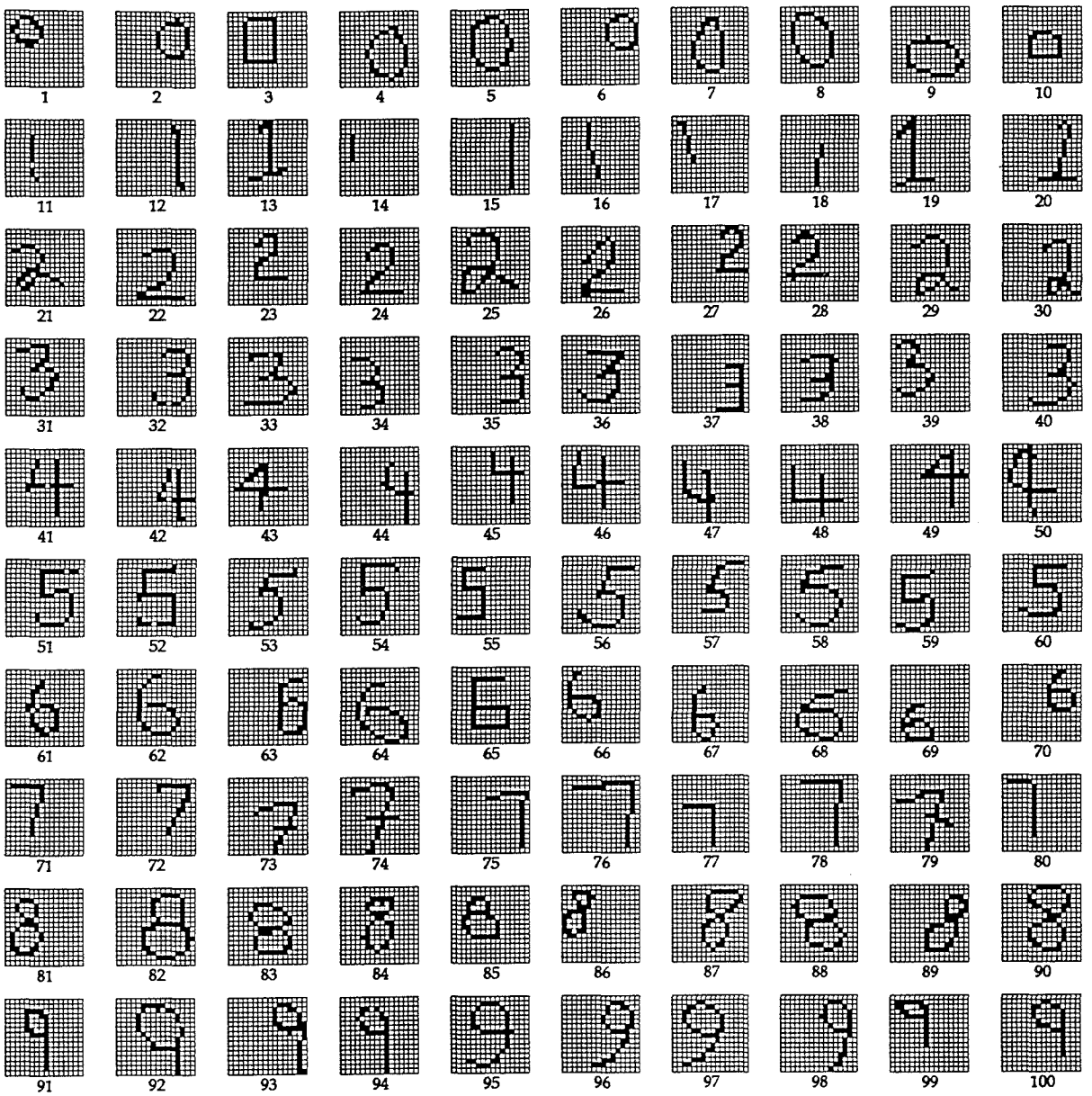
The topmost and bottommost rows of this transparency represent alignment markers which facilitate magnification and position calibration during system operation. The other 30 rows correspond to the 30 preprocessed images from class zero. An example of the raw detector output for the all one input vector is shown in Figure 5.28. The two tallest peaks in this trace correspond to sector markers on the disk and represent the inner product between the all one vector and itself. From this data we can calculate the effective brightness per input pixel as measured at the detector to be 0.6nW. This value is in good agreement with



(a)

Figure 5.26 : Training and testing sets used in 0-9 handwritten character recognition task.

(a) 10 per class from reference library.



(b)

Figure 5.26(cont)

(b) 10 per class from testing set.

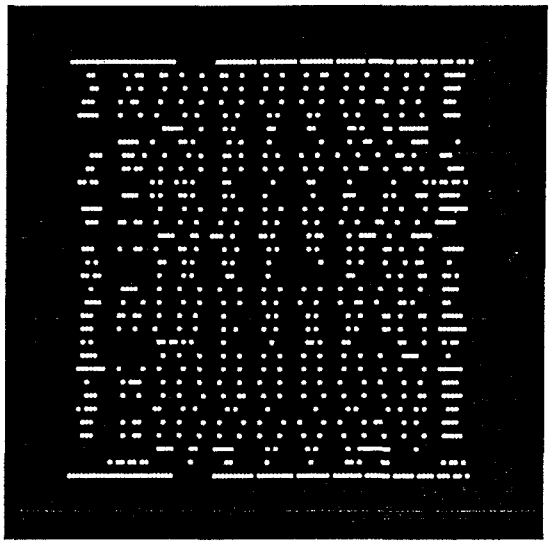


Figure 5.27 : Input transparency used in 0-9 experiment.

the known optical losses in the system. The other peaks in Figure 5.28 provide normalization data as before. An example of the signal following the S/H circuit is shown in Figure 5.29. This trace includes a portion of sector 0 which is empty on this disk. The low-level signal seen in sector 0 is indicative of the system noise level before quantization by the A/D converter in the PC. A discussion of noise sources and their effect on system performance will be included later in this section. The PC samples the signal shown in Figure 5.29 once per level, averages 4 rotations worth of data (total acquisition time ≈0.2s) and computes the Euclidean distances from the inner products as

$$|\underline{y} - \underline{x}|^2 = |\underline{y}|^2 + |\underline{x}|^2 - 2\underline{x} \cdot \underline{y}, \tag{5.7}$$

where $\underline{y}$ is the unknown input image and $\underline{x}$ is a stored reference. Since our optical system actually measures $\underline{x} \cdot \underline{y}$ and $\overline{\underline{x}} \cdot \underline{y}$, we may form the distance for binary

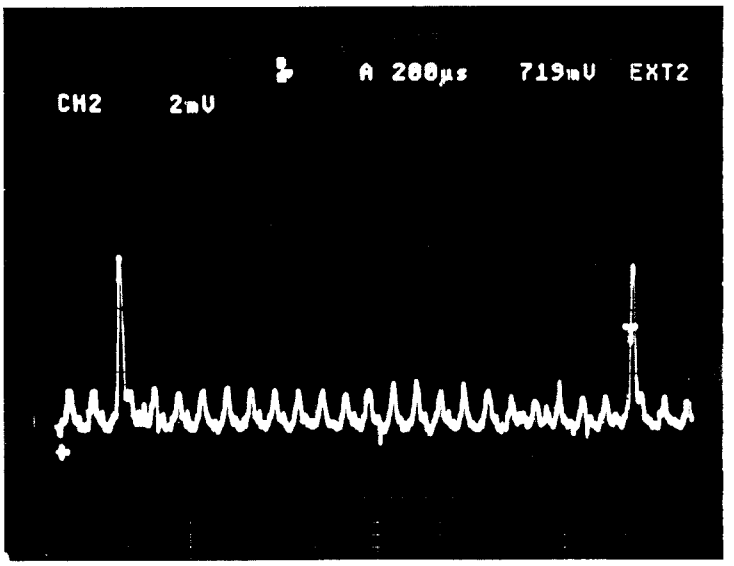


Figure 5.28 : Raw detector output showing normalization signal for 0-9 system.

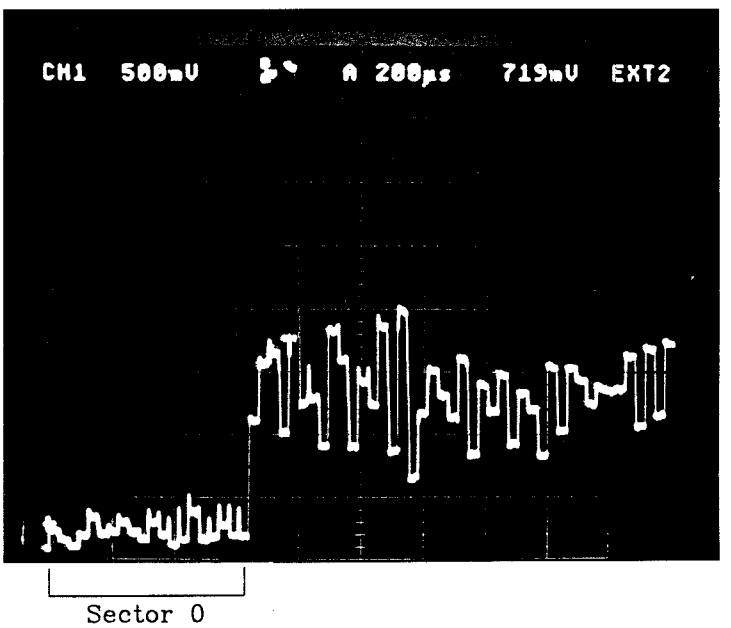


Figure 5.29 : Example of inner product signal obtained after S/H circuit.

Low level signal at left of figure represents system noise level.

vectors $\underline{y}$ as

$$|\underline{y}|^2 = (\underline{y} \cdot \underline{1}) \tag{5.8}$$

$$= (\underline{y} \cdot (\underline{x} + \overline{\underline{x}})) \tag{5.9}$$

$$= (\underline{y} \cdot \underline{x} + \underline{y} \cdot \overline{\underline{x}}), \tag{5.10}$$

so that

$$|\underline{y} - \underline{x}|^2 = |\underline{x}|^2 + \underline{y} \cdot \overline{\underline{x}} - \underline{y} \cdot \underline{x}. \tag{5.11}$$

Once again, $|\underline{x_i}|^2$ for $i = 1, ..., 650$ is stored in normalization memory and read out during the postprocessing stage.

CLASS	SIMULATION	EXPERIMENT	MODEL
0	28	28	28
1	27	19	17
2	24	28	24
3	23	9	15
4	28	24	26
5	24	26	21
6	22	23	20
7	24	26	17
8	23	21	26
9	26	9	24
	83%	71%	73%

Table 5.1 : Recognition results for 0-9 optical KNN network.

The results of classifying the 300 element testing set using the optical system described here, are given in Table 5.1. Entries in the table indicate the number of correct classifications made, out of 30, for each of the 10 classes 0-9 for the

k=5-NN algorithm. The last row in the table indicates the overall classification rate for the experimental system as compared with an ideal simulation of the 5NN algorithm as well as a computer model of the optical system. Regarding the experimental column of Table 5.1, we see that the majority of errors occurred for classes 3 and 9. This is consistent with the observation that in simulation, threes are often confused with fives and eights while nines are often confused with fours. These are the primary misclassification trends seen in the optical experiments as well. These sorts of errors imply that a heirarchical or tree-like decision procedure might be used to improve the recognition rate by separating difficult class dichotomies and treating them independently. The simulation column of Table 5.1 is representative of many simulation runs that examined variations in recognition performance with both k and library size T. It was observed that for $1 \leq k \leq 5$ and for $650 \leq T \leq 700$, the recognition rate Σ, did not vary substantially (i.e., $81.3\% \leq \Sigma \leq 84.7\%$). This relative insensitivity with k was observed in the experimental system as well where the recognition rate varied from 67% to 71% for k less than 7. The model column of Table 5.1 was obtained from a detailed simulation of the optical system which incorporates the various implementational error sources as measured from the experimental apparatus. The error sources that were included in the model are :

(1) *Beam nonuniformity* which accounts for the gaussian nature of the SLM illumination and was seen to result in significantly reduced beam intensity at the outermost input pixels. The illuminating intensity pattern was measured and found to be down to 42% at pixel 1 and 56% at pixel 100.

(2) *Electrical noise* in the amplification and S/H circuits was measured to be

200mV p-p out of 3.8V peak which corresponds to a SNR before the A/D converter of ≈32 after averaging.

(3) *Quantization noise* resulting from the 8 bit A/D or 19.5mV per level quantization. This effect reduces the sampled SNR to ≈29.

(4) *Sampling phase error* in the S/H circuit will result in sampling the inner product off peak. This error was measured to be 2μs and was attributable to motor speed nonuniformity. This effect was not critical however since the inner product peak width is equal to the vector width in pixels divided by the pixel rate or 6μs.

(5) *SLM (LCTV) contrast* was measured at the optical disk to be ≤20:1.

According to our model, finite SLM contrast was by far the dominant error source in this experiment. Using the model, we have plotted the expected recognition rate vs. contrast for fixed values of the other error sources as given above. This plot is shown in Figure 5.30. It can be seen from the figure that for a value of contrast less than 50:1 as in our system, the recognition rate is strongly dependent on the contrast. Only when the contrast becomes greater than 50:1 do we see that the asymptotic rate of 81% is achieved. This is less than the 83% ideal simulation rate owing to the presence of the other error sources (items 1-4 above).

