An associative memory model for searching image database by image snippet

Javed I. Khan and D. Y. Y. Yun

Laboratories of Intelligent and Parallel Systems
Department of Electrical Engineering
492 Holmes Hall, 2540 Dole Street
University of Hawaii at Manoa, HI-96822
javed|dyun@wiliki.eng.hawaii.edu

ABSTRACT

This paper presents an associative memory called as ***multidimensional holographic associative computing (MHAC)****, which can be potentially used to perform feature based image database query using image snippet. MHAC has the unique capability to selectively focus on specific segments of a query frame during associative retrieval. As a result, this model can perform search on the basis of featural significance described by a subset of the snippet pixels. This capability is critical for visual query in image database because quite often the cognitive index features in the snippet are statistically weak. Unlike, the conventional artificial associative memories, MHAC uses a two level representation and incorporates additional* ***meta-knowledge*** *about the reliability status of segments of information it receives and forwards. In this paper we present the analysis of focus characteristics of MHAC.*

1. INTRODUCTION

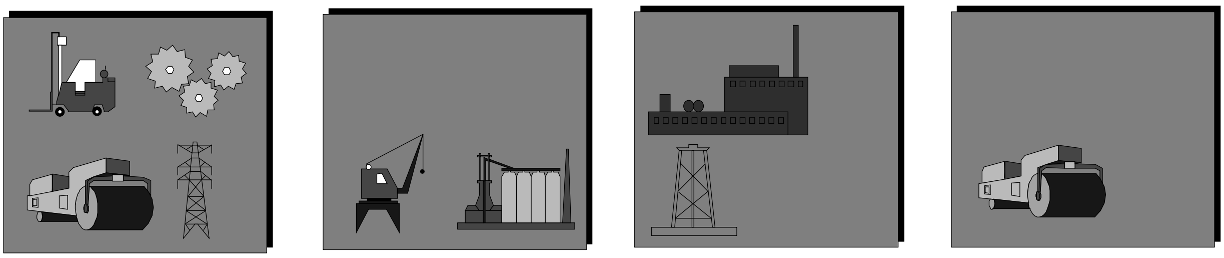
One of the key characteristics of conventional artificial neural networks (CANN) is their robustness and ability to cope with the unreliable information. However, it is important to note the precise nature of imperfect information[6] such a network can manage.

Almost all of the conventional associative memories, such as Hopfield network, Bi-directional Associative Memory (BAM), Back Propagation network, or the recent fuzzy-ART[1] are based on the basic McCulloch and Pitts like cell network. The robustness of such a cell during a recollection process is of statistical nature and is derived from **statistical** element-to-element matching[2,4,6] between the query pattern and the learned patterns. The synoptic weights are learned during training and remains static during recollection. As a result, the basic matching (activation) process during recollection treats all the elements of information with non-alterable significance. As a consequence, the derived robustness of conventional neural computing depends rather on the numerical distribution of the "correct" versus the "incorrect" information elements.

However, if some particular elements of information are knowingly missing during recollection, or if some are unreliable with a meta-knowledge about their unreliability, then there is no built-in mechanism to use this additional meta knowledge within the frame work of conventional artificial neural computing[1].

Such **meta-knowledge weighted matching** is critical in visual query with example image snippet. Because, the features in the snippet, those are important from cognitive point of view, are statistically weak. The same is true for any application where, the useful elements in the stimulus pattern is statistically weak, in compared to its size.

Let us consider the example of Fig-1. If an associative memory is trained with the image frames A, B and C, then during associative query with template-D, the conventional associative memories will produce Frame-B or Frame-C as closet match. While, from the perspective of most human the closest cognitive match is Frame-A. The reason for such unexpected result is that the statistical weakness of the feature pixels (roller) in contrast to the statistical strength of the flat background. However, what is more important is that in more complex examples, the cognitive weight can vary dynamically among the users based on varying view points. There is no way of apriori-modelling such cognitive importance during the learning.



Frame-A *Frame-B* *Frame-C* *Template-D*

Fig-1

In this paper we demonstrate a new associative network which has the capability to dynamically make use of such additional meta-knowledge about the status of information and to support specific view-points during query. The model is based on an enhanced representation of information. Internally, the information is represented as a multidimensional complex number (MCN). The phase set of MCN represents the basic information while the magnitude represents the meta-knowledge. Unlike, CANNs, Multidimensional associative computation (MHAC) considers (and processes) not only the basis information but also additional **meta-knowledge** about the reliability status of segments of information it receives and forwards. MHAC takes advantage of this additional knowledge from the input during its internal associative learning and searching and performs **meta-weighted** matching. However, when the cognitive weights are not available, then this cell performs similar to the conventional networks.

