Accepted for publication in the
Journal of Visual Languages and Computing
Academic Press, UK, special Issue on Visual Information Systems, will appear in Vol 8, no 2, in June 1997.

A PARALLEL, DISTRIBUTED AND ASSOCIATIVE APPROACH FOR PATTERN MATCHING WITH COMPLEX DYNAMICS

Javed I. Khan & David Y. Y. Yun

Contact:
Laboratories of Intelligent and Parallel Systems
Department of Electrical Engineering
493 Holmes Hall, 2540 Dole Street
University of Hawaii at Manoa
HI-96822, USA

Phone: (808)-956-3868
Fax: (808)-941-1399
javed@hawaii.edu

SUMMARY

This paper presents a new associative pattern matching network based on a digital adaptation of optical holography. Unlike any existing neural network or associative memories, it can dynamically localize its search on any subset of the pattern space and at the same time can generate a feedback on the quality of match. Current associative memories based on neuro computing is unable to support such meta-interaction. Conceptually, this scheme involves adaptive "enfolding" of the raw massive search space and direct regeneration of the matching pattern during search. The search process is virtually a constant time operation compared to traditional algorithm approaches, inherently parallelizable, and an excellent candidate for hardware or optical implementation. This new approach is expected to significantly facilitate applications that require direct pattern matching in massive image repositories in real-time. This is also the first pattern matching system which utilizes the content-based retrieval abilipp
ty of associative computing.

A PARALLEL, DISTRIBUTED AND ASSOCIATIVE APPROACH FOR SEARCHING IMAGE PATTERN WITH COMPLEX DYNAMICS

Javed I. Khan & D. Y. Y. Yun

Laboratories of Intelligent and Parallel Systems
Electrical Engineering Department
University of Hawaii at Manoa
javed@hawaii.edu

ABSTRACT

This paper presents a new associative pattern matching network based on a digital adaptation of optical holography. Unlike any existing neural network or associative memories, it can dynamically localize its search on any subset of the pattern space and at the same time can generate a feedback on the quality of match. Current associative memories based on neuro computing is unable to support such meta-interaction. The scheme involves adaptive "enfolding" of the raw massive search space into a holograph and direct regeneration of the matched target pattern during search. The search process is a constant time operation compared to traditional algorithm approaches, inherently parallelizable, and an excellent candidate for hardware or optical implementation. This new technique is expected to significantly facilitate applications that require direct pattern matching in massive image repositories in real-time. *Target recognition*, *visual-query*, *content-based image retrieve*, and *automatic index-extraction* are just few such applications.

1 INTRODUCTION

This research describes a new computational technique for real-time pattern matching into massive digital information repositories. It is based on a pattern matching network constructed on the principles of optical holography. This technique "enfolds" the massive search space into a single holographic memory substrate of exponentially reduced dimension. During search, instead of actually accessing the large raw pattern space, the holographic regeneration technique directly computes the pattern through associative recollection into this holograph.

The key to this technique is a new holographic pattern matching network which unlike any existing Artificial Neural Network (ANN) or Artificial Associative Memories (AAM), has two important interaction capabilities. First, it can dynamically localize its search on any sub-pattern space. Secondly, it can provide a feedback measure on the quality of match, in addition to the regenerated pattern. Both of these capabilities are essential for effective pattern matching. But neither are supported by current AAMs or ANNs models.

In recent years optical holography has attracted great interest for its high information storage density and speed of access [PsMo95]. Because of the rapid advancements in key technologies during the last five years, specifically in finer optical signal detection, new polymer invention, and innovative optical assembly [Sinc94, HMCC95] now it is conceivable that hundreds of billions of bytes of data can be stored and transferred at a rate of billion bits per second or faster and any data element can be randomly accessed in 100 microseconds or less. If compared to the best technology available today, it promises boost of storage density and retrieval time in the order of 100 to 1000 times[1]. No other memory technology that offers such performance is as close to commercialization.

Already, in early 1993, a CalTech group demonstrated storage of 100 MB data on a single crystal. In 1994 another Stanford group displayed a crystal where data can be retrieved with a reliability of less than one bit error per million bits. Another group at Rockwell demonstrated a compact system with 1000 holograms, where arbitrary pages can be accessed in less than 40 microseconds and data can be retrieved without error. More recently a new startup company Holoplex developed a highspeed memory system capable of storing 1000 finger prints (approximately half of a CD) for a security system and yet retrieving all data in less than a second. In the past three years major companies like Rockwell, IBM and GTE have launched or expanded their initiative to develop holographic memories [PsMo95, HMCC95].

In the backdrop of such exciting and phenomenal advancement of holographic storage memories (both theoretical and practical), we have recently expanded our investigation into more advanced use of holographic memories based on their associative recollection ability. The holographic storage techniques mentioned above perform retrieval of pattern-waveforms from index waveforms. In this research we investigate its ability to retrieve pattern waveforms from sub-pattern waveforms. In this paper we demonstrate how memories based on holographic principles can be used to perform associative sub-pattern matching with the capability of dynamic localization at amazing speed by manipulating the modulus constraints of holographic representation.

Why Need Search Localization? Ability of dynamic search localization is fundamental to any search process. In visual perception, it is natural that when we match information, we always focus on some specific objects in the pattern. We never do an indiscriminate pixel to pixel matching on the entire pattern. This new computing net provides such ability of dynamic focusing on a pattern sub-space with the speed and parallelizability of associative computing and can be put into cellular hardware[2]. The crucial role of search localization in visual search problems is illustrated next through an example of image perception. But such search localization is not supported by existing cellular computation models such as AAMs or ANNs.

Consider the images in frames A, B, and C. Let these three frames are stored in an associative memory. Let us now use template-D as a sample for retrieval. A natural expectation is that the memory should retrieve frame A as the closest match corresponding to this sample. This expectation is based on the implied cognitive significance of the object roller in the image. However, any conventional AAMs will retrieve frame C as the closest match. This is because frames B or C are closer to D than A both in *least mean square* (LMS) and *maximum dot-product* sense. The reason for such an awkward response of conventional AAMs is the comparative statistical strength of the index region. Although the region denoting the roller is cognitivically more important but their statistical dominance is obscured by the large cover of the cognitivically less important background area. In contrast to such response of conventional AAMs, the internal matching process of natural memory seems to be immune to such statistical weakness and can retrieve information by localizing attention on cognitivically important zones.

Dynamic Attention: From retrieval point of view, even more revealing limitation of current AAMs is that apparently most natural associative matching process can change the locality of attention dynamically during recall. Consider template-E which has two objects of focus and

1 An Optical Compact Disk (CD) stores about 640 MBytes. The retrieval time of its mechanical head is in the order of milliseconds. The retrieval mechanism of holographic storage is non-mechanical.

2 This technique also has excellent potential for fast optical realization, because the underlying hyperspherical computations can be naturally implemented on optical computations.

two ostensible matches. A natural memory can easily shift its focus to the Plant or any other objects (or if needed even to the background) and retrieve variety of corresponding matches apparently without any significant internal reorganization of its learned state. Such flexibility is missing in a conventional AAM. For a given AAM, after the learning it acts as a deterministic machine where each initial state flows into a pre-determined single attractor without accommodating dynamic (post-learning) change in the locality of focus in the pattern element space.

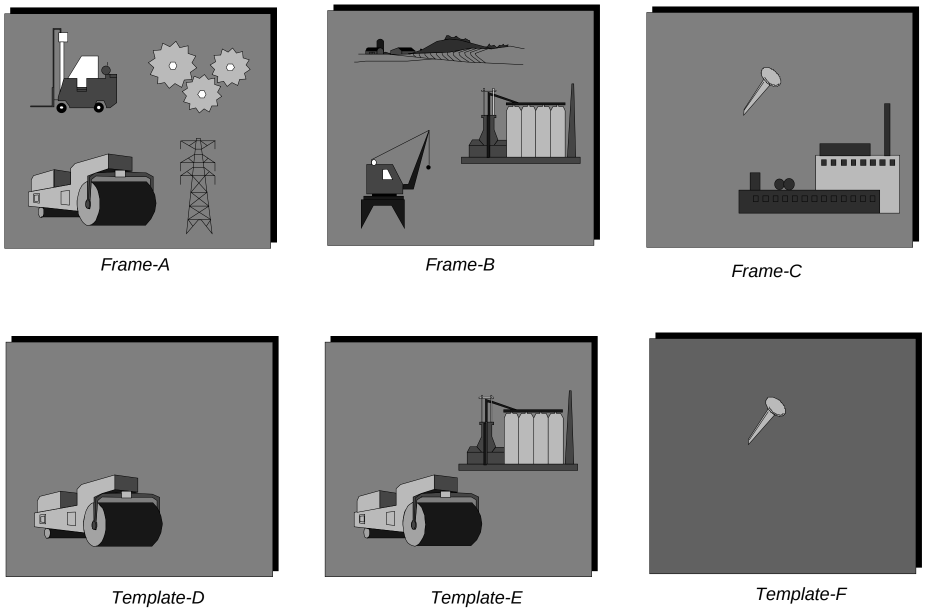


Fig-1 Attention Modulated Retrieval

Statistical Strength of Effective Cue: A particularly serious consequence of the inability to localize search is the inability to work with a small cue (such as the nail in template-F). Conventional AAMs require the index to be statistically large in the pattern. As soon as the amount of "incorrect" part becomes statistically dominant over the "correct" part of input information these networks cease to converge correctly. Any conventional AAM requires the effective cue to be at least half (50%[3]) the pattern size for correct retrieval. This is quite stringent restriction for most pattern matching applications. A memory with the ability of search localization can obtain match only with local dominance within the field of attention instead of the global dominance, and thus can avoid such limitation.

Managing Incomplete Information: Dynamically localizable search is also critically important in the processing of *incomplete information*. Let us consider a decision problem where out of 100 data points, 15 specific sensors are unavailable, and yet the best possible decision has to be made with respect to the available amount of information. Notably, this situation is quite different from the case of noisy information where it is not known which

3 It is an well observed phenomenon in almost all neural networks and AAMs that, as the number of error bits approach approximately just 40%, the probability of correct recall vanishes sharply irrespective of the sophistication of the learning algorithms. Experiments published by many researchers contains the fingerprint of such pure statistical nature of neural network convergence [TaJo90, Hopf82, MiFa90, KuWo91] (although, it apparently always managed to escape persuasion).