Although the experimental recognition rate of 71% does not agree well with the simulation result, it does agree quite well with our model. The plot shown in Figure 5.30 seems to imply that by replacing the input SLM and imaging system so as to achieve >50:1 contrast at the disk, the performance of the experimental

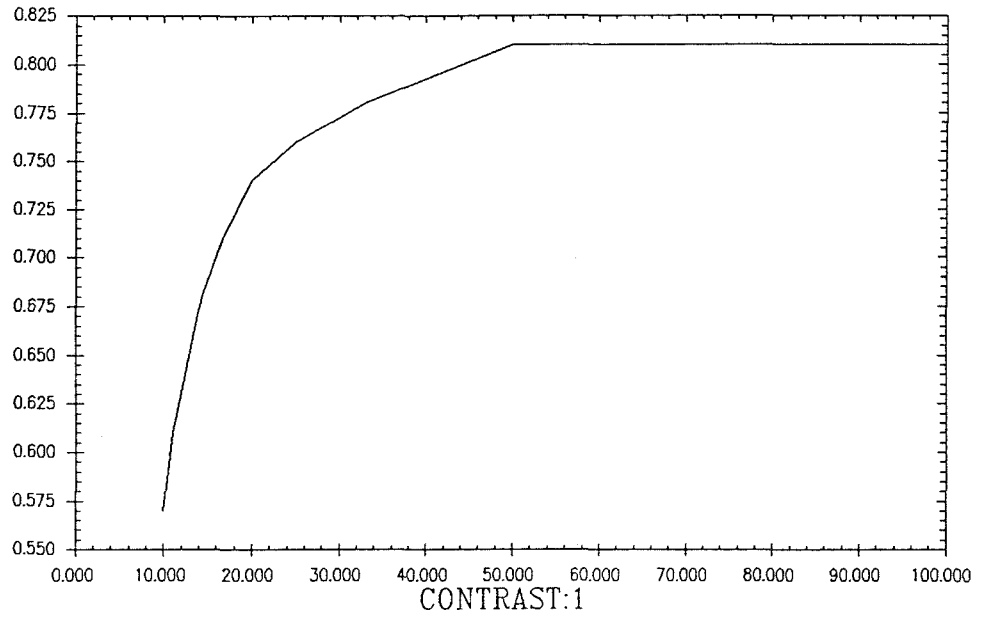


Figure 5.30 : Plot of classification rate vs. contrast as predicted using the model described in the text.

system will approach the noise limited performance of 81%.

We have replaced the input transparencies used in this experiment with an Epson Liquid Crystal Television (LCTV) which exhibits a measured contrast of 90:1. A photograph of the new optical system is shown in Figure 5.31. According to the previous discussion, this modification should cause the classification rate of the optical system to approach the noise limited rate of 81%. The optical system was once again tested using the 300 character testing set and the classification results from the modified system are given in Table 5.2. As we expected, the result of improved SLM contrast is to improve the optical system recognition rate from 71% to 79% in reasonable agreement with the performance predicted by our model.

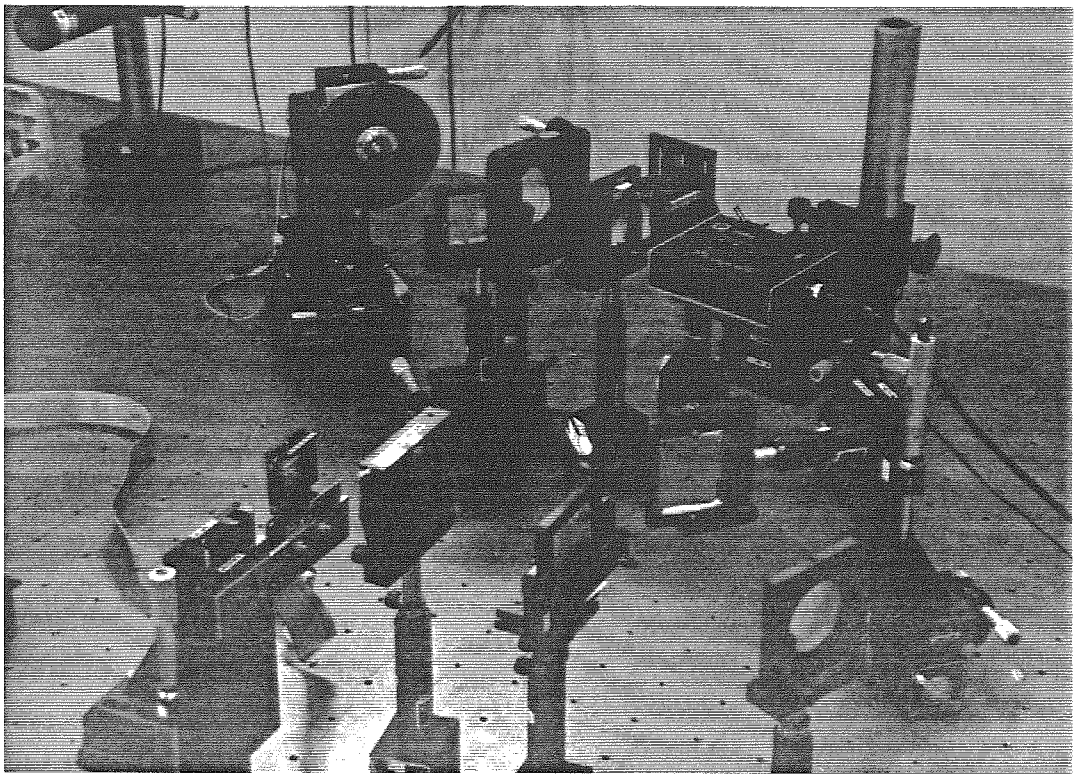


Figure 5.31 Photograph of high contrast optical KNN classifier utilizing Epson LCTV as an input SLM.

Class	Experiment	Simulation
0	29	28
1	25	27
2	29	24
3	22	23
4	24	28
5	18	24
6	15	22
7	23	24
8	26	23
9	26	26
Total	237	249
	79%	83%

Table 5.2 : Recognition results for high contrast 0-9 optical KNN system.

5.6 RBF Character Recognition Network

The final optical pattern recognition system we will demonstrate is based on the radial basis function approach described earlier. The actual architecture we have implemented using the optical disk is represented schematically in Figure 5.32. A preprocessed 100 bit binary vector $\underline{x}$, is presented to the system as before and the first layer of RBF units compute the RBF projections $y_i = exp(-|\underline{x} - \underline{t^i}|^2/\sigma_i^2)$. We will once again use the 10 class handwritten character recognition problem as a test for this system and have chosen to use as RBF centers $\{\underline{t^i}\}$, the 650 reference images used in the KNN system described in the previous section. This choice of centers was primarily motivated by the fact that the required vectors had already been recorded on the optical disk and that the SONY disks are somewhat scarce. In addition, this approach will allow us to make a purely architectural comparison between the 1-layer KNN network and the 2-layer RBF approach. After the RBF projections are calculated in the middle layer, this 650-dimensional intermediate representation is transformed using the interconnection matrix $\underline{\underline{W}}$ to arrive at a 10-dimensional output representation as shown. Each output neuron corresponds to one of the classes 0-9 and a winner-take-all network then computes the final classification. Since we have chosen to use the 650 templates from the previous system as RBF centers, the only learning required for the first layer of this network is for the widths $\{\sigma_i\}$. The second layer of course must be trained to perform the desired classification on the intermediate RBF representations.

There are many potential training algorithms for $\{\sigma_i\}$ and $\underline{\underline{W}}$. We began to investigate several of these and during these investigations the most successful

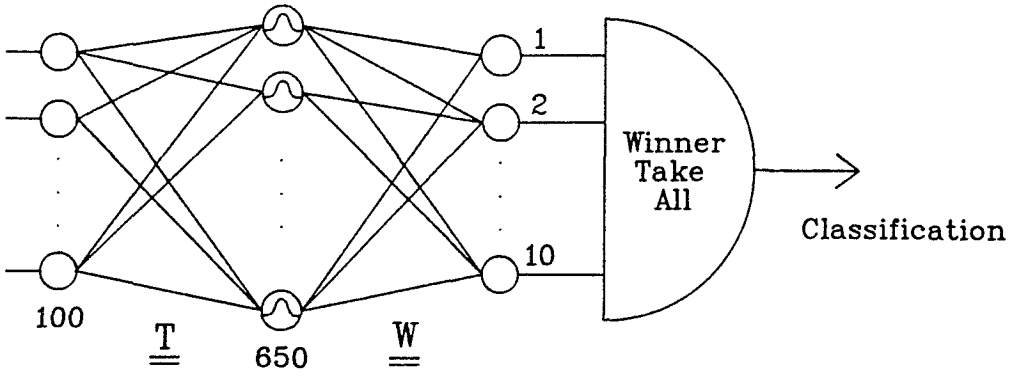


Figure 5.32 : Multilayer RBF network implemented using optical disk-based first layer.

algorithm was evolved in a stepwise fashion. Consider the simple case of setting the RBF widths to be constant

$$\sigma_i = \sigma \quad \forall \quad i = 1, 650. \tag{5.12}$$

We may now specify the second layer weights using the simple sum rule as

$$\underline{w^i} = \sum_{j=1}^{650} \beta \underline{y^j} \quad i = 1, 10, \tag{5.13}$$

where β is given by

$$\beta = \begin{cases} 1 & \text{if } \underline{x^j} \in \Omega_i \\ -1 & \text{otherwise} , \end{cases} \tag{5.14}$$

and $\underline{x^j}$ is the jth element of the training set, $\underline{w^i}$ is the weight vector associated with the ith output neuron $1 \leq i \leq 10$ and Ω_i is one of the classifications 0-9,

Width σ	Training Set (Correct out of 650)	Testing Set (Corrrect out of 300)
0.1	650	38
1.0	650	48
2.0	649	73
3.0	628	115
5.0	368	157

Table 5.3 : RBF recognition results obtained using fixed widths and the simple sum algorithm for second layer weights.

and $\underline{y^j}$ is the 650-dimensional RBF representation associated with the jth training vector $\underline{x^j}$. This algorithm defines $\underline{w^i}$ simply as the sum of same class vectors minus the sum of other class vectors for each class i. Using these specifications for $\{\sigma_i\}$ and $\underline{\underline{W}}$, rather poor results were obtained as shown in Table 5.3. In the table we show the performance of the RBF network on both the training and testing sets for several values of σ. Notice that as the performance on the testing set improves, classification rate for the training set declines. We were unable to find a network that exhibited good performance on both the training and testing sets using this algorithm. A useful adjustment to the above algorithm is to replace the simple sum with a binary address algorithm for specification of the output weights. This algorithm does not require specific knowledge of the intermediate representations generated during training, it only requires knowledge of the class assignment of each of the 650 RBF centers. This reduces second layer computation time and improves network performance. The binary address algorithm defines $\underline{\underline{W}}$ as

$$w_{ij} = \begin{cases} 1 & \text{if } \underline{t^j} \in \Omega_i \\ -1 & \text{otherwise.} \end{cases} \tag{5.15}$$

Width σ	Training Set (Correct out of 650)	Testing Set (Corrrect out of 300)
0.25	650	247
0.5	650	246
1.0	650	245
2.0	650	222
3.0	650	160

Table 5.4 : RBF recognition results obtained using fixed widths and the binary address algorithm for second layer weights.

The network performance resulting from this training procedure is shown in Table 5.4. The best such network achieved a recognition rate of 82% which is comparable to the results obtained using the KNN approach. In both of the above algorithms, we found that broader basis functions resulted in more training set misclassifications. This is understandable behavior since broader basis functions correspond to greater overlap between centers from different classes which can result in confusion of class assignment for the centers themselves. A simple modification of the above algorithm which incorporates detailed knowledge of the training representations is to follow the above procedure with the perceptron learning algorithm for the output weights. The results obtained in three runs using this approach are shown in Table 5.5. This approach results in the best fixed width performance of 87.7% on the testing set with 100% recognition of the training set.

From the above results we may conclude that improved recognition performance will require variable RBF widths. The most successful variable width

Width σ	Training Set (Correct out of 650)	Testing Set (Corrrect out of 300)
5.0	650	192
7.0	650	244
9.0	650	263

Table 5.5 : RBF recognition results obtained by following the previous algorithm with the perceptron rule to improve second layer performance.

$\tilde{\sigma}$	Training Set (Correct out of 650)	Testing Set (Corrrect out of 300)
0.5	650	226
0.7	650	257
1.0	650	266
1.2	650	267

Table 5.6 : RBF recognition results obtained using the previous algorithm with variable centers.

approach we found was to make σ_i proportional to the distance between template $\underline{t^i}$ and its nearest neighbor. That is

$$\sigma_i = \tilde{\sigma} \min_j |\underline{t^j} - \underline{t^i}| \quad j \neq i. \tag{5.16}$$

Using this approach together with the previous algorithm for training the second layer, we have a best RBF performance of 89% as shown in Table 5.6. Although the trend with increasing $\tilde{\sigma}$ is an improvement in network performance, we found that in general, the broader the basis functions the longer the perceptron algorithm will take to converge. For this reason, Table 5.6 does not contain any entries for $\tilde{\sigma} > 1.2$. Note that the best RBF network performance of 89% is better that the best KNN system performance of 83% using the same template library. This performance can also be compared with a single layer of 10 neurons, each trained with the perceptron using the 650 image reference library. The recognition rate in this case is 75% on the 300 element testing set. In general, we would expect an improvement in RBF network performance with variable centers where both the *number* and *location* of those centers are optimized. This case was not studied here as we are primarily interested in the performance of the optical implementation.