---

1 A careful analysis will reveal that setting zero (or any other value) for the missing input is not the solution to this problem. Zero-setting will contribute an error of the order of the expected value of the element.

The basic computational model is physically a conceptual generalization of optical holographic principles[3] and computational an instance of the neural network associative memory model. Among the previous works, to our knowledge Sutherland[7] first proposed a 2-D version of a computationally close network, however, without any consideration to the meta-weighted processing capability of such representation. In this work, we use a generalized multidimensional network based on the behavior of unit vectors in multidimensional Riemann space, which surpasses the performance of the former, as well as of most CANNs.

In this model, a large number of image frames are first "folded" into a correlation memory substrate of MHAC, called the holograph, using a generalized multidimensional differential Hebbian learning algorithm. During the visual query, the model first identifies the probable feature fields in the query frame to extract the desired meta-knoweledge and then performs feature-weighted search.

In this paper, we first explain the representation formalism and then present the computing model. In section 3 we present the analysis of the performance of this model, in relation to the dynamically adjustable meta-weights of the query. Finally, in section 4 we present the result from experiments with an image data base.

2. THE MODEL

2.1 Imperfect Information

A stimulus pattern is a suit of elements $S = \{s_1, s_2, \ldots s_n\}$. Each of the individual elements represents a piece of *information*. In conventional AAM, each of these pieces is expressed and processed as a scalar valued real number extending over a given range. We include the *meta-knowledge* about the state (imperfection) of a given piece of information as a part of the basic notion of information. Each piece of information is modelled as a bi-modal pair (λ, α). Where, α is a set of basic information elements and λ represents the meta-knowledge associated with this set. Thus, an element of stimulus suit is defined as:

$$s_k = (\lambda_k, \{\alpha_1^k, \alpha_2^k, ..\alpha_d^k\})$$

Meta-knowledge λ is interpreted as additional *knowledge about the state of imperfection* of the knowledge contained in associated set α. The above interpretation, in the context of a general memory (irrespective of its implementation mechanism) which computes on imperfect knowledge, generates expectations about the operational behavior of this quantity. We summarize them as following:

> **Expectation-1:** *The matching criterion should put more importance to a pieces of information that is attributed with high degree of λ than to a piece attributed with low λ in the query.*

> **Expectation-2:** *If during a memory query, the α-suit presented with a high degree of λ, and if α-suit of this query demonstrates high degree of resemblance to the α-suit of a priory encoded stimulus pattern, then memory should retrieve the associated response α-suit with*

higher degree of accuracy and high degree of λ. On the other hand, if it does not resemble any of the α-suit among the priory stored stimulus patterns in the memory, then it should generate a response with low degree of λ. Table-1 details the expectation.

Query		Response	
λ_{query}	α_{query}	$\lambda_{response}$	$\alpha_{response}$
HIGH	**CLOSE**	**HIGH**	**CLOSER**
LOW	**CLOSE**	**HIGHER|LOW**	**CLOSER**
HIGH	**NOT-CLOSE**	**LOW**	**CLOSEST| DONT CARE**
LOW	**NOT-CLOSE**	**LOWER**	**CLOSEST| DONT CARE**

Table-1

Expectation (1) relates to the forward communication of the meta-knowledge. The memory can utilize the additional meta-knowledge supplied by external body during a query to improve its search. Expectation (2) relates to the reverse communication, where the external body is communicated back not only the retrieved contents but also meta-knowledge about its status. Both of the communications are essential in the context of transferring imperfect knowledge.

We demonstrate a generalized associative memory scheme which realizes above operational expectations. In the rest of this section we show, the operational representation and the computational framework which satisfies the above expectations.

2.2 Representation

We use multidimensional complex numbers as operational representation to map the bi-modal information. Each piece of information α_i^k are mapped onto a phase element θ_i^k in the range of $\pi \geq \theta \geq \pi$ through a suitable transformation, and corresponding λ_k becomes its magnitude.

$$s_k = (\lambda_k, \alpha_k) \Rightarrow \lambda_k e^{\left(\sum_j^{d-1} \hat{i}_j \theta_j^k\right)}$$

where, each $s(\lambda_k, \theta_1^k, \theta_2^k, \ldots \theta_{d-1}^k)$ is a vector which is expanded inside a unit sphere in a d-dimensional spherical space. Each of the θ_j^k is the spherical projection (or phase component) of the vector along the dimension $\hat{i}_j$ expressing the *content* of information, and magnitude representing λ_k of the information inscribed in the phase components.