15 are missing. In the former case, some extra knowledge is available about the location of the missing elements. A pattern matching machine which can not localize its search, can not take advantage of this extra information despite their availability and consequently will converge to a solution of lesser optimality for the available amount of knowledge. A memory with the ability of search localization, in contrast, can localize its optimization only to the available data points leading to a better solution.

Computation Advantages for Real-Time Image Matching: Conventional and classical pattern matching techniques (see appendix-A for comparison) requires too many computations to realize dynamic attention. This is because the dynamic assignability of the relative significances of pattern elements prohibits any post-ordering of the element space. Without internal ordering, algorithmic approaches reduce to exhaustive linear search. Associative techniques characteristically enjoy efficiency advantage almost by exponential order with respect to the number of patterns.

As an instance of associative computing machines, like other AAMs, this new approach is (i) adaptive, (ii) capable of dealing with imprecise information, and (iii) readily scalable. Notably, image information is also inherently (i) difficult to model, (ii) subjective and imprecise, and (iii) content-wise sparse but representationally massive. These characteristics naturally suggest suitability of associative computing for real-time image processing. However, as explained earlier the capabilities of current models are limited to non-localizble search with unary attention only.

[Khan95] has recently investigated the associative recollection ability of holographic memories. It has been demonstrated in this work that an associative computation model called as *Multidimensional Holographic Associative Computing* (MHAC), based on a novel adaptation of optical holography can overcome aforementioned limitations. This paper presents this holographic technique from the perspective of its potential for real-time pattern matching in massive image repositories.

The following section first presents the conceptual model of this new memory and its capabilities. Section 3 then briefly discusses the relative strength and weakness of other associative memory approaches those are available today. Sections 4 and 5 then present the computational model of MHAC and analysis of its performance. Finally section 6 provides the characterization test results of this new technique in real-time pattern matching, along with an actual example application involving image matching in real-time.

2 ASSOCIATIVE MEMORY AND SEARCH

In this section first some very basic concepts about memory and search will be revisited to clarify the distinguishing aspects of MHAC. We will also establish some notations along with which will be used throughout this paper.

A memory stores and retrieves a collection of patterns. A pattern is a collection of elements. Let a stimulus and corresponding response pattern be denoted by the symbolic vectors $S^{\mu} = [s_1^{\mu} \quad s_2^{\mu} \quad \dots s_n^{\mu}]$ and $R^{\mu} = [r_1^{\mu} \quad r_2^{\mu} \quad \dots r_M^{\mu}]$. Here subscript refers to the element index and superscript refers to pattern index. Individual elements in these vectors can be considered to represents a piece of *information*. The values of these elements generally correspond to a measurement obtained by some physical sensor.

An associative memory system can be considered to be comprised of (i) a learning algorithm $\mathbf{A}_{\text{learn}}$ which converts all the $\{S^{\mu}, R^{\mu}\}$ associations into some internal representation, (ii) a physical storage medium and representation formalism **AM** to store the associations, (iii) a decoding algorithm $\mathbf{A}_{\text{retrieve}}$ to recollect stored information R^R from a given query stimulus S^Q, and (iv) a matching criterion D to measure the closeness of stimulus patterns to the query pattern. The operation of an associative memory can be put together as:

Definition (Associative memory): *Given, a set of stimulus pattern vectors* $\mathbf{S} = \{S^{\mu} \mid 1 \le \mu \le P\}$ *and a set of response pattern vectors* $\mathbf{R} = \{R^{\mu} \mid 1 \le \mu \le P\}$ *an associative memory is capable of learning the relationship between a stimulus member* $S^{\mu} \in \mathbf{S}$ *and the corresponding response member* $R^{\mu} \in \mathbf{R}$ *in such a way that, given a query pattern* S^{Q}, *it can retrieve a pattern* $R^{R} \approx R^{T}$ *such that* $R^{T} \in \mathbf{R}$, *and* S^{Q} *is closest (best match) to* $S^{T} \in \mathbf{S}$ *according to a matching criterion* D.

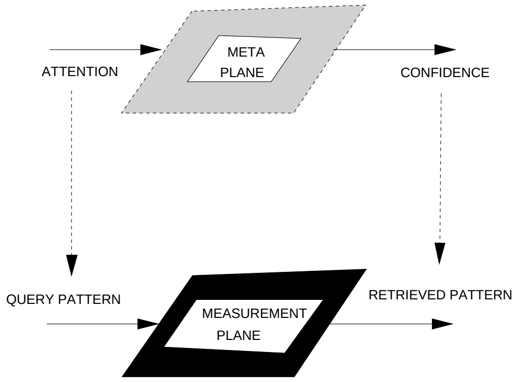


Fig-2 Information Flow Model of Bimodal Memory

Bimodality of Information: The search process in an associative memory, in its simplest form can also be considered as an input output system with two ports (as shown in Fig-2 bottom plane). It has an input where query stimulus pattern is received from inquirer, and an output where for each query the memory generates the response.

Now we will introduce the difference between MHAC and conventional AAMs. MHAC is a super-class of conventional AAMs in the following sense. A conventional memory formalism processes the *measurement* components of information elements. In contrast, MHAC includes a second component called *meta-knowledge* about the state of each given piece of information (measurement).

The exact linguistic meaning of the meta quantity varies. At the input side, it corresponds to a form of *attention* defined on the query pattern (or more accurately, on the measurement of the query pattern) given by the inquirer, and at the output side, it corresponds to the *confidence* on the retrieved pattern (i.e. on the measurements of the retrieved pattern) as assessed by the memory itself. The upper plane in Fig-2 shows this correspondence. MHAC makes such status information an integral part of its representation.

Formally, each of the elements of information in MHAC is represented as a bi-modal pair $\{$*measurement,meta-information*$\}$ ($s_k^{\mu} = \{\alpha_k^{\mu}, \beta_k^{\mu}\}$). Where α represents the *measurement* of the information elements and β represents the *meta-knowledge* (i.e. confidence or attention) associated with this measurement.

Behavioral Definition of β: To obtain a functional definition of this new representation of an information element, we now specify how this new meta-quantity should behave and functionally relate to usual measurements. In the context of any general memory (irrespective of its implementation mechanism), the above formalism indeed generates some specific expectations about the operational behavior of these meta quantities. These are summarized in the form of following two expectations, one in input side and the other in output side.

Inflow Expectation (1): *The memory matching criterion should put more importance to a piece of information that is attributed with high degree of inflow* β *than to a piece attributed with low* β *in the query. The expectation can be stated as a matching criterion:*

$$D\left(S^{Q}, S^{T_a}, \mathrm{B}\right) = \underset{i}{\overset{N}{\mathrm{M}}}\, \beta_i dist\left(\alpha_i^{Q} \alpha_i^{T_a}\right) \qquad \text{....(1)}$$

Here $\mathrm{B} = [\beta_1 \beta_2 \ldots \beta_n]$ denotes the *meta-vector*. As evident, this meta-vector represents a form of importance or attention on the pattern vector.

From the context of query and dynamic specifyability of B^{Q}, this expectation corresponds to the matching criterion that realizes *retrieval with changeable attention* (RCA).

Here, M is a set operator with scope N. generally a summation is used such that $\mathrm{M} = \sum\{\cdot\}$. $dist(\cdot)$ is the *distance measure function* between any two pattern elements. It can be any arbitrary function with the constraint that it is monotonic with respect to $|s_i^a - s_i^b|$, and symmetric for all (s_i^a, s_i^b) pairs.

The other symmetric expectation related to the outflow of meta-knowledge provides completeness to our attempt of delineating the behavior of a new memory which deals with meta-knowledge.

Outflow Expectation (2): *If α values of query demonstrate high degree of resemblance to the α values of a priory encoded stimulus pattern, then memory should retrieve the associated response α with higher degree of accuracy and high degree of β. On the other hand, if it does not then it should generate a response with low degree of β.*

This outflow meta-knowledge of a retrieved response element R^R should be a function of the distance computed by the matching criterion.

$$\beta_j^R = S(D) \qquad \ldots(2)$$

Conceptually, the *similarity measure function S()* should be a monotonic function of distance *D()* with negative derivative.

Inflow Expectation relates to the inward communication of the meta-knowledge into the memory system. An external querying system supplies the stimulus elements and the additional significance level of each stimulus element. Outflow Expectation relates to the outward communication from the memory where the external querying body is given back not only the retrieved measurements but also the meta-knowledge *confidence* about the status of the retrieved content. As we will see, both of the transfers are essential in the context of imperfect knowledge transaction.

The above expectations now essentially constitute the behavioral definition of a memory system which incorporates possibility of dynamic search localization of the given measurements. In the rest of this paper such a memory will be referred as *Attentive Memory*. In section 3 we will investigate the limitations of current techniques in realizing such an Attentive Memory. Section 4 presents the a computational technique MHAC which can realize both of these expectations of an attentive memory.

Search Types: Now we identify some search classes that stem from the various forms the meta vector can assume. The general class is where the elements of the meta-vectors are analog valued. Two special cases are (i) where the vector elements have binary enumerations denoting a pixel's status as completely within or outside of the window of focus. This matching criterion can be stated alternately with a scope limiter set F^Q instead of N on the summation ($F^Q \subseteq N$), where N contains the elements which has $\beta_i = 1$. The other case is where (ii) all the vector elements have a unary value 1. In the subsequent discussion we will refer to the general class (denoted by equation (1)) where meta-vectors can have any analog value as Type-A (analog), and the above two special classes of it as type-B (binary) and type-U (unary) matching criteria.

Clearly, for type-U memory the attention is fixed for all elements. This is the case with conventional AAMs. On the other hand, type-A and B memory provides the opportunity to dynamically vary attention over the pattern elements by varying the β values. In Type-B memory (where β is either 0 or 1) the retrieval corresponds to dynamically constructed sub-pattern search. Our objective is the realization of type-A and/or Type-B memory.

3 CURRENT APPROACHES

The computational techniques available today can be classified into two major categories (i) associative computing based, and (ii) procedural search. Below, we analyze each of these approaches in performing search of Type-A or B.