In Figure 5.33 we show the RBF widths computed using the best case procedure described above. Each row in the figure represents the widths associated with centers in a *single* class. There are therefore 65 blocks per row and 10 rows in Figure 5.33. Each small rectangular block in the figure is a grey scale coding of the width associated with the corresponding template where dark represents zero width. It is interesting that the second row in Figure 5.33, corresponding

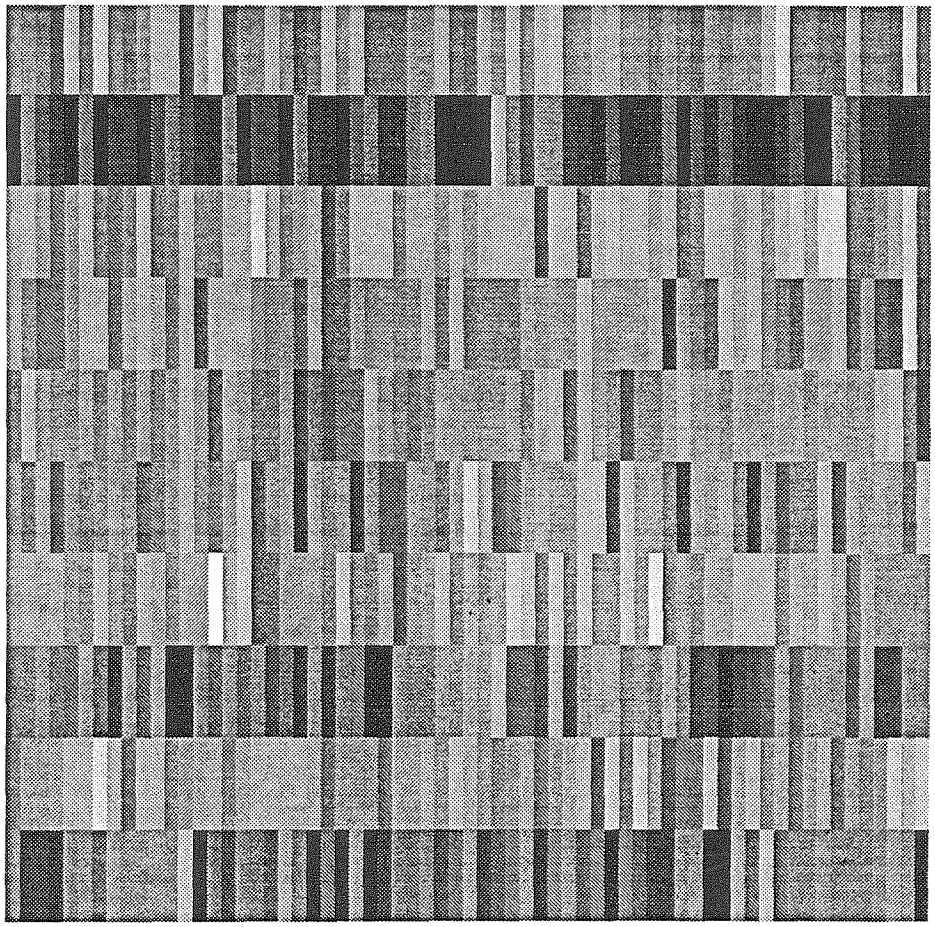


Figure 5.33 : RBF widths associated with *best* RBF network.

to handwritten ones, is particularly dark, indicating that these vectors tend to be clustered or, more generally, tend to be located close to other vectors. Also in Figure 5.33 we can see that the width associated with one particular template representing a handwritten six (template 15 in row 7), is quite broad, indicating that this vector is basically isolated in the input space. Using the same display format as in Figure 5.33, Figure 5.34 shows the second layer weights generated for the *best* RBF network found above. The single bright row in each weight vector indicates that the weight vector is basically tuned to intermediate representations from only one class. The 89% recognition rate achieved using this RBF network corresponds to 33 misclassifications out of 300. An example of one misclassification from each class along with the intermediate representation associated with the misclassified input is shown in Figure 5.35. Next to each input character in the figure is the erroneous class to which it was assigned. Most of these misclassifications, although they would most likely not be made by a human classifier, can be easily understood by inspection of Figure 5.35.

The optical system used to implement the RBF classifier is identical to that used in the KNN implementation since the first layer of the present system is also based on Euclidean distances. The optical disk-based inner product calculations are once again collected by a postprocessing system which now computes the required gaussian weighting and simulates the output layer where a classification is made. The postprocessing steps were carried out in software for our experiments. Using the low contrast version of the previous KNN optical system, the classification rate for the optical RBF system was 67.7%, indicating that this algorithm is more sensitive to the various optical system imperfections than was

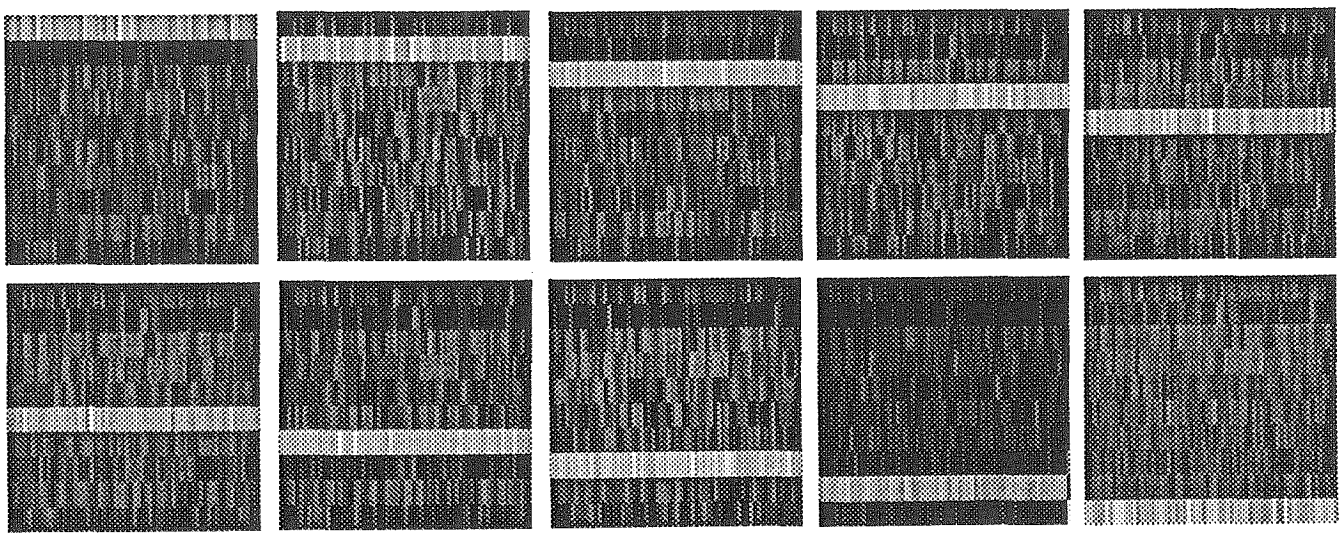


Figure 5.34 : Second layer weights from best RBF network.

Each large block represents the weight vector for a single output neuron.

Weight vectors in top row correspond to neurons 0-4 with 5-9 on the bottom.

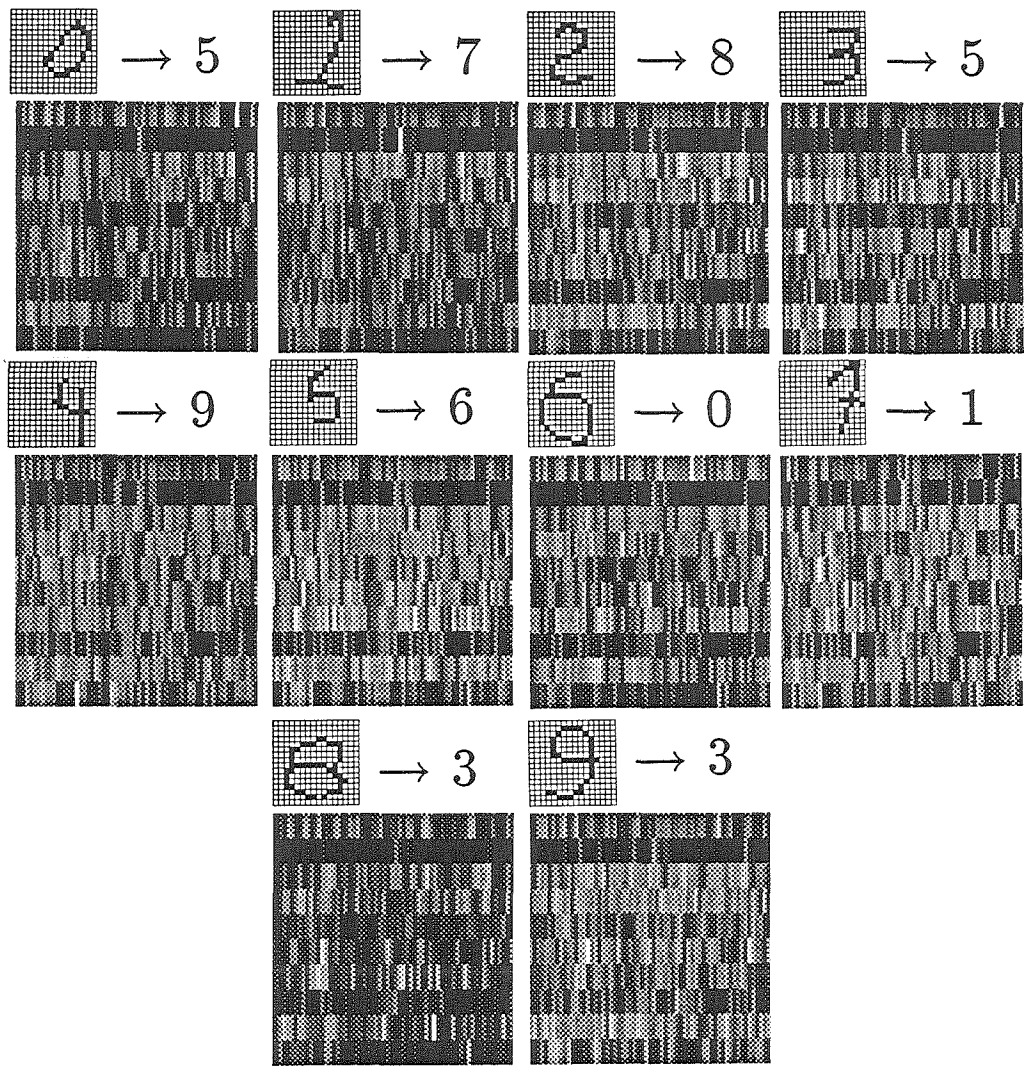


Figure 5.35 : Example of one misclassification per class.

Original 16×16 input vector → class to which it was assigned.

Below each pair is the corresponding intermediate representation.

CLASS	SIMULATION	EXPERIMENT	MODEL
0	29	24	30
1	29	18	26
2	28	26	24
3	27	7	19
4	23	20	12
5	25	24	27
6	24	23	17
7	28	27	26
8	25	21	12
9	29	13	21
	89%	68%	71%

Table 5.7 : Recognition results for optical RBF network.

the KNN approach. A comparison between the performances of the low contrast optical system, a simulation and an experimental model is shown in Table 5.7.

A model once again provides a technique by which we can examine the optical system performance as a function of the various implementational error sources. In the present case it was not possible to isolate a single critical variable responsible for the degradation in performance as all variables appear to play a role. In Figure 5.36 we show the recognition rate vs. contrast for noise levels above, below and equal to the measured noise levels in the optical system. We notice a sharp decrease in performance below a contrast level of 10:1 which is an improvement in contrast tolerance over the KNN approach; however, for a contrast of 20:1 and a noise level slightly higher than measured, we have a recognition rate of only 72% which is close to the experimental rate. A slightly more critical variable for this system is the gaussian profile width. A plot of recognition rate vs. log of the $1/e^2$ gaussian profile width is shown in Figure 5.37 and points out the

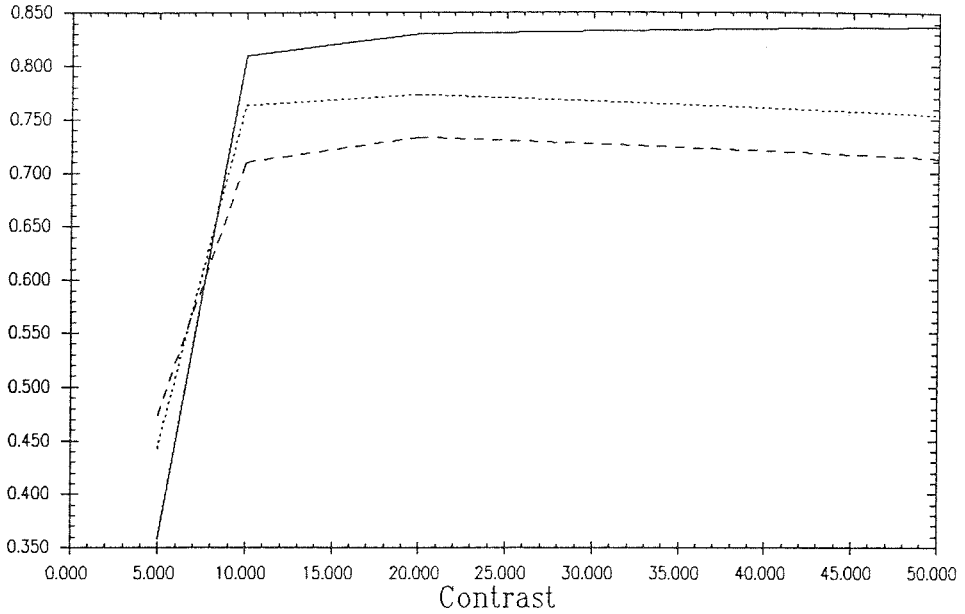


Figure 5.36 : RBF recognition rate vs. contrast for three different noise levels as predicted using model.