Sutherland's representation[7] is a direct 2-D special case of this representation scheme. Recently, we have shown that the generalized multidimensional representation is not only compact but also demonstrates improved response recall characteristics[5]. The multidimensional mapping of stimulus element from the external scalar field intensities is performed by some non-linear mapping. In many physical cases, the phase components can be directly obtained from the sensors. For, example in color images, the three basic color intensities can be directly translated into three dimensional components resulting in a 4-D representation.

Thus, a stimulus pattern is represented as:

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

Similarly, the response patterns are also obtained. A similar mapping on the external scalar response field intensities provides the response representation:

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

2.3 Encoding

In the encoding process, the association between each individual stimulus and its corresponding response are defined in the form of a correlation matrix by the inner product of the conjugate transpose of the stimulus and the response vectors:

$$[X^{\mu}] = [\overline{S}^{\mu}]^T \cdot [R^{\mu}] \qquad \ldots .(1)$$

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.

The strength of this new paradigm is derived from the fact that an enormous number of such associations can be learned and stored on the same $n \times m$ space by superimposing the individual correlation matrices onto the same storage elements.

A suit of associations derived from a set of stimulus and corresponding response is stored in the following correlation matrix X. The resulting memory substrate containing the correlation matrix is referred as Holograph.

$$[X] = \sum_{\mu}^{P} [X^{\mu}] = \sum_{\mu}^{P} [\overline{S}^{\mu}]^T [R^{\mu}] \qquad \ldots .(2)$$

2.4 Retrieval

During recall, an excitory stimulus pattern $[S^e]$ is obtained from the query pattern:

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

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

$$[R^e] = \frac{1}{c}[S^e]\cdot[X] \qquad \text{....(3)}$$

$$where, \quad c = \sum_k^n \lambda_k$$

3. ANALYSIS

By combining, the encoding and decoding operations expressed in (1) and (2), the retrieved association can be decomposed into principal and cross-talk components.

$$[R^e] = \frac{1}{c}\cdot[S^e]\,[\overline{S}^t]^T\,[R^t] + \frac{1}{c}\cdot\sum_{\mu\neq t}^{P}[S^e]\,[\overline{S}^\mu]^T\,[R^\mu]$$

$$= [R^e_{principal}] + [R^e_{crosstalk}] \qquad \text{....(4)}$$

Where, S^t is considered the candidate match. From (4) it can be deduced:

1. If, the excitory stimulus $[S^e]$, bears similarity to any priory encoded stimulus $[S^t]$, in their α-suit then the principal component of generated response $[R^e]$ resembles its corresponding response pattern $[R^t]$.

2. If $[S^e]$ does not correspond to any of the enfolded associations then the elements of $[R^e]$ demonstrates distinctive low magnitude indicating absence of the requested *information* in its enfolded memory.

3. The growth of the cross talk component decides the capacity of the network. For, acceptable recall performance, the magnitude of the cross talk must remain well below unity. The magnitude of the cross-talk component indicates the saturation level of the holograph.

A query is performed with incomplete knowledge (about the expected pattern) and some associated meta-knowledge about the nature of imperfection. The operational memory returns a best possible but yet imperfect answer with some degree of associated meta-knowledge about the nature of its imperfection. The following section investigates the relationship between the involved knowledge components.

3.1 FOCUS (λ) CHARACTERISTICS

We analyze the principal component given by (4) to estimate the λ of the recalled response in terms of the closeness of the query stimulus S^e and the candidate match with a priory encoded pattern S^t. Let us consider the retrieval of the j^{Th} component of the response. The retrieval of individual response components are independent and identical. We also assume that all the encoded stimulus pattern have $\lambda = 1$. This assignment merely sets a reference focus for the initially encoded information. For the sake of notational simplicity we also assume *d=2*.

$$r^e_{j(principal)} = \frac{1}{c}[S^e]\,[\overline{S}^t]^T r^t_j$$

$$= \frac{i}{c}\left[\lambda_1 e^{\hat{i}\theta^e_1}, \lambda_2 e^{\hat{i}\theta^e_2}, \ldots. \lambda_n e^{\hat{i}\theta^e_n}\right] \begin{bmatrix} 1.e^{\hat{i}\theta^t_1} \\ 1.e^{\hat{i}\theta^t_2} \\ . \\ . \\ 1.e^{\hat{i}\theta^t_n} \end{bmatrix} r^t_j$$

$$= \frac{1}{c}\sum_k^n \lambda_k e^{\hat{i}\left(\theta^e_k - \theta^t_k\right)} r^t_j \qquad \ldots(5)$$

(5) asserts the validity of expectation (a) within this computational frame work.