3.1 Associative Search

The optimization criteria of existing neural models directly belong to type-U category. Models those use Hebbian class of learning maximize global dot-product of the patterns [Koho89, CGMR92, Klop87]. On the other hand, the models those use LMS class of learning maximize global mean square error [WiHo60]. There are also few other variant measures (for example *entropy, likelihood-ratio*). Hopfield has given a unified perspective to analyze all the matching criterion of conventional neuro-computing. He demonstrated that all the neural networks minimize some form of energy function [Hopf82, Oja82].

However, their fundamental limitation in supporting either type-A or B search arise from the following commonalty. The key features to note in *all* of the functions those are used by current neural network models are (i) the set operator is a summation process $M \equiv \sum$, (ii) the scope N of the set operator is all inclusive and is based on entire element space, and (iii) the modulator function is only the function of distance $\delta \equiv dist(.)$. These properties of existing neural computers together makes them a type-U memory.

The reason that conventional AAMs cannot support dynamically changeable attention lies right at the heart of conventional neurocomputing: the *scalar product rule of synaptic efficacy*, which is the basic building block of almost all conventional AAMs. In [Khan95] it has been shown that:

Theorem (limitation of existing AAMs): *An associative memory constructed by interconnecting cells with the scalar product rule of synaptic transmission specified by the equation below can not realize the retrieval of type-B, or type-A. Where, f() is any single variate function, and s_j is a real valued number in the range I=[0,1], and the weights w_{ij} contains the learned pattern.*

$$r = f\left(\sum_{i}^{n} w_{ij} . s_j + b_i\right) \quad \text{....(3)}$$

The fact that a neural network can not dynamically localize search can be intuitively explained in several ways. The discrete sum of equation (3) is the foundation stone of the synaptic efficacy rule. But, any finite summation requires all its inputs to be present. A summing output can tolerate some random statistical distortion of the input values but it can not tolerate deliberate (full of partial) absense of inputs[4]. For example, if half of the inputs are missing then the generated sum will be far away from the exepected (expectation developed during learning) sum and will require fundametally different treatment of the inputs. A second perspective will be given in section 4.1 when we discuss representational limitations of scalar neural networks.

Notably, despite the immense volume of work in Neural Networks in last 20 years, according to the knowledge of the authors, no literature mentions or analyzes such critical limitations of AANs and AAMs. Interested readers may want to read [Khan95] which contains revealing analysis of the limitations and capabilities of current AAMs and ANNs.

3.2 Procedural Search

The principal challenge for the procedural approaches in type-A/B/U searches is the computational complexity. Misnky & Papert investigated the exact match as well as best match versions for the matching criterion of type-U [MiPa72] case by case for all the major classes of procedural search algorithms. [Khan95] has extended the analysis for type-B and type-A search. Appendix-A provides a table summarizing the complexities of principal known strategies.

Intutively, the efficiency of these algorithms is derived from the fact that the words can be ordered according to some linear measure. A linear ordering of the patterns can make search a logarithmic process. However words can not be sorted in any meaningful way if the *n*-digits, those make the words, themselves are not ordered. This is the case for multidimensional patterns.

4 Sherrington's [Sher06] observation on the existence of some form of integration process in the nervous sites is generally used to rationalize the use of linear weighted sum as something fundamental to conventional artificial neural computing. However still now the theory itself has not been decidedly validated or refuted. More importantly, our suggestion of weighted average does not imply the absence of integration.

Only in the case of type-U exact match, an arbitrary order of the digits can be assumed. This is because, any ordering converges to the same solution for tupe-U exact match. However once the problem becomes that of best match, all the digits require equal importance (errors at all digits bear equal weight). So the imposed ordering becomes meaningless. The problem grows another step out of hand for type-A and type-B search. These searches not only defy any pre-imposed order but impose its own new order every time a new query is launched with a new meta-vector.

3.3 Associative vs. Procedural Search

The distributed associative search technique based on artificial neural computing, is fundamentally different from the conventional search. It will be relevant to look into two important distinctions. On one hand, artificial associative memories are inherently best match machines. Exact match is guaranteed only under restricted situation. In contrast, conventional algorithmic search techniques are primarily exact match machine. On the other hand, retrieval in associative memories is generally faster.

The information in an image is representationally large but sparse in content. Thus, from both of these distinguishing points, associative search better suits image application. However the inability of previous associative memories to perform type-A and B search becomes an insurmountable obstacle to their usefulness for image specially if the constraint of real-time processing is added.

4 COMPUTATIONAL MODEL

This section now provides the computational embodiment of the bi-modal memory defined in section 2, and shows how such a memory can actually be realized within neural network like adaptive cellular computation paradigm. The proposed MHAC is conceptually based on optical holography [Gabo48, Weny78]. Further details can be found in [Khan95]. In the brief description of this section, first in section 4.1 the representation is explained. Then in sections 4.2 and 4.3 respectively, the encoding and decoding techniques are explained. Together, this computing model realizes the bi-modal memory with attention and confidence and is capable of meeting both the expectations.

4.1 Representation

In the holographic approach a multidimensional complex number is used as the computational representation of each element of information. Each α_k is mapped onto a set of phase elements $\theta_{j,k}$ in the range of $\pi \geq \theta_{j,k} \geq \pi$ through a mapping transformation $m^{+\alpha}(x)$, and corresponding meta information β_k is mapped as its magnitude λ_k through another transform $m^{+\beta}(x)^5$.

$$s_k = (\alpha_k, \beta_k) \Rightarrow \lambda_k e^{\left(\sum_j^{d-1} \hat{i}_j \theta_{j,k}\right)} \qquad \text{....(4)}$$

Where, each element $s_k(\lambda_k, \theta_{1,k}, \theta_{2,k}, \ldots\theta_{d-1,k})$ is a vector inside a unit sphere in a d-dimensional spherical space. Each $\theta_{j,k}$ is the spherical projection (or phase component) of the vector along the dimension $\hat{i}_j$. This computational representation is called **multidimensional complex numeric** (MCN) representation of information. Fig-3 shows 4 MCN elements A, B, C, and D on a 3D space with a projection.

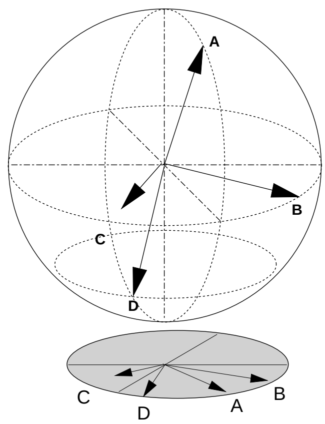


Fig-3 Points on Hyperspherical Surface

5 The inverse transformations to revert back from MCN representation are respectively denoted by $m^{-\alpha}()$ and $m^{-\beta}()$.

The function $m^{+\alpha}()$ should be single valued and continuous. For discrete inputs, continuity is required at the defined points. Preferable this mapping transform should maximize the symmetry at the phase domain. Phases of the points A,B,C and D in Fig-3 represent their measurement-values in the 3D MCN space.

Almost any positive valued assignment function can be used as $m^{+\beta}()$ as long as they conform to the following two constraints. First, elements with same magnitude (equi-significant) are required to contribute equally to the subsequent matching. Secondly, elements with 0.0 magnitude (at the center) should have no effect on the outcome of the computing irrespective of the phase. When all the elements of a pattern are placed on the surface of any internal sphere it represents equal importance for all of them. Such iso-surface case is functionally equivalent to that of conventional AAMs. In Fig-3 the points A,B, and D have meta-value 1 and point C is of lesser significance.

Combined representation: Thus, each of the information elements is represented as a vector bounded in the unit multidimensional spherical space. A stimulus pattern is computationally represented as:

$$[S^{\mu}] = \left[\lambda_1^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \theta_{j,1}^{\mu}\right)}, \lambda_2^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \theta_{j,2}^{\mu}\right)}, ..\lambda_n^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \theta_{j,n}^{\mu}\right)} \right] \ldots (5a)$$

A similar mapping on the external scalar response field intensities provides the response representation:

$$[R^{\mu}] = \left[\gamma_1^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \phi_{j,1}^{\mu}\right)}, \gamma_2^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \phi_{j,2}^{\mu}\right)}, ..\gamma_m^{\mu} e^{\left(\sum_j^{d-1} \hat{i}_j \phi_{j,m}^{\mu}\right)} \right] \ldots (5b)$$

Here, the phasor ϕ represents the measurement of the retrieved response and γ represents the expected *confidence* (system assigned significance) on ϕ.

Since, we have introduced the holographic representation which is complex valued, it will be interesting to add some perspective on its evolution._Complex number is not completely an alien concept in neuro computing. In 1990, Sutherland [Suth90], in his pioneering work presented the first truly holographic associative memory which use holography like complex representation. On the other hand, the term 'holographic' had often been associated with some early AAMs (such as Willshaw's correlograph). However, these models resemble very little with the memory proposed here. These early memories are based on scalar representation and thus suffer from the same limitations of today's ANNs. More recently Timothy Masters [Mast94] reported another 2D-complex valued network with a learning algorithm analogous to Backpropagation. However both of these complex representation based works remained focused on the relative performance issue of these networks as a conventional neural adaptive filter (with type-U retrieval). None of these works explored the profound and fundamentally different attention phenomena (type-A/B retrieval) associated with such representation.

The first artificial system to demonstrate associative recollection phenomena ever, optical holography itself [Gabo48] is a complex valued computation mechanism. When pioneering researchers ventured to recreate such fascinating optical transforms artificially on digital computers, they embraced some simplifications in digital adaptation. One of those early simplifications was the use of scalar numbers instead of complex ones. The main enchantment and research emphasis of this period was artificial learning rather then the entire process of artificial associative recollection. Consequently, early researchers focused on the correlation transform. In subsequent years our understanding of computational learning grew. But subsequent research stayed affixed with the simplified representation. From the evolutionary perspective of artificial associative computing, this particular work is a visit back to the lost dimensionality of representation; and a step beyond. It explores further into a computational model based on multidimensional phasor (instead of only 2-D phasor) representation.