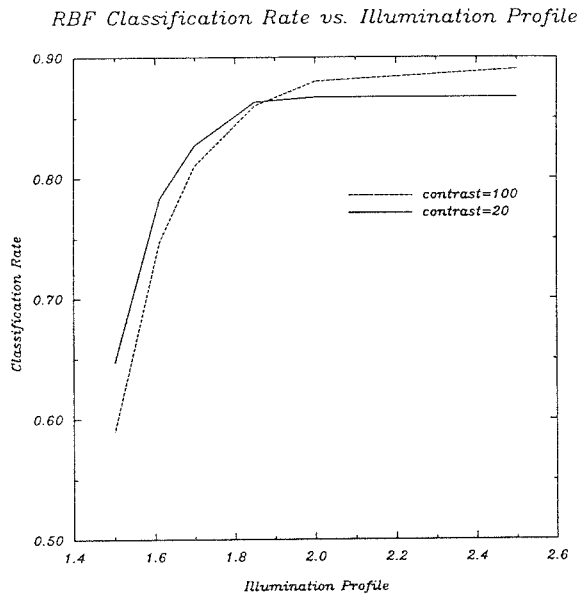


Figure 5.37 : RBF recognition rate vs. log of the $1/e^2$ gaussian illumination profile width (from model).

importance of illumination uniformity for the optical RBF classifier. Referring once again to Figure 5.37, if the measured decline in recognition performance was due entirely to beam profile nonuniformity then the required beam width parameter would be 1.52. A plot of the corresponding profile together with the experimentally measured profile and the profile fit used in the model is given in Figure 5.38. These profiles match quite well, indicating that an improvement in beam uniformity together with an improvement in contrast might increase the optical RBF classification rate to near 90%.

Using the high contrast version of the optical KNN system which incorporates a LCTV input SLM, we re-tested the optical RBF network performance. During the analysis of this new data it was discovered that there existed significant variations in written pixel reflectivity over the disk both in the along track and across track directions. This effect is demonstrated in Figure 5.39 where we have plotted normalization data as a function of template number for the *all one* input vector. The data we have plotted is the sum of $(\underline{1} \cdot \underline{t^j})$ and $(\underline{1} \cdot \overline{\underline{t^j}})$ as measured using the optical system. This graph should be constant for all j since $|\underline{t^j} + \overline{\underline{t^j}}|^2 = 100$. Although disk wobble could be responsible for *some* of the variation seen in the figure, the two plots shown were obtained from different optical alignment conditions so that the observed effect (a prominent increase in signal between templates 450 and 550) most likely results from nonuniformities in reflectivity. A direct measurement of the disk reflectivity confirmed that there is as much as 10% variation in reflectivity between opposing portions of the disk. In addition to variations in media reflectivity, errors that occur during the writing process result in reflectivity errors in the written data. This effect

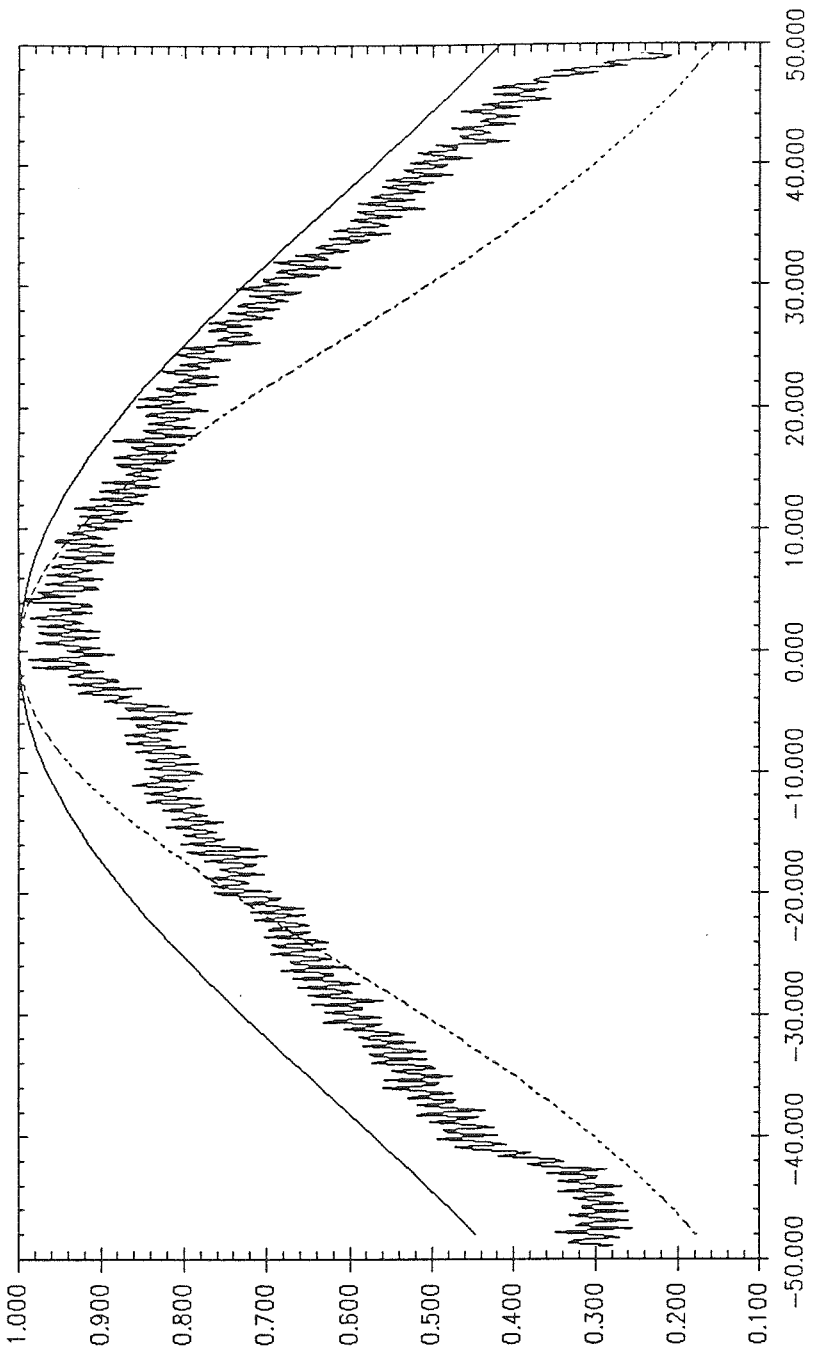


Figure 5.38 : Comparison between measured illumination profile and the gaussian functions (solid) used in the model and (dashed) required to account for recognition performance.

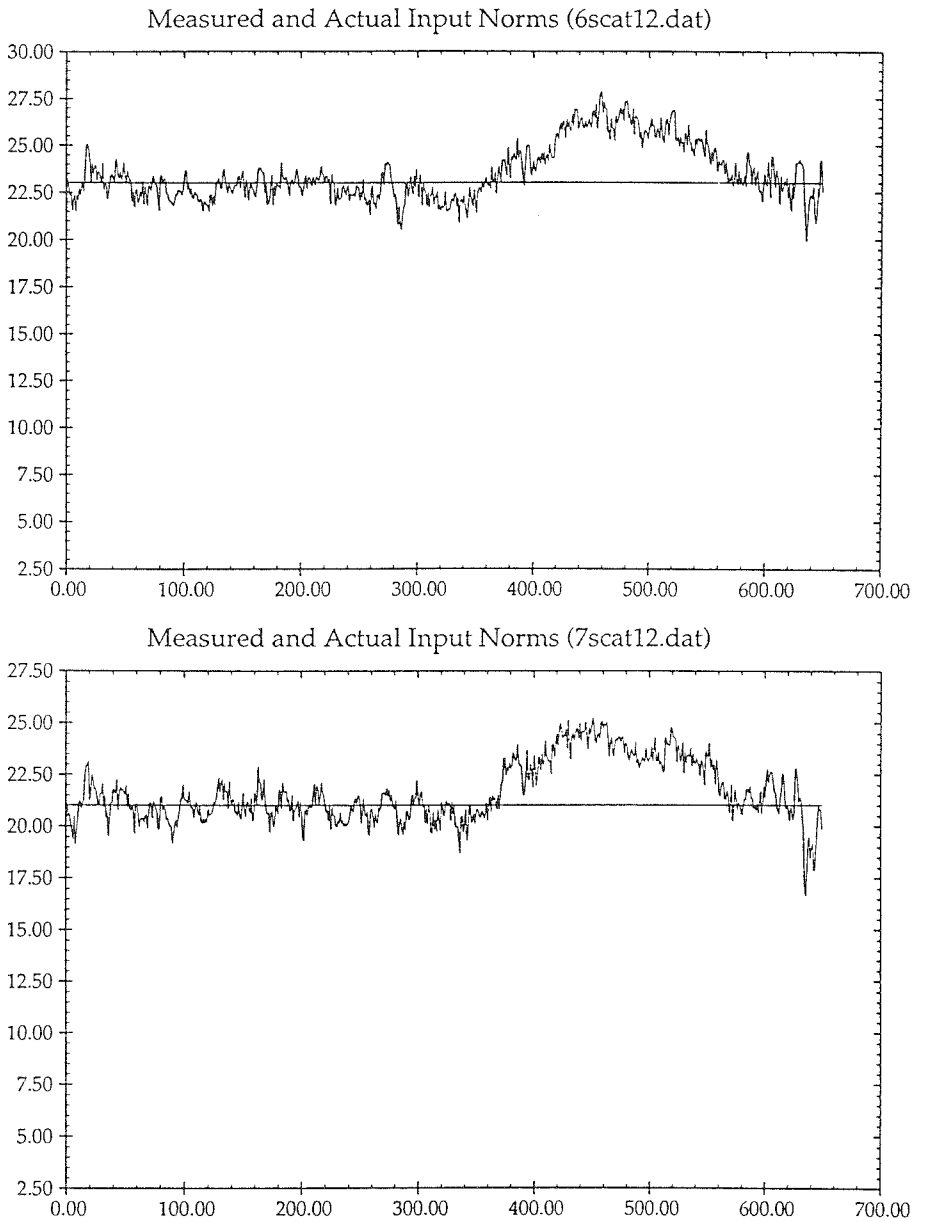


Figure 5.39 : Experimental normalization vs. template number for two different alignment configurations.

was verified by observing such errors on the disk under a microscope. These nonuniformities in reflectivity may be partially compensated through the use of a model for the optically computed inner products which incorporates calibration of these variations. We can estimate no more parameters from the experimental data than the number of known quantities we measure, which in our case is 1300 for normalization data points. Thus, we will assume that the optical system computes inner products as

$$(\underline{x} \cdot \underline{t^j})_{opt} = \beta_j(\underline{x} \cdot \underline{t^j}) + \gamma_j, \tag{5.17}$$

where $\{\beta_j\}$ and $\{\gamma_j\}$ are 1300 unknown constants that quantify the inaccuracy of the optical disk and the subscript *opt* indicates a measurement from the optical system. We can solve for $\{\beta_j, \gamma_j, j = 1, 650\}$ using the normalization data $\{(\underline{1} \cdot \underline{t^j})_{opt}\}$ and $\{(\underline{1} \cdot \overline{\underline{t^j}})_{opt}\}$. If we assume that the disk characteristics do not vary much between adjacent servo areas then we can write

$$(\underline{1} \cdot \underline{t^j})_{opt} = \beta_j |\underline{t^j}|^2 + \gamma_j \tag{5.18}$$

and

$$(\underline{1} \cdot \overline{\underline{t^j}})_{opt} = \beta_j |\overline{\underline{t^j}}|^2 + \gamma_j. \tag{5.19}$$

Subtracting these two measurements yields

$$(\underline{1} \cdot \underline{t^j})_{opt} - (\underline{1} \cdot \overline{\underline{t^j}})_{opt} = \beta_j(|\underline{t^j}|^2 - |\overline{\underline{t^j}}|^2), \tag{5.20}$$

so that the $\{\beta_j\}$ may be derived from the stored normalization data as

$$\beta_j = \frac{(\underline{1} \cdot \underline{t^j})_{opt} - (\underline{1} \cdot \overline{\underline{t^j}})_{opt}}{2|\underline{t^j}|^2 - 100}. \tag{5.21}$$

Using this equation for β_j, γ_j can be derived as

$$\gamma_j = (\underline{1} \cdot \underline{t^j})_{opt} - \beta_j |\underline{t^j}|^2. \tag{5.22}$$

Having derived a relationship between the experimentally measured inner products and the desired values for $\underline{x} \cdot \underline{t^j}$, we can now write an expression for the desired distances as a function of the experimentally measured values as

$$|\underline{x} - \underline{t^j}|^2 = |\underline{t^j}|^2 + |\underline{x}|^2 - 2\underline{x} \cdot \underline{t^j} \tag{5.23}$$

$$= \frac{(\underline{1} \cdot \underline{t^j})_{opt} - \gamma_j}{\beta_j} + \frac{[(\underline{1} \cdot \overline{\underline{t^j}})_{opt} - (\underline{1} \cdot \underline{t^j})_{opt}]|\underline{x}|^2}{[(\underline{1} \cdot \overline{\underline{t^j}})_{opt} + (\underline{1} \cdot \underline{t^j})_{opt} - 2\alpha_j \gamma_j]} \tag{5.24}$$

where the factor α_j is used to account for the fact that classification data and normalization data are typically taken under slightly different experimental conditions. We assume that this effect results in inner product measurements that differ from normalization measurements by a multiplicative factor. α_j is given by

$$\alpha_j = \frac{(\underline{1} \cdot \overline{\underline{t^j}})_{opt} + (\underline{1} \cdot \underline{t^j})_{opt}}{\beta_j |\underline{x}|^2 + 2\gamma_j}. \tag{5.25}$$

In addition to modeling the variations in disk reflectivity and improving the contrast of the input SLM, the illumination profile was also improved. A more uniform illumination profile was generated by careful design of the collimating optics and input aperture. The resulting beam profile parameter was measured as 1.8 corresponding to a recognition rate of 86% as predicted using the plot of Figure 5.37. The results of the modified optical RBF classifier are given in Table 5.8. The overall classification rate has been improved from the previous 68%

Class	Experiment	Simulation
0	25	29
1	28	29
2	28	28
3	25	27
4	28	23
5	20	25
6	25	24
7	23	28
8	24	25
9	22	29
Total	248	267
	83%	89%

Table 5.8 : Recognition results for high contrast optical RBF network.

in the uncorrected system, to 83%, which is in good agreement with the profile limited performance of 86%.