Now, we will compute the magnitude of the above term λ^e_j of the response. With a set of some straightforward trigonometric and algebraic manipulations it is possible to deduce the following expression:

$$\lambda^e_j = \frac{1}{c}\sqrt{\sum_k^n (\lambda_k)^2 + \sum_k^n \sum_{l \neq k}^n \lambda_k \lambda_l \cos(\theta^e_k - \theta^t_k - \theta^e_l + \theta^t_l)} \text{ assuming } |\, r^t_j \,| = 1$$

$$\text{since, } c = \frac{1}{\sum_k^n \lambda_k}, \text{ and let } (\theta^e_k - \theta^t_k - \theta^e_l + \theta^t_l) = \phi^{e-t}_{k-l}$$

$$= \frac{1}{c}\sqrt{\left(\sum_k^n \lambda_k\right)^2 + \sum_k^n \sum_{l \neq k}^n \lambda_k \lambda_l (\cos\phi^{e-t}_{k-l} - 1)}$$

$$= \sqrt{1 + \frac{\sum_k^n \sum_{l \neq k}^n \lambda_k \lambda_l (\cos\phi^{e-t}_{k-l} - 1)}{\left(\sum_k^n \lambda_k\right)^2}} \qquad \ldots.(6)$$

Let us define a distance measure between two patterns d such that, α-suit elements of the stimulus S^e and S^t are bounded by the distance d over the entire set, such that $|\,\theta^e_j - \theta^t_j\,| \leq d$, for all j which implies, $0 \leq |\,\phi^{e-t}_{k-l}\,| \leq 2d$

If $d \rightarrow 0$, then the right term of (6) becomes zero, which indicates that irrespective of distribution of δ in the query pattern, $\lambda^r_j \rightarrow 1$ (this conforms the first two dependencies of expectation (2)). On the other hand, if $d \gg 0$, indicating greater distance between the patterns, then right term of (6) becomes increasingly negative ($(\cos\phi - 1) \ll 0$), which indicates the decay of λ^r_j. (this satisfies other two parts of expectation (2)). The decay characteristics is cosine natured. For small perturbation, λ^r_j decays very little from unity indicating statistical robustness. However, with the increase of inter-pattern distance decay becomes sharper. As shown by (6), for large d, the rate of decay is modulated by the asymmetry in the distribution of λ-suit in the query.

3.2 NOISE CHARACTERISTICS

The magnitude of the responses generated from each of the individual associations represents the confidence about the match. Subsequently, the individual components are added in proportion to these confidence measures in the form of vector summation. Therefore distortion of the formation depends on the relative magnitude of the candidate solution and the sum of all other solutions, which are respectively the principal and cross-talk component of (4). In this section we investigate the signal-to-noise ratio in the retrieved response which is defined as:

$$SNR = \left[\frac{| r^{e}_{j(principal)} |}{| r^{e}_{j(cross-talk)} |} \right]^2$$

From (4), it can be derived:

$$SNR = \frac{\sum_{k}^{n} (\lambda_k)^2 + \sum_{k}^{n} \sum_{l \neq k}^{n} \lambda_k \lambda_l \cos \phi^{e-t}_{k-l}}{(p-1) \sum_{k}^{n} (\lambda_k)^2 + \sum_{\mu \neq t}^{p} \sum_{k}^{n} \sum_{l \neq k}^{n} \lambda_k \lambda_l \cos \phi^{e-\mu}_{k-l}}$$

Assuming, independent, identical and symmetrical distribution of α-suit (θ^{μ}_{j}), over all the element space of all the enfolded patterns:

$$E[\cos \phi^{e-\mu}_{k-l}] = 0 \text{ or for sufficiently large pxn, } \left[\sum_{\mu \neq t}^{p} \sum_{k}^{n} \sum_{l \neq k}^{n} \lambda_k \lambda_l \cos \phi^{e-\mu}_{k-l} \right] \to 0$$

Thus,

$$SNR = \frac{1}{(p-1)} \cdot \left[1 + \frac{\sum_{k}^{n} \sum_{l \neq k}^{n} \lambda_k \lambda_l \cos \phi^{e-t}_{k-l}}{\sum_{k}^{n} (\lambda_k)^2} \right] \qquad \ldots(7)$$

If, the distance between the candidate and query is large ($d^{(e,t)} \gg 0$), then according to (7):

$$SNR \to 1/(p-1)$$

.