4.2 Holographic Enfolding

In associative memory information is stored in the form of associations. In the encoding process, the association between each individual stimulus and its corresponding response is defined in the form of a correlation matrix computed by the inner product of the conjugate transpose of the stimulus and the response vectors:

$$[X^{\mu}] = [\overline{S}^{\mu}]^{T} \cdot [R^{\mu}] \qquad \ldots.(6)$$

If the stimulus is a pattern with n elements and the response is a pattern with m elements, then $[X]$ is a $n \times m$ matrix with d-dimensional complex elements.

A suit of associations derived from a set of stimulus and corresponding response is then "enfolded" into the following correlation matrix X. The resulting memory substrate is referred as Holograph.

$$[X] = \sum_{\mu}^{P} [X^{\mu}] = \sum_{\mu}^{P} [\overline{S}^{\mu}]^{T} [R^{\mu}] \qquad \ldots.(7)$$

4.3 Holographic Regeneration

During recall, the query stimulus pattern $[S^{e}]$ is represented by:

$$[S^{e}] = \left[\lambda_1 e^{\left(\sum_{j}^{d-1} i_j \theta^{e}_{j,1} \right)}, \lambda_2 e^{\left(\sum_{j}^{d-1} i_j \theta^{e}_{j,2} \right)}, \ldots. \lambda_n e^{\left(\sum_{j}^{d-1} i_j \theta^{e}_{j,n} \right)} \right]$$

The decoding operation is performed by computing the inner product between the excitatory stimulus and the correlation matrix X:

$$[R^{e}] = \frac{1}{c} [S^{e}] \cdot [X] \qquad \ldots.(8)$$

$$where, \quad c = \sum_{k}^{n} \lambda_k$$

Although, the above computation appears analogous to conventional associative computing paradigm, but it process the measurement component of information in a fundamentally different way. Next section explains the fundamental distinctions that make this new parallel and distributed computing paradigm capable of supporting type-A & B RCA search.

4.4 Distinction of Holographic Computation

The above encoding and decoding algorithm can be realized as a cellular automata. Just like conventional neural networks, each cell will realize the complex transfer function of the form below:

$$\dot{z}_i = \sum_{j}^{n} \dot{w}_{ij} \dot{s}_j \ldots. \qquad (9)$$

The transformation it realizes on the measurement component of input information is fundamentally different from that of any existing AAM. Let, $\dot{w}_{ij} = \| w_{ij} \| e^{-\hat{i}\omega_{ij}}$. Then, the transformation between the measurement components of input and output is given by:

$$\phi_i = \cos^{-1}\left[\frac{1}{c} \sum_{j}^{n} \| w_{ij} \| \cos(\theta_j - \omega_{ij}) \right]$$

$$\text{where, } c = \sum_{j}^{n} \| w_{ij} \| \qquad \ldots.(10)$$

For comparison, the scalar product rule of synaptic efficacy used by conventional AAMs is given below with equivalent notations:

$$\phi_i = f(y_i), \quad y_i = \sum_{j}^{n} w_{ij} \theta_j + b_i \qquad \ldots.(11)$$

This new transfer function has three distinguishing characteristics. The first is that the transfer function is a weighted *trigonometric (cosine) mean* function, in contrast to the conventional *weighted sum*. Secondly, that there is no explicit activation function. The third distinguishing feature is that instead of a cell wide single threshold, each of the synaptic inputs to the cell has its individual threshold.

The finite summation process used by the transfer function (11) is tolerant to statistical distortion, but is not tolerant to selective and deliberate loss of inputs. But a mean process such as (10) is robust in both the senses. This is the key distinction that enables search localization (RCA type B/A search) in holographic memory.

The other two distinctions together determines the mapping ability of holographic memory and the specific nature of the discriminating hyperplane that distinguishes learned classes. For a holographic cell the trigonometric transformation pairs serve as the local non-linearity, as opposed to the global non-linearity used by conventional neurons. Such localization of non-linearity is essential for dynamic match localization.

Holographic cell can be distinguished from a representational perspective also. If an element is designated as 'dont-care', it should be represented in such a way that all other valid enumeration values should be equipotential from the point of 'dont-care'. Conventional AAMs have a scalar state space. But, on such one-dimensional state-space it is not possible to assign a point which is equidistant from all the possible enumerations of an analog measurement. MCN representation solves this problem by creating hyperspherical state space (Fig-3) where the center creates the representation for an unbiased 'dont-care'.

5 PERFORMANCE ANALYSIS

The basic associative retrieval characteristics, the ability to satisfy the both the behavioral expectations as outlined in previous section, as well as its performance have been formally analyzed in [Khan95]. We briefly state the results.

Result (accuracy of retrieval): *For a MHAC specified with equations (6), (7), and (8) with n stimulus elements and p stored patterns the maximum distortion for (i) $p \gg 1$, and (ii) the input elements are uniformly distributed in phase space is:*

$$|\Phi_{error}|_{\max} = \sin^{-1}\left(\sqrt{\frac{pw}{n}}\right) \qquad \text{....(12)}$$

Here, **w** refers to the 'porosity' of the attention distribution:

$$w = \frac{\left[\sum_k^n(\lambda_k)\right]^2}{n.\sum_k^n(\lambda_k)^2} = \frac{[E\{\lambda\}]^2}{E\{\lambda^2\}} \qquad \text{....(13)}$$

Equation (12) shows that the focus can be effectively (almost linearly) compensated with higher *n* or lower *p*. This result is very significant, because even for a fixed size problem it is possible to design a network with exponentially higher effective stimulus length (*n*), by techniques such as higher order encoding [Khan95, Suth90].

Pattern deviation causes error which is equal to the directed sum of the deviations of individual pattern elements. Thus it linearly moves away from target patterns, with mean of the shift in query $\Phi_\varepsilon = n.\bar{\varepsilon}_i$, when $\theta_i^t - \theta_i^e = \varepsilon_i$ when the error is small. It can be geometrically shown that the error due to pattern deviation grows in the order of $\sqrt{2\sin\left(\frac{\varepsilon}{2}\right)}$.

6 EXPERIMENTS

The characteristics of this memory have also been empirically validated with extensive computer simulation [Khan95]. In this paper, we present the following three characteristics which are specifically important for the real-time image processing. (i) Focus characteristics show the unique ability of search localization. (ii) Loading characteristics show the extent of pattern space that can be searched by a single retrieval of holographic memory. Finally, (iii) the scalability experiments demonstrated the potential ability of this memory to support pattern matching in massive image collections. In these experiments we have used the following parameters.

Definition (Accuracy): *Accuracy of retrieval (SNR) is measured as the peak signal to noise ratio in the measurement component of information over all the elements.*

$$SNR = 20\log\frac{2\pi}{mse} \qquad mse = \sqrt{\frac{1}{m}\sum_i^m[\phi_i^\mu - \phi_i^{T(\mu)}]^2} \qquad \text{...(14)}$$

Definition (Load Factor): *The loading factor (L) is defined as the ratio of the total number of elements (n) in the patterns to the number of stimulus response associations (p) encoded.*

$$L = \frac{p}{n} \quad \text{....(15)}$$

The definition of *focus* (QPD=*1-w*) corresponds to (13). In all these experiments, pattern elements have been generated randomly with *clipped Gaussian*[6] distribution to match natural distributions (such as image intensity). Standard deviation has been varied to generate data with different asymmetry characteristic.

Besides investigating the general relationship among these parameters, these experiments simultaneously examine the specific ranges of these critical parameters within which an effective and cost efficient attentive memory can be constructed.

6.1 Analysis of Experiments

Focus characteristic: For this experiment, a set of holographs have been generated with various numbers of encoded patterns. After the training, by using a randomly selected sub-set of pixels from each of the originally stored patterns as the query pattern, recalls have been performed. The focus strength has been controlled by setting $\lambda_i \approx 0$ for a specified number of random elements in the query pattern. Fig-4 shows the typical performance (in terms of percentage of dynamic error) with the smooth variation of focus strength of the query pattern. The four curves in this graph show the focus characteristics for three different load factors L= .02, .04, .06 and .08. For all cases the patterns have length n=1000.

As can be observed in both of these plots, a typical focus characteristics curve is monotonic and demonstrate three distinct zones. First a high-performance zone, then a linear-zone, and finally a cut-off zone. The high performance zone corresponds to RCA type-U search performance. This zone characterizes regular AAM like high focus and is featured with high accuracy. As evident by the accuracy level of this zone, an attentive memory, even when it acts as a regular type-U memory, it far exceeds the retrieval accuracy of most other analog AAMs. This zone demonstrates accuracy so high that the original encoded pattern suffers less than 2-3% magnitude error (which in other word means over 30 db accuracy). In linear zone, the accuracy gracefully decays with the focus strength. Analytically the characteristics of this zone correspond to equation (12). As can be seen, focus strength can be reduced almost as low as 0.1 till the accuracy falls below 20db (about 10% magnitude error). In marked contrast, a regular AAM shows avalanche degeneration of performance when the focus strength approaches just 0.6-0.5 [TaJo90].

6 Because of the circular nature of phase space, only those random generations have been used which falls between 0 to 2 Π. The phase values, otherwise have been generated with mean 0 and various standard deviations (sd).