The 6% difference seen in Table 5.8 is one measure of the inaccuracies present in the optical computation. Another method by which we can compute the cumulative effect of the many optical system imperfections is to consider the system performance *prior* to the recognition algorithm. A calculation of the error present in the distance computations themselves will provide another measure of the accuracy of the optical system. In Figure 5.40 we show the 650 distances computed for a single input image (a handwritten 3) using both the ideal computer simulation (b) and the optical system (a). From the figure we see that there is a substantial variation between these two plots which may be quantified by computing the RMS distance error over the entire testing set as

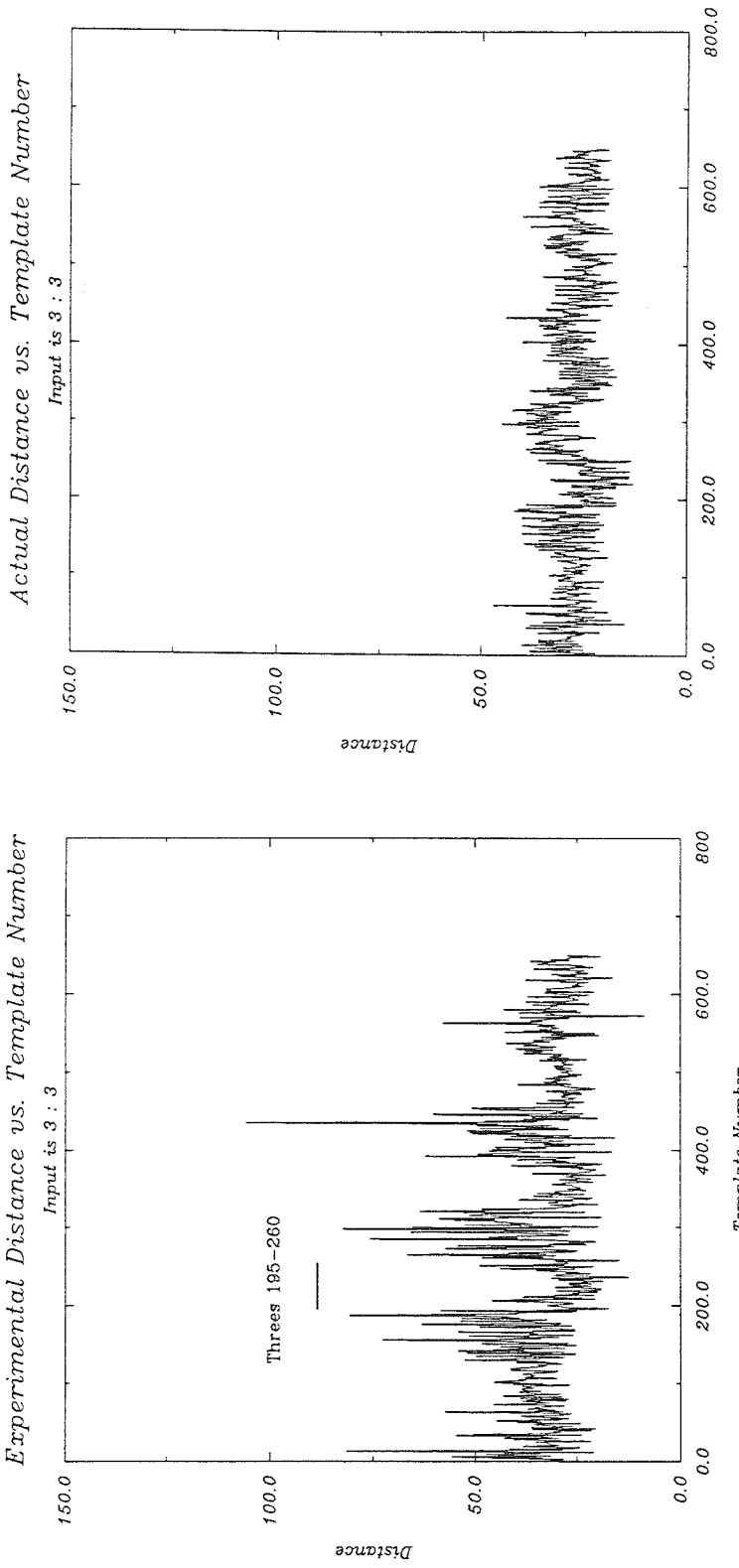


Figure 5.40 : Distance vs. template number for a single input character.

(a) Experimentally computed distances.

(b) Actual distances.

$$\Delta D_{RMS} = \frac{\sqrt{\frac{1}{M}\sum_{i=1}^{M}(d_i^{Opt} - d_i^{Sim})^2}}{\frac{1}{M}\sum_{i=1}^{M} d_i^{Sim}}, \tag{5.26}$$

where d_i^{Sim} and d_i^{Opt} are the Euclidean distances between the 300 input images and the 650 templates calculated from simulation and the optical system respectively. There are M=195,000 such measurements in our case. For the results presented here, the RMS distance error was found to be ΔD_{RMS} =28.5%. Although this error is quite large, the recognition rate obtained using the optical system is in satisfactory agreement with the expected rate, attesting to the robustness of the RBF approach.

The present system as well as the previous shift variant KNN system is capable of operating at very high processing rates. If the upper limit on disk storage is approached wherein 10^6 vectors each of dimension 10^4 are stored as radial lines over the entire disk, then the inner product computation rate for a 100Hz disk rotation rate becomes 10^8 inner products per second. This rate of comparison corresponds to 10^{12} binary operations per second. This capability would clearly be very difficult to duplicate with existing electronic technology.

5.7 Optoelectronic Multilayer Pattern Recognition

In this chapter we have described three architectures based on optical disk technology, that realize optical handwritten character recognition to various degrees. A theme common to each of these systems is an electronic postprocessing stage. These systems might therefore be more appropriately labeled *optoelectronic* handwritten character recognition systems. This approach represents a useful implementation philosophy for multilayer pattern recognition in which high speed, parallel optical preprocessing is used for feature extraction or dimensionality reduction and high accuracy, serial electronics is utilized as a flexible, algorithmic-based postprocessor. In this way we match the parallelism, and in the present case the high storage capacity, of optical media and systems with the maturity and the flexibility of VLSI technology, thereby evolving systems that may be useful in the solution of *real* image recognition and understanding problems.

As an example of this type of hybrid pattern recognition system, consider the diagram shown in Figure 5.41. The lefthand portion of the figure depicts a correlation-based feature extraction subsystem that might project some unknown input image onto feature dimensions such as oriented lines, line stops, angles, ..., etc. These features are stored on disk 1 and can be effectively scanned using the systems that we have demonstrated in this thesis. The peaks of the resulting feature maps, as well as their locations as derived using chip 1, can now be operated on using any number of algorithms, including KNN and RBF as described here, Parzen windows, associative memories or a VLSI-based multilayer neural network as shown in the figure. The subsystem comprising disk 2 and chip

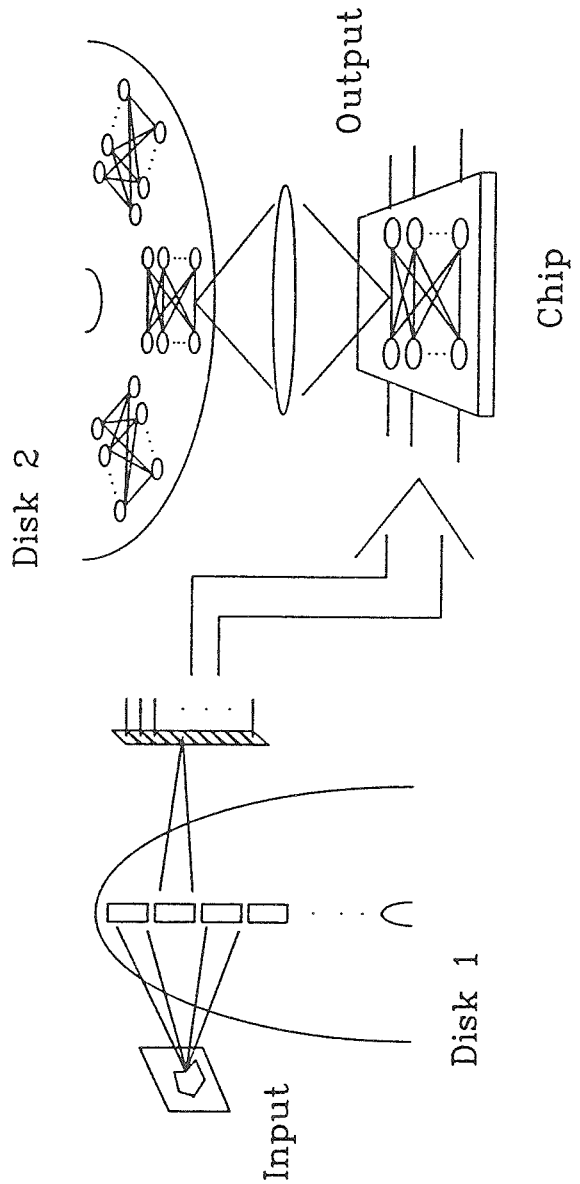


Figure 5.41 : Hybrid optoelectronic pattern recognition system.

2 has been demonstrated and the details of that work can be found in references [130-132]. The basic notion here is that an electronic input vector describing the projections derived using subsystem 1 is presented to chip 2 as shown, on which a single feedforward layer of a network is implemented. The weights of this layer are loaded from disk 2 via the third dimension, using photodetectors fabricated on the surface of chip 2. After the chip has been clocked and one layer has been implemented, the resulting representation is latched and fed back to the input. Disk 2 may now rotate, allowing new weights to be loaded onto the chip whereby on clocking, another layer may be implemented. A severe limitation to the use of VLSI technology in neural network implementations is the wire routing problem that arises in large networks.[133–135] The optical disk-based preprocessor in this system provides the dimensionality reduction required to realize such a VLSI based image recognition system.

6. Conclusions and Future Work

In this thesis we have re-introduced the conventional optical disk as a parallel access memory. In addition to reviewing various mechanisms for serial optical data storage, we have presented systems that use parallel retrieval from optical storage devices to advantage in the context of optical pattern recognition. The implications of parallel access are manyfold. Parallel access provides the potential for extremely high data rate transfer from mass memory to RAM or between mass memory and individual processors. The optical nature of this transfer provides the potential for simultaneous multiprocessor access to shared memory. In the more general case of RWE parallel access, the disk can represent in addition to mass storage, on the order of 10^4 parallel, moderate bandwidth communication channels for parallel computing. Specific examples of such systems will be discussed in more detail below. In this thesis we have seen how parallel access can provide a simple mechanism for searching through a library of data. In the systems we have discussed, the optical disk acts as both a memory device and an SLM.

We have seen how novel pattern recognition systems may be designed to take advantage of parallel access to optical disks. The basic building block for many of these systems, as for most optical pattern recognition systems, is the image correlator. The photorefractive correlator discussed at the end of Chapter 4 was seen to be the most attractive disk-based image correlator owing to its high speed and superior efficiency. The optical approaches we develop realize advantages over their *electronic* counterparts in terms of speed and capacity while also improving on many *optical* architectures owing to their utilization of mature

technologies such as commercially available optical disks and SLMs. With any analog computing paradigm however, limited accuracy is a potential problem. Although these accuracy issues remain a problem in analog *optical* computing systems, we found that some pattern recognition algorithms are robust enough to tolerate rather severe optical system inaccuracies. Furthermore, optical disk imperfections such as wobble, disk center offset, and nonuniform media characteristics including defects, can be compensated using either realtime adaptive optical systems or offline postprocessing based on characterization data.