On the other hand, for close match, ($d^{(e,t)} \to 0$):

$$SNR = \frac{1}{(p-1)} \cdot \left[1 + \frac{\sum_{k}^{n} \sum_{l \neq k}^{n} \lambda_k \lambda_l}{\sum_{k}^{n} (\lambda_k)^2} \right]$$

$$= \frac{n}{(p-1)} . w$$

Where,

$$w = \frac{\left[\sum_{k}^{n} (\lambda_k) \right]^2}{n . \sum_{k}^{n} (\lambda_k)^2} \quad = \frac{[E\{\lambda\}]^2}{E\{\lambda^2\}}$$

w intuitively refers to the 'porosity' of the window frame or the overall λ distribution strength. It varies from 0 to 1 and depending on the distribution of λ in the query field. 0 indicates absolute uncertainty of the information in the query frame and 1 indicating absolute certainty.

when, $\lambda_j \to 1$,

$$SNR = \frac{n}{(p-1)}$$

The above analysis provides a useful design tool to directly control the quality of the retrieval.

4. EXPERIMENTS

4.1 Training Characteristics

The following two figures first presents the training characteristics of the holograph. The average SNR for retrieval after the encoding of stimulus and response pattern associations is shown in Fig-2. The training stimulus set itself has been used as a query during recall to observe the encoding performance. Fig-3 shows the confidence buildup characteristics during the training. One of the principal strength of this network is its fast convergence. In both Fig-2 and Fig-3, the network converged within 7/8 iterations. This result is significantly faster than most conventional networks with iterative learning.

4.2 Focus and λ Characteristics

In Fig-3 and Fig-4 the query pattern distortion (x-axis) represents the degree of focus. Focus is measured as a ratio of the number of the query pattern elements within the focus and the total number of pattern stimulus elements. Fig-4 shows the variation of SNR of the target with the focus of the query. Fig-5 shows the variation of the confidence (Mean Normalized Confidence) of the retrieved targets with respect to the focus of the queries. Both of these figures also show the characteristics for few query patterns, which was never encoded in the hologram. The experiment shows that even with 20% focus frame (=20% query pattern distortion), the more than 10 db SNR can be achieved. More interestingly, the output confidence MNC remains almost unattenuated with the change in focus.

5. CONCLUSION

The proposed cell utilizes additional meta-knowledge about the status of information elements to better manage incompleteness and unreliability of information, when such meta-knowledge is available. In case when such meta-knowledge is not available, its performance becomes equivalent to other conventional cells.

In the context of visual query in image data base, this allows the network to focus on intended aspects of the snippet. The focus fields can be assigned analog values up to pixel level. The region of focus need not to be spatially contiguous. Thus, the attention of the cell can be diverted on almost any image feature, which can be expressed by a collection of pixels.

The proposed paradigm provides an effective framework for a human subject to communicate his cognitive focus to the database during his associative query. However, a separate but related open problem is the automatic detection of such focus fields in the image snippet. Currently we are working to determine the lattitude of dynamic indices which can be constructed on an image snippet. We are also investigating how it can be extracted with auotomated assistance with minimum human effort.