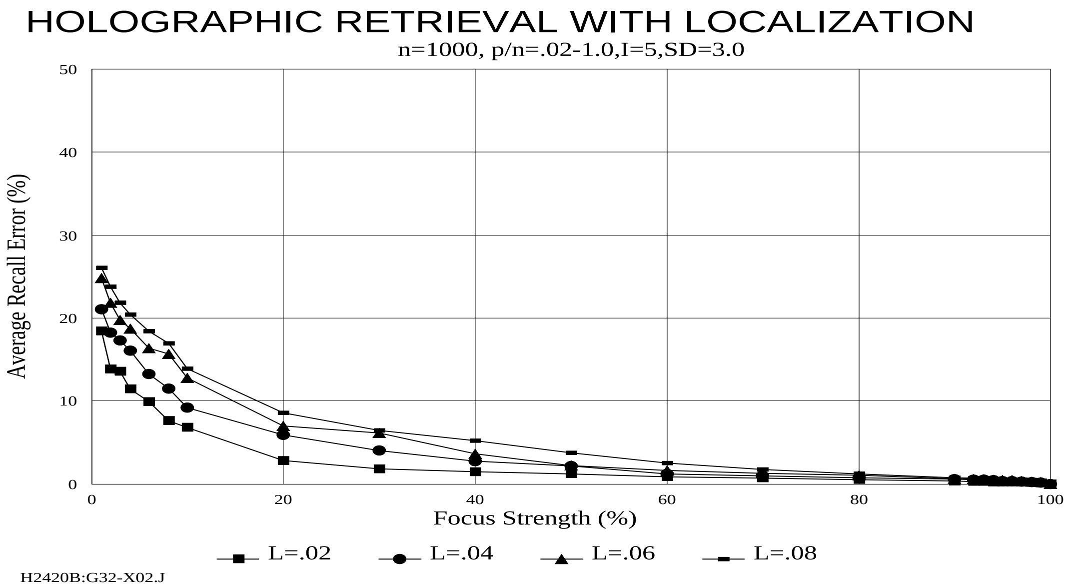

HOLOGRAPHIC RETRIEVAL WITH LOCALIZATION
n=1000, p/n=.02-1.0,I=5,SD=3.0
Average Recall Error (%)
0
10
20
30
40
50
0
20
40
60
80
100
Focus Strength (%)
L=.02
L=.04
L=.06
L=.08
H2420B:G32-X02.J

Fig-4

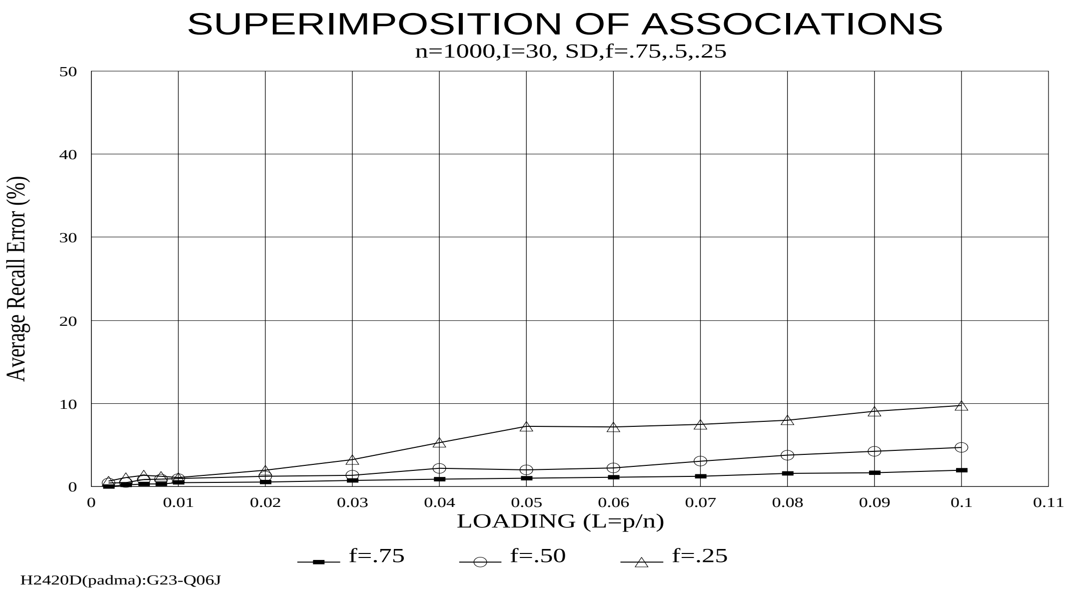

SUPERIMPOSITION OF ASSOCIATIONS
n=1000,I=30, SD,f=.75,.5,.25
Average Recall Error (%)
0
10
20
30
40
50
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.11
LOADING (L=p/n)
f=.75
f=.50
f=.25
H2420D(padma):G23-Q06J

Fig-5

Loading characteristic: Fig-5 shows a typical loading performance, when more and more patterns are added to the same holograph. It plots the average error (y-axis) against the load factors (x-axis) for three RCA type-A cases with focus strengths f=.75, f=50, and f=.25.

As shown in this plot, a typical loading characteristics curve shows monotonically decreasing performance with increased load factor. Quantitatively, for f=.25, the RCA type-A retrieval accuracy drops to 20 db, while the load factor reaches .07. Typically, a load factor of .03 to .10 can be reached maintaining 20 db performance with f=0.3-0.1. This experiment shows that an enormous number of pattern associations can be stored and retrieved from a single holographic memory. For example, a load factor of .02 means that about 5,000 images of size 512x512 can be enfolded into a single holographic attentive memory and can be searched with RCA type-A capability only at the cost of only one complex inner product[7]. Table-1 lists few other loading scenarios.

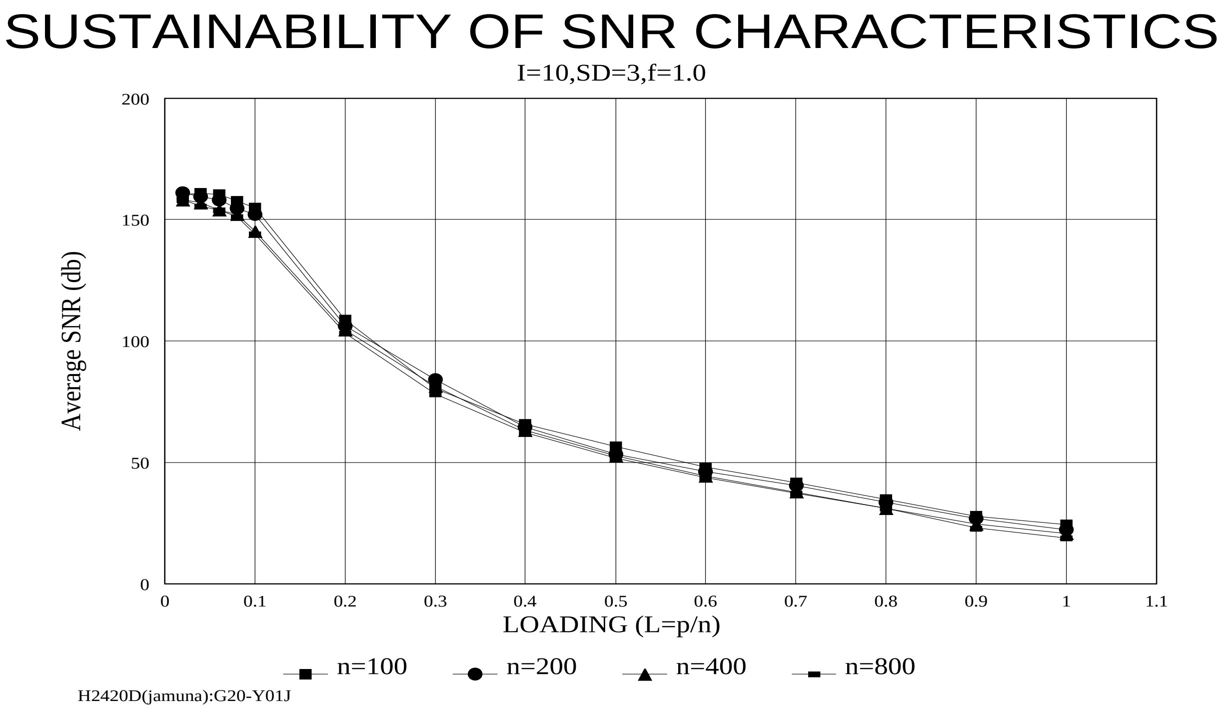


Fig-6

Scalability Characteristics: Fig-11 shows the scale invariancy of loading characteristics for exponentially varying *n*=100, 200, 400, and 800 for a fixed *p/n* ratio. As evident, although the problem scale varies exponentially, these curves overlap on each other demonstrating the performance invariancy of focus characteristics with respect to the scale of *n*.

This result conforms to the analytical derivations of equation (12) which show that most of the performance characteristics are related to the load ratio *p/n*, rather than *n*. Both analytical as well as experimental results indicate the enormous algorithmic scalability of the attentive memory with sustainable performance.

Performance Characterization: Also extensive

7 The size of the matrix is nxm, where m is the response label size and typically is $O(\log p)$, where as a procedural best match search is nxp.

simulation has been performed which determines an *operational range space* (ORS), within which it is possible to guarantee a target performance. As a typical ORS an encoded holograph with load factor .08 or less can guarantee retrieval of patterns from 10% focus windows with more than 20 db accuracy on the average for a image set.

Loading is closely tied to the space efficiency of any associative memory. The dimension of a holograph is determined by the length of the stimulus (*n*) and response patterns (*m*). Load factor provides an estimate how many such patterns can be enfolded on a single holographic memory. Table-1 shows typical estimates on the number of patterns that can be stored (and queried) for few image sizes. However, for patterns with limited size, load factor is not necessarily a hard limitation. The number of stored patterns *p* for relatively small patterns can be increased by higher order encoding.

n	*L*	*p* (1st order)
160×120	.04	768
256×256	.02	1310
512×512	.02	5120
1024×1024	.01	10240
1024×1024	.02	20480

Table-1 Typical Memory Loading

6.2 An Associative Search and Retrieval Example

<u>The Scheme:</u> In this section we now demonstrate the novel capabilities of this holographic image pattern search system with a simple example. Our objective is to search for a specific example pattern in a collection of images. In this scheme, we first "enfold" these images into a holographic memory. When a sample image pattern is given, instead of matching it with every stock image, we now 'resonate' the sample image with the holograph. The localization capability of holographic pattern matching net allows any object or features specifiable by the subset of the pixels to be specified as the basis of similarity in pattern matching. It also allows the definition of fuzzy features or objects. Each decoding operation generates a resonance feedback and a direct construction of a *response label pattern* (RLP) associated with the best matching stimulus pattern. Strong feedback indicates the potential correct hit. By shifting the location of resonance it is also possible to obtain match at spatially translated locations.

<u>Attention Based Query:</u> For this prototype example we have enfolded 64 CT-scan images (of dimension 256x256 pixels) into a holograph. The first two sets of experiments demonstrate the performance of this enfolded holograph for retrieval of various types of localization. Fig-7 respectively show the scheme of attention windowing for these tests. Fig-8(a) shows the result when during the retrieval the attention on the sample pattern has been varied from 60x60 pixels to 110x110 pixels in 6 steps. It is called ZOOM test. Fig-8(b) shows the result when instead of using attention windows of different sizes, they were selected from different locations of the query image frame (with fixed size 55x55 pixels). Both of the figures plot the average retrieval accuracy in the bar-plot (left y-scale) and the corresponding size of the focus window in the line-plot (right y-scale). The three bar sets denote the accuracy for three assertion values for the air segment used during training.