Despite the advantages mentioned above, the disk-based systems we have described suffer from several potential limitations. The WORM nature of the media used in our work, although not *fundamental*, is indeed a limitation. As we have discussed in Chapter 3, many RWE mechanisms exist and show promise for future storage systems. An interesting area of future work is in adaptive parallel access systems based on RWE disks. Also, combining RWE technology with parallel write capabilities might open up new possibilities for *true* parallel access. The second limitation of these systems is the requirement for serial writing. The serial writing mechanism employed in commercial disk systems limits the use of such disks in adaptive systems. A delay of 20,000 revolutions is required before a single radial vector may be completely recorded, corresponding to roughly a 10 minute response time. The use of a parallel write mechanism would make such an operation more feasible by reducing the response time to roughly 1μs. Several commercial possibilities exist for parallel write. Laser diode arrays are routinely fabricated with more than 1000 individually controllable sources. Twenty such source arrays could be used to simultaneously address all 20,000

tracks on the optical disks we use. In order to reduce the complexity of driving each source with an independent electrical signal, optical addressing of these arrays is possible through external optical pumping. This type of system provides a mechanism through which an optical signal can be used to effect parallel write operations which may in turn be detected using an optical parallel readout scheme. Also, commercially available diode-pumped YAG lasers represent a compact, high-power source sufficient to realize parallel write systems using conventional WORM and RWE media. Alternatively, parallel write may become feasible in the near term through the use of alternate media such as photopolymer or nonlinear organic compounds.

We will now overview several potentially interesting disk applications that might take advantage of parallel write and RWE media. We will also discuss several new architectures for optical and optoelectronic pattern recognition based on optical disk style systems. The following discussion represents work in progress as well as interesting candidates for future work in the general area of optical computing and pattern recognition.

6.1 More Optical Disks

Recently, closed loop optical systems have been demonstrated that realize useful memory operations such as autoassociative and heteroassociative recall.[136,137] These systems are based on spatially multiplexed memories in a Vander Lugt correlation architecture and suffer from limited shift invariance owing to the requirement for Fourier plane sampling. In the spirit of these systems, we are interested here in considering an architecture in which images may be accurately recalled

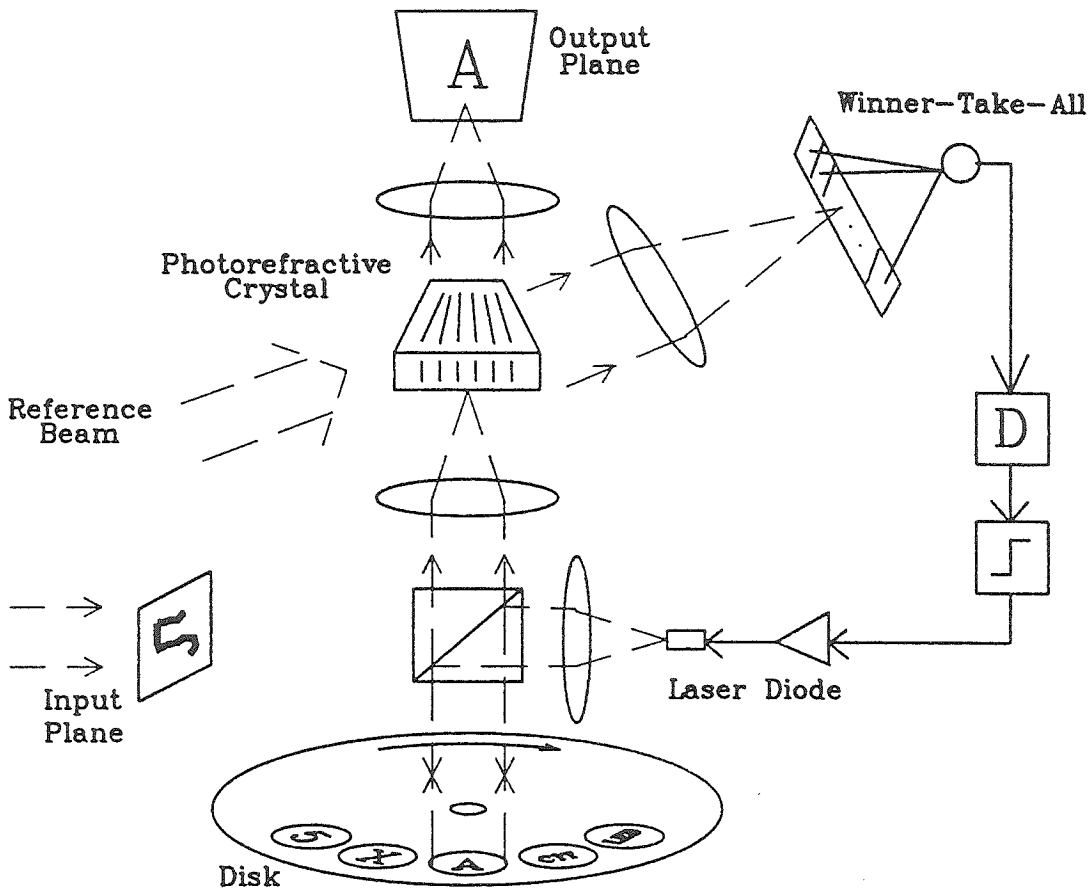


Figure 6.1 : Closed loop associative memory based on the photorefractive correlator described in chapter 2.

from noisy or partial inputs in a shift invariant fashion. An optical disk-based version of such a closed loop associative memory is shown in Figure 6.1. This system is based on the photorefractive correlator described in Chapter 4 and uses temporal multiplexing of stored memories to eliminate the need for Fourier plane sampling, thereby retrieving full 2D shift invariance. The system operation is as follows. As we described in Chapter 4, the unknown input image is illuminated in the input plane of the photorefractive correlator and the Fourier transform hologram of this pattern is formed in the crystal with the reference beam as shown. Following this recording stage the readout stage is initiated by illuminating the optical disk on which the stored memories reside. As the disk rotates, a 2D image correlation is computed between the input and all of the stored memories. As before, the 2D correlation patterns thus computed appear as sequences of 1D vertical slices in the output plane and therefore may be detected using a 1D array of N detectors. The 1D detector array is followed by a winner take all circuit to choose the largest correlation signal over the N output detectors at each instant. The largest signal from the output array is thresholded, delayed and amplified to produce a feedback signal to the disk illumination source. The disk is allowed to rotate a full 360° under continuous illumination, and thereafter, the disk illumination is pulsed according to the feedback signal as shown. The delay time is adjusted so that it corresponds to exactly one disk rotation. In this way, the memories that best match the input image as measured using the correlation peak will be obtained periodically in the readout plane. Fine tuning of the threshold will be required in order to obtain a *single* output memory; however, adaptive methods can also be used to realize the correct threshold. During the

first disk rotation for example, the threshold block in the figure can be used to record the value of the largest peak, and thereafter a value slightly smaller than this stored value can be used as a threshold. This will insure that only one memory is retrieved. In this scheme the total readout time is equal to the time for two full disk rotations or approximately 3 seconds. It is not entirely unreasonable however, to find the readout of a sequence of associations undesirable. In fact, the unknown input may comprise some linear combination of stored memories and the associative system described here could be used to retrieve the strongest components of such a composite image. Consider setting the threshold to some intermediate level. Multiple associations will occur for most inputs; however, due to the nature of the photorefractive retrieval, the recorded hologram will decay with continued readout. During successive rotations then, memories whose match with the input are smallest will drop below threshold, and thereafter will not be retrieved. Eventually, all but one of the stored memories will have dropped below threshold and a single output association will be obtained. The correct output too will eventually fall below threshold unless some refreshing mechanism is invoked. Such a mechanism may involve a simple reference beam shutter which controls reinforcement of the correct pattern in the crystal and is triggered by the existence of only one pulse in the feedback loop per disk rotation.

The system described above is fully shift invariant in the sense that any shifted version of a stored memory will be successfully recalled. Horizontally the photorefractive correlator is shift invariant owing to the disk rotation. The input image shift therefore is encoded on the feedback signal as a delay. This delay will result in the retrieval of a shifted memory from the disk. Vertically,

the winner-take-all circuit insures that the correlation *peak* is fedback regardless of its position on the output detector array. In this way, the correct memory is recalled regardless of the vertical position of the input. The vertical position of the input image may of course be retained by storing the location of the winning pixel in the winner-take-all circuit for later readout if the original position of the input object is required.

The above architecture is in the same vein as those we have been considering throughout the bulk of this thesis. This is an example of yet another system that may take advantage of WORM media and in which serial writing is tolerable since the memory storage phase takes place only once, off line. In order to see the potential usefulness of parallel write on RWE media, consider a simple system in which parallel across track writing is used to simulate a multichannel Bragg cell that can support as many channels as there are tracks on the disk. Readout can be achieved in this mode by illuminating along track for each channel. Each track therefore represents a single 1D SLM which can be used to modulate an incident beam in the same manner as are conventional AO devices. The resulting modulated beam is obtained diffracted along the direction of rotation and contains a doppler shift as with conventional Bragg cells. Each channel in this configuration is binary; however, 1-D area modulation can be used to associate a group of sources and a corresponding group of tracks with a single channel. Analog readout is obtained by illuminating such an array of tracks and collapsing the diffracted beam to a line that will then display the desired gray levels. Notice that in this scheme, only alignment between the writing source array and the readout beams is necessary. Since no mechanical alignment

between sources and the disk is required, the disk rotation rate can easily be increased above 100Hz, resulting in a modulator channel bandwidth > 50MHz. In principle, each channel will have an aperture equal to the disk rotation time, or a SBP= 10^6; however, utilization of all of this SBP will be difficult, owing to the circular nature of the tracks. A reasonable modulator SBP of 10^4 however, will only require readout illumination that covers 3.6° of a single track. This device is an interesting alternative to conventional AO cells. Using chalcogenide film-based RWE disks, diffraction efficiencies much greater than those achievable with MO media will be possible, making the light efficiency of systems using the disk based device comparable with those using conventional Bragg cells with the cost of such a disk-based 10^4 channel AOD being comparable to that of a single channel conventional modulator. In the same across track write configuration, a 1D *across* track readout scheme would correspond to realizing a large array (20,000 elements) of electrically addressable point modulators. As mentioned in the beginning of this chapter, these schemes can be made optically addressable through source-detector integration or through external optical pumping of the source laser diodes.

The above system is a simple example of how multiprocessor access to mass memory may be facilitated through the use of optical disk-based architectures that take advantage of parallel write and RWE media. In a more complex environment, the disk can provide a method for simultaneous multiprocessor reading and writing to portions of mass memory as well as a communication channel over which moderate bandwidth information can be transferred between processors. An example of such a complex multiprocessor environment operating within a

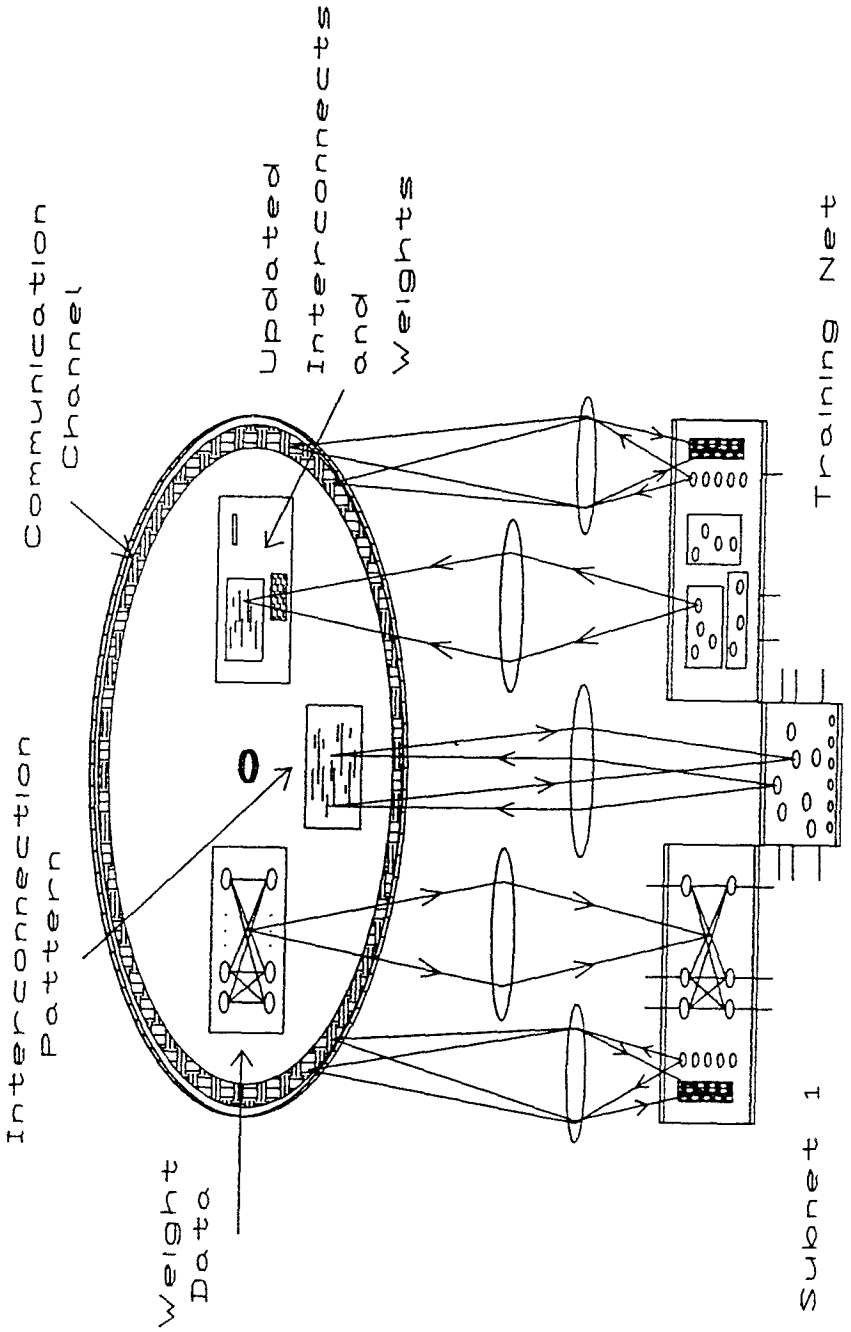


Figure 6.2 : Complex multiprocessor environment with storage and communication facilitated by the optical disk.