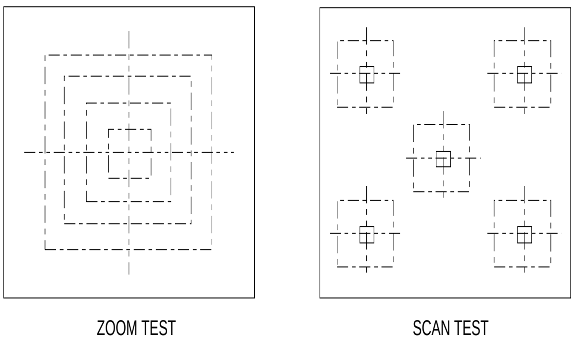


Fig-7 Characterization Tests

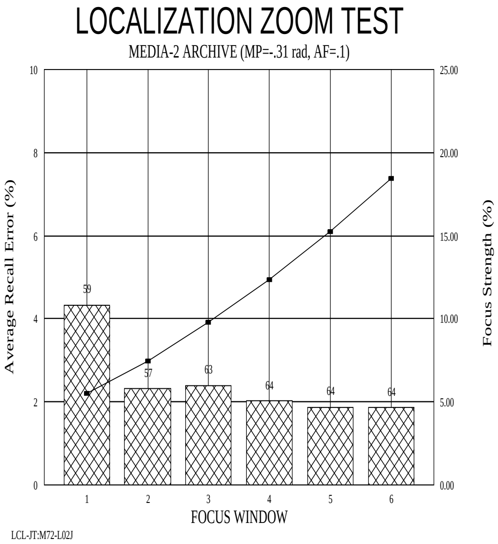


Fig-8(a) Zoom Test

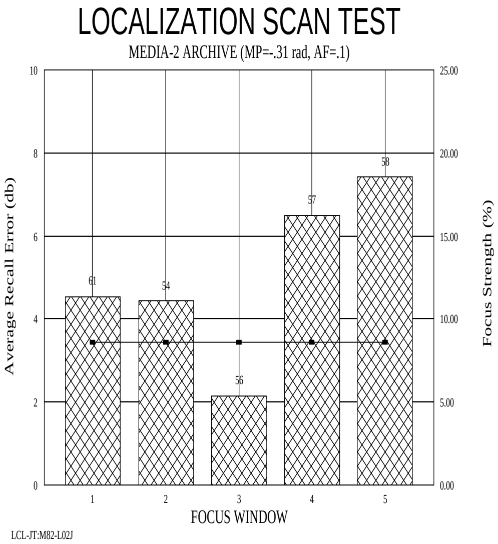


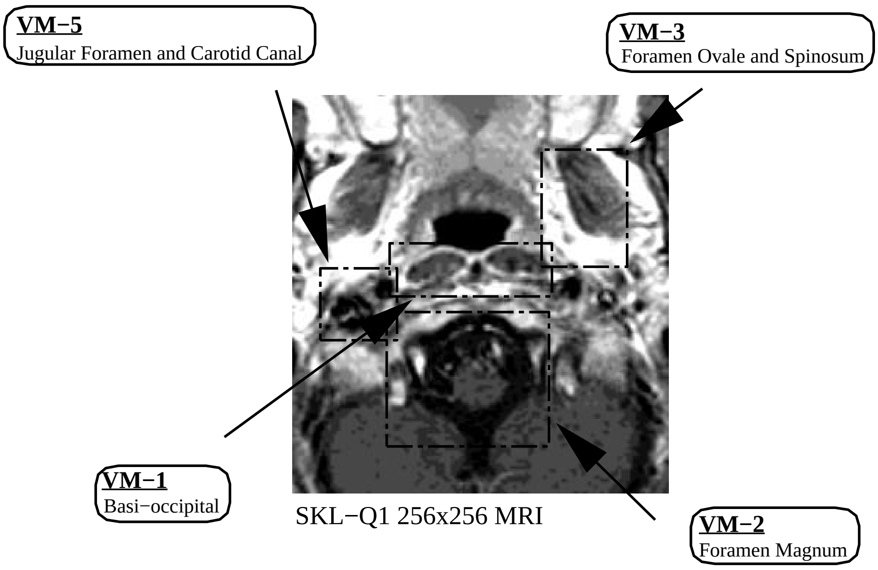
Fig-8(b) Scan Test

As evident, the retrieval accuracy remained well above 20-25db (10% magnitude error) even for attention window lower than 10% of the total frame. Also the relationship between the performance and focus is near-linear. This is remarkably different from conventional AAMs, which show cutoff when the number of dissimilar bits between the target and sample images reaches 50% irrespective of their sophistication. In Fig-8(b) it is also evident that the accuracy of retrieval is in general high in the center (number 3) window. This conforms to the general nature of CT and MRI images where the objects of cognitive importance are always centered.

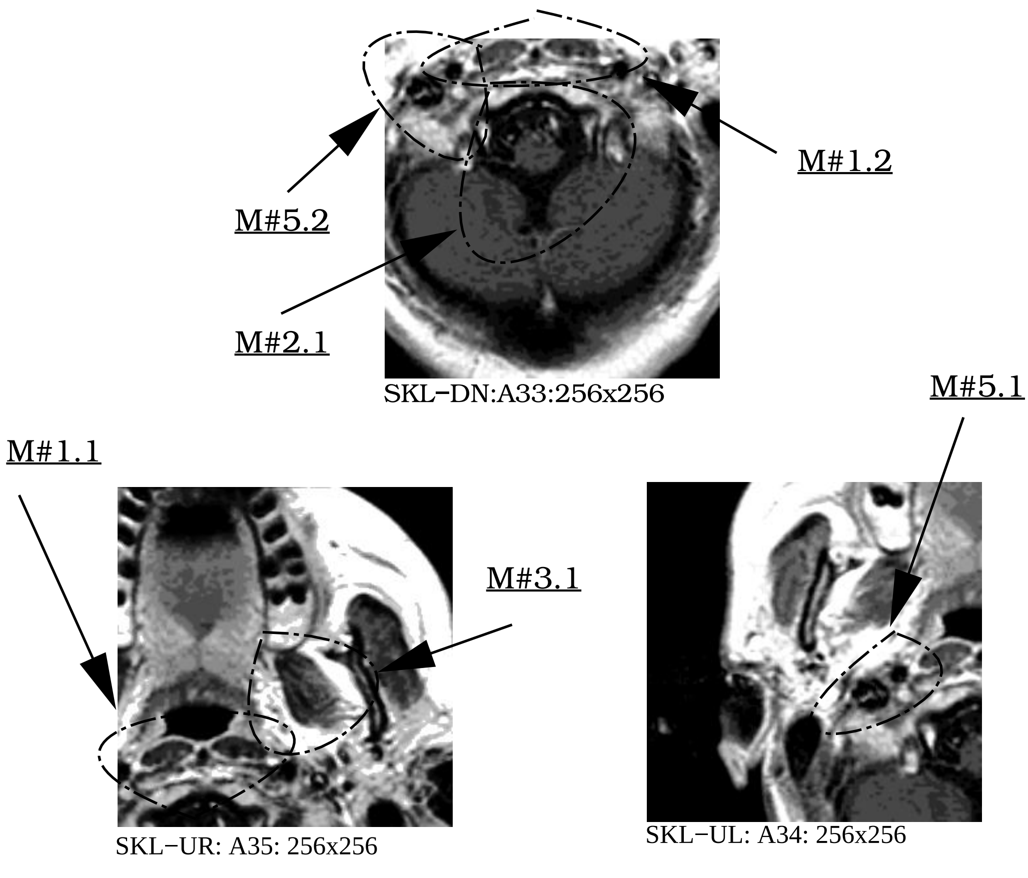
Fig-9(a) shows a sample pattern with 4 attention windows (VM-1, VM-2, VM-3, VM-5). A clipped Gaussian range filter analogous to (8) has been used to gracefully smoothaway the boundaries of attention cover. The images retrieved by the search method corresponding to these attention shields are shown in Fig-9(b),(c), and (d). For example, for the attention shield emphasizing *Jugular-Foramen and Carotid Canal* (VM-5) region of the sample, two correct matches M#5.1 at plat #34, and M#5.2 at plate #33 have been found. A conventional AAM cannot be used for such search. Because, in the first place, without the ability of match space localization, it will never converge to different matching pattern images beginning from the one sample pattern image.

MNC feedback based interactive Search: The critical role of MNC in associative pattern search is also demonstrated in this retrieval example. As can be seen in Fig-9, that the object of interest in the sample pattern is spatially translated compared to the matching sub-pattern in the target. Any translated search, requiring multiple decoding needs a feedback to evaluate the presence of match. For associative search, this feedback is needed in addition to the regenerated retrieved pattern.

We have used a simple *spirally descending search* (SDS) on a 10x10 pixel grid space to converge to the correct spatial location in the holograph. The match was sensed by a process of resonance through MNC feedback. It is illustrated though an example search with pattern VM-5. Fig-10 plots the MNC feedback obtained from the memory by resonating the VM-5 sample at various spatially translated grid locations of the holograph. Sharp resonances are detected from the holograph at displacements -10,100 and 130,-20. These indicate the presence



Fig− 9(a) Sample Queries in MEDIA (Skull)



Fig− 9(b) Retrived Patterns and Matches

of matching patterns in these locations in the images enfolded beneath. As evident visually in Fig-9 that these correspond to the two correct matches (plates #34, and #33). On one hand, the MNC feedback at the meta-plane provides the resonance strength, on the other hand, the phase plane provides the original response pattern that was enfolded. Conversely, the characteristically low resonance at other grid locations tells external searcher about the absence of matches in those locations.

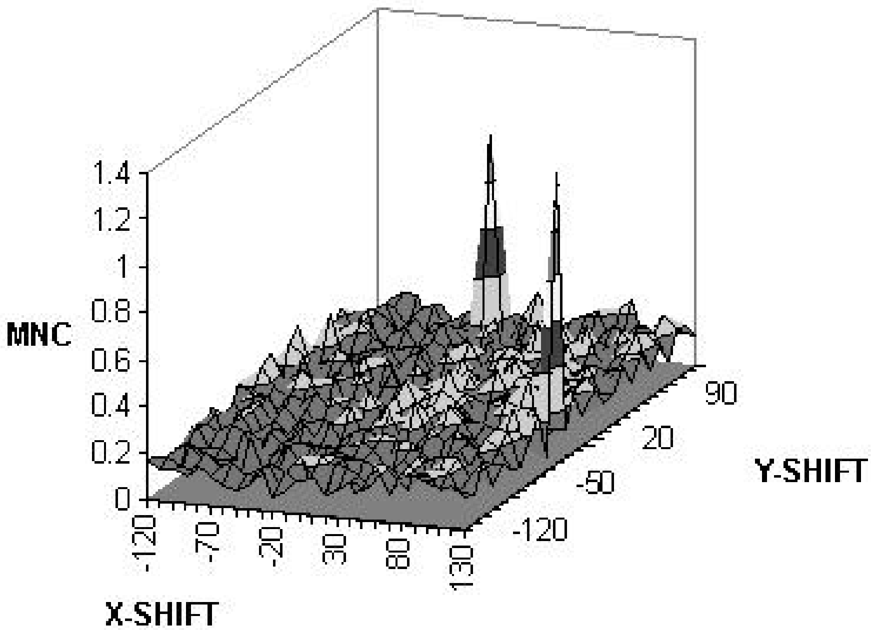


Fig-10 MNC Response at Grid Points

Table-2 Retrieval

MASK#	Object	Match#	Image	xsft	ysft	SNR (db)	MNC
1	Basi-occipital (f=.101)	1.1 1.2	A35 A33	-60 -10	-90 100	36.21 33.42	0.956 0.895
2	Foramen Magnum (f=.122)	2.1	A33	-10	100	32.27	1.123
3	Foramen Ovale and Spinosul (f=.072)	3.1	A35	-60	-90	33.33	1.092
5	Jugular Foramen and Carotid Canal (f=.044)	5.1 5.2	A34 A33	130 -10	-20 100	31.92 37.46	1.095 1.183

The most important advantage of this new search mechanism is evident in this example. In a conventional search, a translated pattern match needs searches in all locations of all the images. In contrast, in holographic method that process is dramatically reduced into the process of searching only on a single holographic plane. The dimension of a holograph is dramatically less than the entire image space. Table-2 shows the search performance for the four regions of objects shown in Fig-9. As evident that whereas the cover (f) for these attention fields are 5-15%, the corresponding response patterns are retrieved with more than 30 db signal to noise ratio.

Feedback like MNC resonance is a natural as well as essential part of any search mechanism that would deal with imperfect information or needs to be interactively controlled by external autonomous systems (an expert system or another controller neural network). Current AAMs do not have this critical capability within its mono-modal framework of information representation, thus can not be used for multistage interactive search.

7 TIMING ANALYSIS & ESTIMATES

Time Complexity: *For search complexity with p patterns (images) each with n elements (pixels), and RLP length m, is* $O(m \cdot n) \approx O(n.\log p)$.

Derivation: Let us consider, that each image has n pixels, and there are p such images. Let us also consider that the length of RLP is m. The search process involves (a) computation of pattern, (b) holographic decoding, and (c) RLP matching. The complexities of the corresponding stages are:

The cost of pattern computation = $O(n)$.

The cost of holographic decoding = $O(nm)$. This is an inner product matrix operation involving complex matrix multiplication.

RLP matching = $O(mp)$. It is a linear search with relatively very small pattern length. Thus, the cost is negligible.

Generally, $m \approx O(\log p)$. Thus the complexity of the overall search process is.

$$O(n) + O(n.\log p) + O(p.\log p) = O(n.\log p) \quad \ldots(16a)$$

Thus the search *computation enfolding factor* is:

$$\frac{O(n.\log p)}{n \cdot p} = O\left(\frac{\log p}{p}\right) \quad \ldots(16b)$$

Space Complexity: *The search with p patterns (images) each with n elements (pixels), and RLP length m, requires* $M = m.d.(n+p)$ *space for holographic encoding.*

Derivation: Let us consider, that each image has n pixels, and there are p such images. Let us also consider that the length of RLP is *m* and that each complex element requires *d* bytes for representation.

The space, required by the holograph is *m.n.d*. Some additional space is also required by the RLPs. Which is *p.m.d*. Thus the total space requirement is:

$$M = m.d.(n+p) \quad \ldots(16a)$$

In practice $p \approx n$, and 4-12 bytes are sufficient for images with 256x256x256 full colors.

Similarly, the *search space enfolding* factor is:

$$= \frac{n \cdot m \cdot d}{p \cdot n} = O\left(\frac{\log p}{p}\right) \quad \ldots(16b)$$

Parallel Performance: Both the encoding and decoding operations involve complex matrix like computation and can be parallelized with almost linear speedup on conventional parallel machnies. Very recently a parallel system has been implemented on the IBM SP2 MIMD machine at the Maui High Performance Computing Center. Fig-11 demonstrated the speed up and efficiency of the parallel system that has been achieved. A maximum constraint block parallel mapping has been adoted to determine the worst case. Here each of the associations are encoded or decoded completely before a second association is processed. The number of processors has been varied from 1 to 24 for both encoding and retreival operations. As observed, almost linear speedup can be achieved.

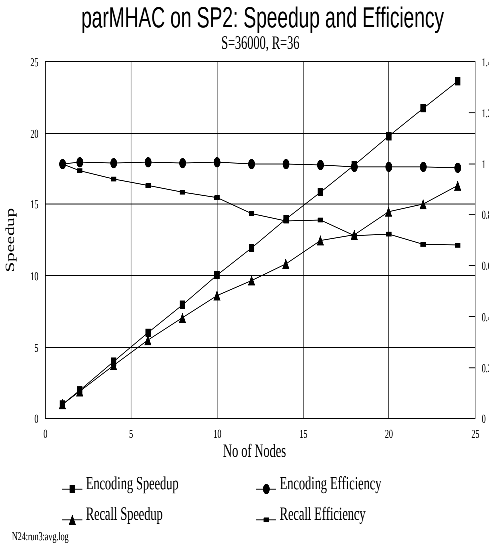


Fig-11 Algorithmic Scalability

8 CONCLUSIONS

In recent years, optical holography has attracted great interest for its high information storage density and access speed. We have expanded the investigation into more advanced use of holographic memories based on their associative recollection ability. In this paper we demonstrate how memories based on holographic principles can be used to perform associative sub-pattern matching with the capability of dynamic localization at amazing speed by manipulating the modulus constraints of holographic representation.

To search into massive information space, this technique adaptively (and automatically) "enfolds" the massive search space into a single holographic memory substrate of exponentially reduced dimension. During search, instead of actually accessing the large raw pattern space, the holographic regeneration technique "simulates" the search through associative recollection into the holograph.

Associative Pattern Matching: As demonstrated, this memory can perform associative recollection using an effective cue as small as 5-10% of the pattern size. Whereas, existing AAMs require the valid part of the query patterns even theoretically to be at least 50% (in practice they generally stop recollection at about 70%) the original pattern size [Tajo90, Hopf82]. This is a fundamental improvement over the capabilities of existing AAMs. The performance ranges given by the ORS experiments for focus (f=1-.1) and loading characteristics (L=.01-.04) suggest the designability of real associative computing based search engines with this model. As an example, a sample estimate of the performance (details shown in appendix-B) suggests that about 100-75 GB image pattern (about 300,000 images) can be potentially searched in 3-10 seconds using holographic technique on existing general purpose machines. It is about 80 times faster than what is possible by other methods.

Computational and Architectural Scalability: Before we comment about the structural scalability, it is important to investigate the brittleness of the solution. Any pseudo exact/ optimization algorithm, besides the sustenance of speedup (architectural or structural scalability), also requires sustenance of the quality of the solution for scalability (computational scalability). Both analytical and empirical evidences obtained in this work suggest that the performance of the network is sustainable for larger scale of the problem size (characterized by n and p). It is a well-observed phenomenon that the scalability of even the other most successful ANN models (such as backpropagation, counterpropagation networks) are rather limited. Not only the amount of computation increases, but also the convergence speed, accuracy of any conventional ANN degrades steeply when problem size increases.

In addition to the demonstrated computational scalability, in a parallel realization of this model on IBM-SP2 machine we have recently demonstrated its sustained linear speedup. In general, two factors contributed to such architectural scalability. The first is its highly structured matrix like operations. The other is the heavy grain complex valued computations.

Hardware Realizability: The method has been analyzed extensively in simulation environment. As evident in equations (6), (7) and (8), that the computational primitives of this memory are based on multidimensional matrix computations. The over all computations and highly structured and repetitive. Such characteristics make this entire model micro implementable with easily cascadable and reusable VLSI blocks. At the macro level, this same property favors its highly parallel and distributed implementation.

In addition, this technique also has excellent potential for fast optical realization. The retrieval process of this memory is little different from that of the storage holographs. In optical implementation, the information can be retrieved by non-mechanical means. Consequently, this new memory also promises information retrieval speed that of storage holographs in the range of several microseconds (which is 100-1000 faster than current CDs).

Applications: Like other AAMs this new technique is adaptive and capable of dealing with imprecise information. In addition, this new memory provides the novel RCA type-A and type-B capability within the framework of associative computing. Consequently, any pattern matching application which requires fast statistical matching with search localization can benefit from this new technique.

It can potentially facilitate the solution of quite a large set of applications requiring both adaptability of model acquisition and dynamic associative recollection. *Multi-object target recognition, content based retrieval in image archive, pattern analysis in multidimensional spectral data, real time speech synthesis* [KhYu94b, ChHs92] are just few of the daunting problems which fit in this class and can directly benefit from this new memory model with attention. It can play an important role for image pattern matching where invariancy can not be obtained easily by any normalized representation of the target object, and sheer rote learning is generally the only resort (this is the case for almost all multi-object targets where moment-based normalized features can not be used). For example, many currently operational target recognition systems relies on matching of the unknown pattern with a massive number of pre-sampled images of the target objects generated at fixed rotation/translation intervals for invariancy. Holographic enfolding can easily enfold all the transformed image patterns into a single holograph and can dramatically reduce the cost of matching (by almost exponential order).