neural computing paradigm is depicted in Figure 6.2. Subnet 1 in the figure is an example of the sort of neural chip we introduced at the end of Chapter 5. As we described there, subnet 1 implements a single layer of a feedforward neural network.[130,131] The connection strengths for this network are loaded via the third dimension through photodetectors on the chip which encode the connections between neurons. In the scheme shown here these weights are stored on the disk, and are loaded onto the chip in parallel. Examples of VLSI circuits that implement relatively small 15×15 subnets of this type have been demonstrated. Subnet 2 is an alternate, optoelectronic neural network based on GaAs technology.[138] This chip consists of an array of neurons, each neuron comprising a photodetector, thresholding amplifier, and an optical source (LD or LED). Communication among these neurons is achieved optically and the interconnections are specified by holograms stored on the disk. In reality we envision each subnet operating in some combination of these two modes; however, use of the later mode should be minimized in order to make the best use of disk SBP since CGH encoding is costly. A special training network is shown in Figure 6.2 and may serve as a teacher or central control unit for this processor. The training net is used to update the weight data stored on the disk as well as the interconnection holograms thereby *programming* the future behavior of the system. In addition to providing interconnect and weight data *memory*, the disk may also serve as a moderate bandwidth *communication channel.* In the scheme envisioned here, each chip shown would be dedicated one or several communication tracks to which it may write data. The array of such communication tracks comprise the disk-based communication channel that may be interrogated by any processor in

the environment.

The operation of the system may be envisioned as follows. The disk is initialized with both weight data and interconnection patterns for each subnet. Control data that may be used to configure the subnets could also be pre-recorded on the disk communication tracks. On receiving external input then, the array of subnetworks can operate in unison to perform some predetermined task. During the computation, two communication modes are supported by the disk. High data rate exchange via disk-based holograms may operate at neuron response times and represents short range propagation of neuron activities. The other communication mode is the slow mode facilitated by the disk-based communication channel. Communications of this sort with the training net represents a facility through which learning may occur. The time scale of this adaptation is consistent with the notion that learning is a relatively slow phenomenon as compared with the neuron response time. Communication among subnets via the disk-based communication channel may represent long range interactions between clusters of neurons. Such interactions are believed to be partly responsible for attention and concept formation mechanisms in biological systems.[139] Once again, the time scale of these phenomena are believed to be slow and therefore may indeed be facilitated by this disk-based scheme. This scheme may also facilitate the representation of *global* variables such as the state of the intercellular fluid in biological systems. In this way, such variables may be broadcast with appropriate latencies so that finite global resources may affect subnet behavior.

Although the above notions are rather heuristic and are specific to a neural computing paradigm, the capabilities facilitated by the type of system depicted

in figure 6.2 are intriguing from the perspective of parallel computing in general. In order to understand the potential of systems similar to those outlined above, an investigation of the requirements for the interaction between parallel mass memory and processors is necessary. A deeper understanding of these relationships in the context of specific classes of problems will represent a major step in the direction of realizing useful parallel computational paradigms.

7. References

[1] *IEEE Computer Magazine*, Special Issue on "Parallel Computers," Jan., 1982.

[2] C.G. Bell, "Multis : A New Class of Multiprocessor Computers," *Science*, Vol. 228, pp.462-467, April, 1985.

[3] S.H. Fuller, *et al.*, "Multi-Microprocessors : An Overview and Working Example," *Proceedings of the IEEE*, Vol. 66, No. 2, pp.216-228, Feb., 1978.

[4] D.P. Bhandarkar, "Analysis of Memory Interface in Multiprocessors," *IEEE Transactions on Computers*, Vol. c-24, No. 9, pp.897-908, Sept., 1975.

[5] D.H. Lawrie, "Access and Alignment of Data in an Array Processor," *IEEE Transactions on Computers*, Vol. c-24, No. 12, pp.1145-1155, Dec., 1975.

[6] *Applied Optics*, Special Issue on "Optical Interconnections," Vol. 29, No. 8, March, 1990.

[7] L. Giles, K. Jenkins, "Models of Parallel Computation and Optical Computing," *Proceedings of OSA Annual Meeting*, Paper Number ML1, Oct., 1986.

[8] G. Eichmann, H.J. Caulfield and I. Kadar, "Optical artificial intelligence and symbolic computing : an introduction," *Applied Optics*, Vol. 26, No. 10, pp.1827-1829, May, 1987.

[9] K. Brenner, A. Huang and N. Streibl, "Digital optical computing with symbolic substitution," *Applied Optics*, Vol. 25, No. 18, pp.3054-3060,

Sept., 1986.

[10] R.W. Keyes, "Optical logic – in the light of computer technology," *Optica Acta*, Vol. 32, No. 5, pp.525-535, 1985.

[11] A.A. Sawchuk and T.C. Strand, "Digital Optical Computing," *Proceedings of the IEEE*, Vol. 72, No. 7, pp.758-779, July, 1984.

[12] *Applied Optics*, Special issue on "Optical Computing," Vol. 29, No. 14, May, 1990.

[13] *Proceedings of the IEEE*, Special issue on "Optical Computing," Vol. 72, No. 7, July, 1984.

[14] *Applied Optics*, Special Issue on "Neural Networks," Vol. 26, No. 23, Dec., 1987.

[15] Rumelhart and McClelland, "Parallel Distributed Processing," Vol. 1, MIT Press, Boston, 1986.

[16] R.P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, pp.4-22, April, 1987.

[17] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Science, USA*, Vol. 79, pp.2554-2558, April, 1982.

[18] G.M. Sheperd, "The Synaptic Organization of the Brain," Oxford University Press, New York, 1979.

[19] S.W. Kuffler, J.G. Nicholls and A.R. Martin, "From Neuron to Brain," Sinauer Associates, Sunderland, Mass., 1984.

[20] R.V. Jones, "Photoneural systems : an introduction," *Applied Optics*, Vol. 26, No. 10, pp.1948-1958, May, 1987.

[21] T. Williams, "Optics and neural nets : trying to model the human brain," *Computer Design*, pp.47-62, March, 1987.

[22] K. Wagner and D. Psaltis, "Multilayer optical Learning networks," *Applied Optics*, Vol. 26, No. 23, pp.5061-5076, Dec., 1987.

[23] D. Psaltis and N. Farhat, "Optical information processing based on an associative memory model of neural nets with thresholding and feedback," *Optics Letters*, Vol. 10, pp.98-100, Jan., 1985.

[24] D. Psaltis, D. Brady and K. Wagner, "Adaptive optical networks using photorefractive crystals," *Applied Optics*, Vol. 27, No. 9, pp.1752-1758, May, 1988.

[25] K. Kyuma, *et al.*, "Optical Neural Networks : System and Device Technologies," *Proceedings of the SPIE*, Vol. 963, Optical Computing 88, pp.475-484, 1988.

[26] D. Psaltis, *et al.*, "Holography in artificial neural networks," *Nature*, Vol. 343, pp.325-330, Jan., 1990.

[27] R.P. Lippmann, "Pattern Classification using Neural Networks," *IEEE Communications Magazine*, pp.47-64, Nov., 1989.

[28] T.J. Sejnowski and C.R. Rosenberg, "Parallel Networks that Learn to Pronounce English Text," *Complex Systems*, Vol. 1, pp.145-168, 1987.

[29] K. Fukushima, "Neural network model for selective attention in visual pat-

tern recognition and associative recall," *Applied Optics*, Vol. 26, No. 23, pp.4985-4992, Dec., 1987.

[30] J.M. Oyster, F. Vicuna and W. Broadwell, "Associative network applications to low-level machine vision," *Applied Optics*, Vol. 26, No. 10, pp.1919-1926, May, 1987.

[31] D. Tank and J.J. Hopfield, "Simple 'Neural' Optimization Networks : An A/D Converter, Signal Decision Circuit and a Linear Programming Circuit," *IEEE Transactions on Circuits and Systems*, Vol. CAS-33, No. 5, pp.533-541, May, 1986.

[32] Y. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," *Scientific American*, pp.88-95, March, 1987.

[33] D. Psaltis, A.A. Yamamura, M.A. Neifeld and S. Kobayashi, "Parallel Readout of Optical Disks," *OSA Optical Computing 1989 Technical Series Digest*, Vol. 9, pp.58-61, Feb., 1989.

[34] D. Psaltis, M. A. Neifeld, A. A. Yamamura and S. Kobayashi, "Optical memory disks in optical information processing," *Applied Optics*, Vol. 29, No. 14, May, 1990.

[35] J.H. Rilum, A.R. Tanguay, Jr., "Utilization of optical memory disks for optical information processing," *OSA 1988 Annual Meeting Technical Digest*, pp.43, 1988.

[36] *Applied Optics*, Special issue on "Optical Storage of Digital Data," Vol. 13, No. 4, April, 1974.

[37] J. Hecht, "Optical Memory for Personal Computers," *Lasers and Applications*, pp.71-76, Aug., 1985.

[38] *IEEE Computer Magazine*, Special issue on "Mass Storge Systems," July, 1985.

[39] R. Duda and P. Hart, "Pattern Classification and Scene Analysis," pp.141–147, John Wiley and Sons, New York, 1973.

[40] Neocognitron

[41] Y. Le Cun, et al., "Handwritten Digit Recognition: Applications of Neural Network Chips and Automatic Learning," *IEEE Communications Magazine*, pp.41-46, Nov., 1989.

[42] Neural Speech Recognition

[43] K. Fukanaga, "Introduction to Statistical Pattern Recognition," Academic Press, New York, 1972.

[44] S. Watanabe ed., "Methodologies of Pattern Recognition," Academic Press, New York, 1969.

[45] Y. S. Abu-Mostafa and D. Psaltis, "Image Normalization by Complex Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 1, Jan., 1985.

[46] D. Casasent and D. Psaltis, "Position, rotation and scale invariant optical correlation," *Applied Optics*, Vol. 15, No. 7, pp.1795-1799, July, 1976.

[47] M.O. Freeman and B.E.A. Saleh, "Moments invariants in the space and frequency domains," *Journal of the Optical Society of America A*, Vol. 5,

No. 7, pp.1073-1084, July, 1988.

[48] Y.S. Abu-Mostafa and D. Psaltis, "Recognitive Aspects of Moment Invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 6, pp.698-706, Nov., 1984.

[49] B. Guenther, C. Chritensen and J. Upatnieks, "Coherent Optical Processing : Another Approach," *IEEE Journal of Quantuum Electronics*, Vol. QE-15, No. 12, pp.1348-1362, 1979.

[50] A. Vander Lugt, "Signal Detection by Complex Spatial Filtering," *IEEE Transactions on Information Theory*, Vol. IT-10, pp.139, 1964.

[51] D. Casasent, "Coherent Optical Pattern Recognition," *Proceedings of the IEEE*, Vol. 67, No. 5, pp.813-825, May, 1979.

[52] E.B. Baum and D. Haussler, "What Size Net Gives Valid Generalization," *Advances in Neural Information Processing Systems 1*, D.S. Touretzky, ed., pp.81-90, Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[53] L.G. Valiant, "A Theory of the Learnable," *Communications of the ACM*, Vol. 27, No. 11, pp.1134-1142, Nov., 1984.

[54] C.L. Giles and T. Maxwell, "Learning, invariance and generalization in high-order neural networks," *Applied Optics*, Vol. 26, No. 23, pp.4972-4978, Dec., 1987.

[55] E. Baum, "On the Capabilitites of Multilayer Perceptrons," *Journal of Complexity*, Vol. 4, pp.193-215, 1988.

[56] F.J. Pineda, "Generalization of Back-Propagation to Recurrent Neural Net-

works," *Physical Review Letters*, Vol. 59, No. 19, pp.2229-2232, Nov., 1987.

[57] T. Grossman, R. Meir and E. Domany, "Learning by Choice of Internal Representations," *Complex Systems*, Vol. 2, pp.555-575, 1988.

[58] N.J. Nilsson, "Learning Machines," McGraw Hill, New York, 1965.

[59] T. Poggio and F. Girosi, "A Theory of Networks for Approximation and Learning," *MIT AI Laboratory and Center for Biological Information Processing, Whitaker College*, AI Memo No. 1140, CBIP paper No. 31, July, 1989.

[60] V.N. Vapnik and A. Chervonenkis, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," *Theory of Probability and its Applications*, Vol. XVI, No. 2, pp.264-280, 1971.

[61] H.D. Block, "The Perceptron : A Model for Brain Functioning," *Reviews of Modern Physics*, Vol. 34, No. 1, pp.123-135, Jan., 1962.

[62] H.D. Block, B.W. Knight and F. Rosenblatt, "Analysis of a Four-Layer Series-Coupled Perceptron," *Reviews of Modern Physics*, Vol. 34, No. 1, pp.135-142, Jan., 1962.

[63] F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, N.Y., 1962.