MHAC has already been successfully used to develop the first associative memory based approach for a content-based image archival and retrieval system, which can overcome subjective incoherence of traditional intermediate model based approaches [KhYu94c].

The authors would also like to acknowledge Maui High Performance Computing Center for resource support in parallel implementation. During this research, the first author has been supported by a doctoral fellowship provided by the East West Center of Hawaii. Also, partial support was provided through the DARPA research grant DABT 63-93-C-0056.

APPENDIX-A
ALGORITHMIC DIRECT SEARCH

In this table the complexities of various algorithmic search methods are briefly summarized. Previously, Minsky and Papert [MiPa72] have computed the capabilities of various search algorithms for exact and best match problem. Following analysis is an extension of their result for type-A and type-B searches. The formal specification of the search strategies and the derivations of these complexities can be found in [Khan95]. Here, we assume $p = 2^a$ patterns are to be searched with n elements each (a is index size). For type-B search we will assume focus window size to be c.

The analysis summarized in Table-A.1 clearly shows the limitation of current strategies in handling type-A (or type-B) search. Exhaustive search, (either in raw or in lossless compressed form) remains the only viable strategy.

What is the reason behind the brittleness of conventional logarithmic search approaches? The efficiency of these algorithms is derived from the fact that the words can be ordered according to some linear measure. However words can not be sorted in any meaningful way if the n-digits, those make the words, themselves are not ordered. In the type-U exact match case, an arbitrary discriminate order of the digits is assumed (or imposed). There is no problem so far if it is exact match. Any ordering converges to same solution for such exact match. However once the problem becomes that of best match, all the digits require equal importance (errors at all digits bear equal weight). So the imposed ordering becomes meaningless. The problem grows more out of hand for type-A and type-B search. These searches not only defy any pre-imposed order but impose its own new order every time they perform a query. Thus, logarithmic search strategy falls apart for type-A/B (even when they are only looking for exact match).

Table-A.1 Complexities of Procedural Search Methods

Strategy	**Memory required**	**Type-U (exact)**	**Type-U (best)**	**Type-B & A (exact)**	**Type-B (best)**	**Type-A (best)**
Table	–	1	n	$\frac{1}{2}2^{(n-c)}$	n	–
Lookup	$\geq 2^n$	$M = 2^n$	$M = n.2^n$	$M = 2^n$	$M = n.\sum_{k=1}^{n}\frac{n!}{(n-k)!}$	$M = \infty$
Hashing	$2.n.2^a$	$4+\varepsilon$	-	-	-	-
Ordered set	$n.2^a$	$a.n$	-	-	-	-
Data-table	$n.2^a$	$\frac{1}{2}n.2^a$	$n.2^a$	$\frac{1}{2}c.2^a$	$n.2^a$	$c.2^a$
Exhaustive	$(n-a).2^a$	$\frac{1}{2}(n-a).2^a$	$(n-a).2^a$	$\frac{1}{2}(n-a).2^a$	$(n-a).2^a$	$(n-a).2^a$
Impossible	$< (b-a)2^a$	-	-	-	-	-

APPENDIX-B
PERFORMANCE ESTIMATION

This appendix demonstrates some estimation of the performance for a holographic associative search system. The estimation is based on the ORS results. The estimation process is illustrated through an example image pattern matching problem assuming about 300,000 (512x512 pixels) patterns. (The situation is approximately equivalent to the trademark database in Japan Patent Office, [Kato92]). Detail specifications are given below. Table B.1 lists the computational performance and Table-B.2 provides time estimates on some parallel computers.

Pattern length (n)Pattern/holograph	= 512x512
RLP mod	= 4096 (L=.015)
RLP length (m)	= 3
Complex number b	= 8 bytes
Space s	= 4 bytes/complex
Number of holograph h	= 512x512x8x4 = 8 MBYTE/holograph
Total space S	= 80 holographs/ 320,000 images
Raw Space	= 640 MBYTE/320,000 images
Space Factor (SF)	= 78 GBYTE
	= .008
Retrieval cost Cr	= 12.5 MFLOP/holograph
Encoding cost Ce	= 40 MFLOP/iteration

Table-B.1 Key Results

Number of frames (p)	.3 million frames
Size of each frame (n)	512x512 pixels
Holographic Space	640 MBYTE
Space Enfolding Factor	.008
Speedup	80

Table-C.2 Server Performance

System	Nodes	Power[8] MFLOPS	$T_{retrieve}$ sec/loc	T_{encode} sec/image	$T_{regular}$ sec/loc
HITACHI S-3800/180 (2ns)	1	408	2.41	.97	3.2
NEC SX-3/14 (2.9ns)	1	314	3.14	1.27	4.16
FUJITSU VP2600	10	249	3.95	1.6	5.25

8 As System capabilities are based on their performance against LINPACK Benchmark [Dongarra, 1995]

11 REFERENCES

[Carp89] Carpenter, G. A., "Neural Network Models for Pattern Recognition and Associative Memory", *Neural Networks*, v.2, 1989.

[CGMR92] Carpenter G. A., S. Grossberg, N. Markuzon, J. H. Reynolds, & D. B. Rosen, "Attentive Supervised learning and Recognition by Adaptive Resonance Systems", *Neural Networks for Vision and Image Processing*, Ed. G. A. Carpenter, S. Grossberg, MIT Press, 1992, pp364-383.

[ChHs92] Chang, S.K., Arding Hsu, "Image Information Systems: Where Do We Go From Here?", *IEEE Trans. on Knowledge and Data Engineering*, v.4, n.5, October 1992, pp431.

[Gabo48] Gabor, D., "A New Microscopic Principle", *Nature*, v.161, 1948, pp777-778.

[Gabo69] Gabor, D., "Associative Holographic Memories", *IBM Journal of Research and Development*, I3, 1969, pp156-159.

[HMCC95] Hong, J.H., I. McMicheal, T. Y. Chang, W. Christian and E.G. Paek, "Volume Holographic Memory Systems: Techniques and Architectures", *Optical Engineering*, v.34, n.8, August 1995, pp2193-2203.

[Hopf82] Hopfield, J. J., "Neural Networks and Systems with Emergent Collective Computational Abilities", *Proc. of National Academy of Science*, USA, v.79, April 1982, pp2554-2558.

[Jaco93] Jacobson, M, *Foundations of Neuroscience*, Plenum Press, New York, 1993, pp173.

[Kato92] Kato, T., "Database Architecture for Content-Based Image Retrieval", *SPIE, V1662 Image Storage and Retrieval Systems*, 1992, pp112.

[Khan95] Khan, Javed. I., "Attention Modulated Associative Computing and Content Associative Search in Images", *Ph.D. Dissertation*, Department of Electrical Enginnering, University of Hawaii, July, 1995.

[KhYu94a] Khan Javed. I., and D. Y. Y. Yun, "Chaotic Vectors and a Proposal for Multidimensional Complex Associative Network", *Proceedings of SPIE/IS&T Symposium on Electronic Imaging Science & Technology '94, Conference 2185*, San Jose, CA, February 1994, pp95-106.

[KhYu94b] Khan Javed. I.,& D. Yun, "Searching into Amorphous Information Archive", *International Conference on Neural Information Processing*, ICONIP'94, Seoul, October, 1994, pp739-749.

[KhYu94c] Khan J. I.,& D. Yun, "An Associative Memory Model for searching Image Database by Image Snippet", *Proceedings SPIE Conference on Visual Communication*, VisCom'94, Chicago, September, 1994, pp591-601.

[Klop87] Klopf, A. H., "Drive-Reinforcement Learning: A Real Time Learning Mechanism for Unsupervised Learning", *Proc. of 1st IEEE Conf. on Neural Networks*, Vol.II, N.J., 1987, pp441-445.

[Koho89] Kohonen, T., *Self-Organization and Associative Memory*, 3rd Ed., Springler Verlag, Berlin, 1989.

[KuWo91] Kumar, B. V. K., P. H. Wong, "Optical Associative Memories", *Artificial Neural Networks and Statistical Pattern Recognition*, I. K. Sethi and A.K. Jain (Eds.), Elsevier Science Publisheres, 1991, pp219-241.

[LeUp65] Leigth, E. N., and J. Upatnieks, "Photography by Laser", *Sceintific Amearican*, June 1965.

[Mast94] Masters, T., *Signal and Image Processing with Neural Networks*, John Wiley & Sons, New York, 1994.

[McPi43] McCulloch, W. S., Walter H. Pitts, "A Logical Calculus of the ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics* v.5, 1943, pp115-133.

[MiFa90] Michel, A. N., J.A. Farrel, "Associative memories via Artificial Neural Networks", *IEEE Control Systems*,v.10,no.3, April 1990, pp6-17.

[MiPa72] Minsky, M., S. Papert, *Perceptrons: An Introduction to Computational Geometry*, 2nd ed., MIT Press, Cambridge, 1972.

[Oja82] Oja, E., "A Simplified Neuron Model as a principal Component Analyzer", *Journal of Mathematical Biology*, v.15, 1982, pp267-273.

[PsMo95] Psaltis, D, Fai Mok, Holographic Memories, *Scientific American*, November 1995, pp70-76.

[Sher06] Sherrington, C.S., *The Integrative Actions of Nervous System*, Yale Univ. Press, New Haven, 1906.

[Sinc94] Sincerbox, G. T., "Selectd Papers on Holographic Storage", SPIE Milestone Series, v. MS 95, SPIE Optical Engineering Press, 1994.

[Suth90] Sutherland, J., "Holographic Models of Memory, learning and Expression", *International J. Of Neural Systems*, 1(3), 1990, pp356-267.

[Tajo90] Tai, Heng-Ming, T. L., Jong, "Information Storage in High-order Neural Networks with Unequal Neural Activity", *J. of Franklin Institute*, v.327, n.1, 1990, pp16-32.

[Weny78] Wenyon, Michael, Understanding Holography, Arco Publishing Inc., NY 1978.

[WiHo60] Widrow, B., M.E. Hoff, "Adaptive Switching Circuits", *IRE WESCON Convention Record*, part 4, 1960, pp96-104.