[64] M. Minsky and S. Papert, *Perceptrons : an Introduction to Computational Geometry*, (expanded Edition), MIT Press, Boston, 1988.

[65] M.A. Neifeld and D. Psaltis, "The Emergence of Generalization in Networks with Constrained Representations," *Proceedings of the Second International Conference on Neural Networks*, IEEE ICNN pp.123, San Diego, 1988.

[66] D. Psaltis, "Acousto-optic processing of two-dimensional signals," *Journal of the Optical Society of America*, Vol. 71, No. 2, pp.198-200, Feb., 1981.

[67] D. Psaltis, "Optical image correlation using acoustooptic and charge-coupled devices," *Applied Optics*, Vol. 21, No. 3, pp.491-95, Feb., 1982.

[68] D. Psaltis, "Two-Dimensional Optical Processing using One-Dimensional Input Devices," *Proceedings of IEEE*, Vol. 72, No. 7, pp.962-974, 1984.

[69] D. Psaltis, "Incoherent Electrooptic Image Correlator," *Optical Engineering*, Vol. 23, No. 12, 1984.

[70] Perry A. Molley, K. Terry Stalker, "A High Dynamic Range Acoustooptic Image Correlator for Realtime Pattern Recognition," *Proceedings of the SPIE*, Vol. 938, Orlando Fla., April, 1988.

[71] W. Rhodes, "Acousto-Optic Signal Processing : Convolution and Correlation," *Proceedings of the IEEE*, Vol. 69, No. 1, pp.65-79, Jan., 1981.

[72] J. Heemskerk and K. Schouhamer Immink, "Compact disk : system aspects and modulation," *Phillips Technical Review*, Vol. 40, No. 6, pp.157-164, 1982.

[73] H. Hopkins, "Diffraction theory of laser readout system for optical video discs," *Journal of the Optical Society of America*, Vol. 69, No. 1, pp.4-24, Jan., 1979.

[74] F. Smits and L. Gallaher, "Design considerations for a semipermanent optical memory," *The Bell System Technical Journal*, pp.1267-1278, July, 1967.

[75] D. Chen and J. Zook, "An overview of optical data storage technology,"

Proceedings of the IEEE, Vol. 63, No. 8, pp.1207-1230, Aug., 1975.

[76] G. Thomas, "Future trends in optical recording," *Phillips Technical Review*, Vol. 44, No. 2, pp.51-57, April, 1988.

[77] P.B. Berra and N.B. Troullinos, "Optical Techniques and Data/Knowledge Base Machines," *IEEE Computer*, pp.59-70, October, 1987.

[78] P.B. Berra, "Optics and Supercomputing," *Proceedings of the IEEE*, Vol. 77, No. 12, pp.1797-1815, Dec., 1989.

[79] R. Bartolini, H. Weakliem and B. Williams," Review and analysis of optical recording media," *Optical Engineering*, Vol. 15, No. 2, pp.99-108, March, 1976.

[80] R. Bartolini, "Media for high density optical recording," *Optical Engineering*, Vol. 20, No. 3, pp.382-386, May, 1981.

[81] B. Brown, "Optical data storage potential of six materials," *Applied Optics*, Vol. 13, No. 4, pp.761-766, April, 1974.

[82] D. Chen, "Magnetic materials for optical recording," *Applied Optics*, Vol. 13, No. 4, pp.767-778, April, 1974.

[83] I. Hatakeyama, *et al.*, "Magnetooptic thin film reading head," *Applied Optics*, Vol. 25, No. 1, pp.146-150, Jan., 1986.

[84] B. Huth, "Calculations of stable domain radii produced by thermomagnetic writing," *IBM Journal of Research and Development*, pp.100-109, March, 1974.

[85] M. Ojima, *et al.*, "Compact magnetooptic disk for coded data storage,"

Applied Optics, Vol. 25, No. 4, pp.483-489, Feb., 1986.

[86] S. Esho, *et al.*, "Optimum thickness of MnBi films for magnetooptical memory," *Applied Optics*, Vol. 13, No. 4, pp.779-783, April, 1974.

[87] J. Feinleib, *et al.*, "Rapid reversible light induced crystallization of amorphous semiconductors," *Applied Physics Letters*, Vol. 18, No. 6, pp.254-257, March, 1971.

[88] R.J. von Gutfeld and P. Chaudhari, "Laser writing and erasing on chalcogenide films," *Journal of Applied Physics*, Vol. 43, No. 11, pp.4688-4693, Nov., 1972.

[89] D. Strand, "Optical recording with amorphous materials," *Optical Engineering*, Vol. 20, No. 3, pp.379-381, May, 1981.

[90] Photopolymer

[91] W. Moerner, ed. , "Persistent spectral hole burning," *Topics in Current Physics*, Vol. 44, Springer Verlag, Berlin, Heidelberg, 1988.

[92] A. Renn and U. Wild, "Spectral hole burning and holography : multiple storage of images," *Proceedings of ISOM89*, Paper No. 28B-8, Kobe, Japan, 1989.

[93] T. Allen and G. Ash, "Optical properties of tellurium films used for data recording," *Optical Engineering*, Vol. 20, No. 3, pp.373-376, 1981.

[94] M. Cunningham, "Recording mechanism studies in organic media," *Proceedings of Joint IEEE, OSA, SPIE Topical Meeting on optical data storage*, pp.207-210, Vancouver, Canada, March, 1990.

[95] Optical Storage Technology Group Sony Corp. of America, "Recording Principle and Reliability Study of Sony Century Media," *Sony Corp. Technical Report*, April, 1988.

[96] Y. Nakane, *et al.*, "Principle of laser recording mechanism by forming an alloy in the multilayer of thin metallic films," *Proceedings of the SPIE*, Vol. 529, pp.76-82, July, 1985.

[97] B. Brown and A. Lohmann, "Complex Spatial Filtering with Binary Masks," *Applied Optics*, Vol. 5, pp.967, 1966.

[98] W. Lee, "Binary Computer Generated Holograms," *Applied Optics*, Vol. 18, pp.3661, 1979.

[99] G. Tricoles, "Computer generated holograms: an historical review," *Applied Optics*, Vol. 26, No. 20, pp.4351-4360, Oct., 1987.

[100] D. Psaltis, M. Neifeld, A. Yamamura and S. Kobayashi, "Holographic Recording on Optical Disks," *Proceedings of the International Symposium on Optical Memory* , Kobe, Japan, Sept., 1989.

[101] Y. Tsunoda, *et al.*, "Holographic video disk: an alternative approach to optical video disks," *Applied Optics*, Vol. 15, No. 6, pp.1398-1403, June, 1976.

[102] I. Satoh and M. Kato, "Holographic disk recording of digital data with fringe stabilization," *Applied Optics*, Vol. 27, No. 14, pp.2987-2992, July, 1988.

[103] T. Yatagai, J.G. Camacho-Basilio and H. Onda, "Recording of Computer-Generated Holograms on an Optical Disk Master," *Proceedings of the SPIE*,

Vol. 1052, No. 16, Los Angeles, 1989.

[104] D. Psaltis, M.A. Neifeld and A.A. Yamamura, "Optical Disk Based Correlation Architectures," *OSA Optical Computing 1989 Technical Series Digest*, Vol. 9, pp.206-209, Feb., 1989.

[105] D. Psaltis, M. A. Neifeld, A. A. Yamamura, "Image Correlators Using Optical Memory Disks," *Optics Letters*, Vol. 14, No. 9, pp.429–431, 1989.

[106] A.W. Warner, D.L White and W.A. Bonner, "Acousto-optics light deflectors using optical activity in paratellurite," *Journal of Applied Physics*, Vol. 43, No. 11, pp.4489-4495, Nov., 1972.

[107] R.W. Dixon, "Acoustic Diffraction of Light in Anisotropic Media," *IEEE Journal of Quantum Electronics*, Vol. QE-3, No. 2, pp.85-93, Feb., 1967.

[108] J. Sapriel, "Acousto-Optics," John Wiley and Sons, New York, 1979.

[109] J.F. Jarvis, C.N. Judice and W.H. Ninke, "A survey of techniques for the display of continuous tone pictures on bilevel displays," *Computer Graphics and Image Processing*, Vol. 5, pp.13-40, 1976.

[110] T.J. Hall, *et al.*, " The Photorefractive Effect - A Review," *Progress in Quantum Electronics*, Vol. 10, pp.77-146, 1985.

[111] D. von der Linde and A.M. Glass, "Photorefractive Effects for Reversible Holographic Storage of Information," *Applied Physics*, Vol. 8, pp.85-100, Springer Verlag, 1975.

[112] A. Yariv, "Phase Conjugate Optics and Real-Time Holography," *IEEE Journal of Quantum Electronics*, Vol. QE-14, No. 9, pp.650-660, Sept.,

1978.

[113] D. Psaltis, J. Yu and J. Hong, "Bias-free time-integrating optical correlator using a photorefractive crystal," *Applied Optics*, Vol. 24, No. 2, pp.3860-3865, Nov., 1985.

[114] M.G. Nicholson, *et al.*, "Simple computational model of image correlation by four-wave mixing in photorefractive media," *Applied Optics*, Vol. 26, No. 2, pp. 278-286, Jan., 1987.

[115] J. Yu, "Optical Processing using Photorefractive Crystals," Ph.D. Thesis, Chapter 5, Caltech, 1988.

[116] C.X. Gu, "Optical Neural Networks using Volume Holograms," Ph.D. Thesis, Chapter 3, Caltech, 1990.

[117] R.R. Neurgaonkar and W.K. Cory, "Progress in photorefractive tungsten bronze crystals," *Journal of the Optical Society of America B*, Vol. 3, No. 2, pp.274-282, Feb., 1986.

[118] J.B. Thaxter and M. Kestigian, "Unique Properties of SBN and Their Use in a Layered Optical Memory," *Applied Optics*, Vol. 13, No. 4, pp.913-924, April, 1974.

[119] Sebestyen, "Decision Making Processes in Pattern Recognition," The Macmillan Co., New York, 1962.

[120] T. Cover, "Estimation by the Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, Vol. IT-14, No. 1, pp.50-55, Jan., 1968.

[121] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification,"

IEEE Transactions on Information Theory, Vol. IT-13, No. 1, pp.21-27, Jan., 1967.

[122] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the Fifth Berkeley Symposium on Math. Stat. and Prob. I.*, Berkeley and L. A. California, 1967.

[123] T. Cover, "Rates of Convergence for Nearest Neighbor Procedures," *Proceedings of the First Annual Hawaii Conference on Systems Theory*, pp.413-415, 1968.

[124] P. Hart, "The Condensed Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, Vol. IT-14, pp.515-516, May, 1968.

[125] T. Wagner, "Convergence of the Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, Vol. IT-17, No. 5, pp.566-571, Sept., 1971.

[126] John Moody and Christian Darken, "Fast Learning in Networks of Locally Tuned Processing Units," *Neural Computation*, Vol. 1, pp.281–294, 1989.

[127] W.P. Bleha, "Progress in Liquid Crystal Light Valves," *Laser Focus Magazine*, Oct., 1983.

[128] W.P. Bleha, *et al.*, "Application of the Liquid Crystal Light Valve to Real-Time Optical Data Processing," *Optical Engineering*, Vol. 17, No. 4, pp.371-384, July, 1978.

[129] R.A. Athale and S.H. Lee, "Bistability and thresholding by a new photoconductor-twisted nematic liquid crystal device with optical feedback," *Applied Optics*, Vol. 20, No. 8, pp.1424-1432, April, 1981.

[130] A. A. Yamamura, S. Kobayashi, M. A. Neifeld and D. Psaltis, "An Optoelectronic Multilayer Network," presented at SPIE OE LASE '90, Los Angeles, Jan., 1990.

[131] D. Psaltis, A. A. Yamamura, M. A. Neifeld and S. Kobayashi, "Optoelectronic Implementations of Neural Networks, " *IEEE Communications Magazine*, Vol. 27, No. 11, Nov., 1989.

[132] M. Neifeld, D. Psaltis, A. Yamamura and S. Kobayashi, "Optical Disk Based Architectures for Pattern Recognition and Neural Network Implementation," *Annual Meeting of the Optical Society of America Technical Digest*, pp.46, Orlando, Oct. 1989.

[133] J. Alspector and R.B. Allen, "A Neuromorphic VLSI Learning System," *Advanced Research in VLSI Proc. 1987 Stanford Research Conf.*, pp.313-349, MIT Press, Boston, 1987.

[134] C. Mead, "Analog VLSI and Neural Systems," Addison-Wesley, New York, 1989.

[135] M.R. Haskard and I.C. May, "Analog VLSI Design nMOS and cMOS," Prentice Hall, New York, 1988.

[136] D. Psaltis and J. Hong, "Shift Invariant Optical Associative Memories," *Optical Engineering*, Vol. 26, pp.10-15, Jan. 1987.

[137] K. Hsu, D. Brady and D. Psaltis, "Experimental Demonstrations of Optical Neural Computers," *Neural Information Processing Systems*, D.Z. Anderson ed., American Institute of Physics, N.Y., 1988.

[138] J.H. Kim, S.H. Lin, J. Katz and D. Psaltis, "Monolithically Integrated 2-D Arrays of Optoelectronic Devices for Neural Network Applications," *Proceedings of the SPIE*, Vol. 1043, Los Angeles, Jan., 1989.

[139] D. Wang, J. Buhmann and C. von der Malsburg, "Pattern Segmentation in Associative Memory," *Neural Computation*, Vol. 2, No. 1, pp.94-106, 1990.