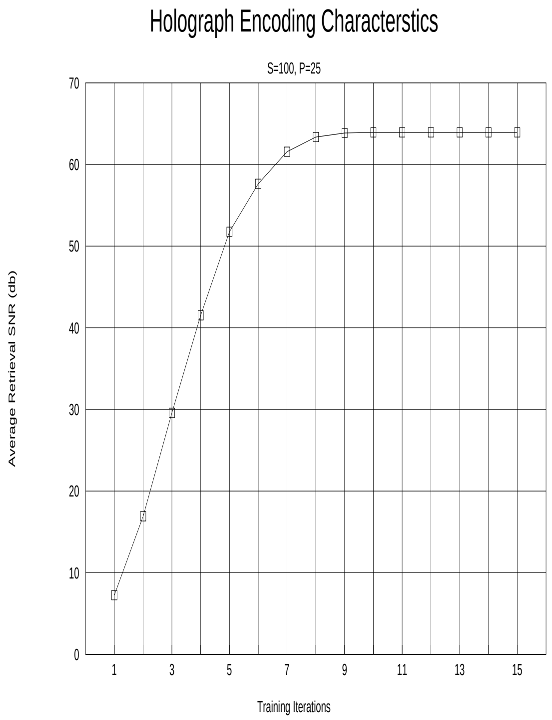


Fig-2

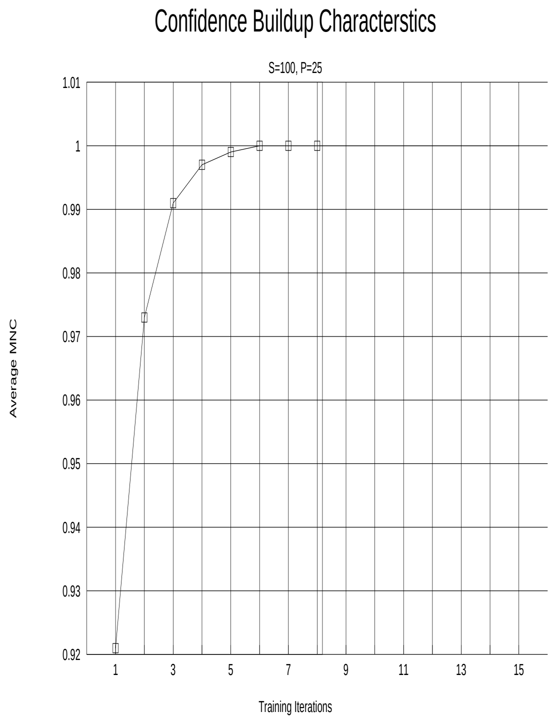
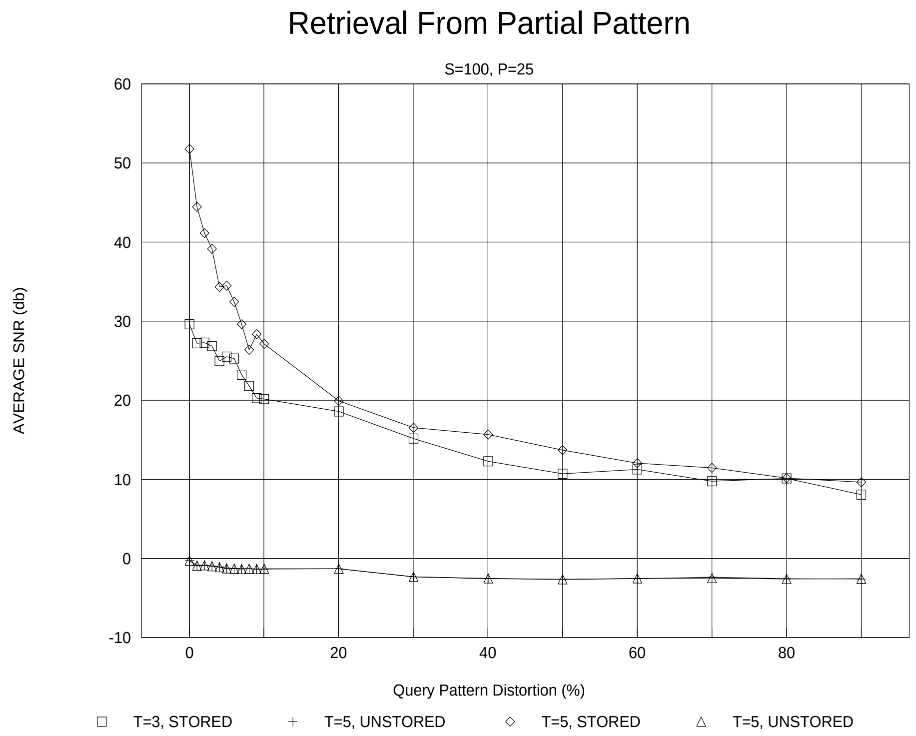


Fig-3



Retrieval From Partial Pattern
S=100, P=25
AVERAGE SNR (db)
60
50
40
30
20
10
0
-10
0
20
40
60
80
Query Pattern Distortion (%)
T=3, STORED
T=5, UNSTORED
T=5, STORED
T=5, UNSTORED

Fig-4

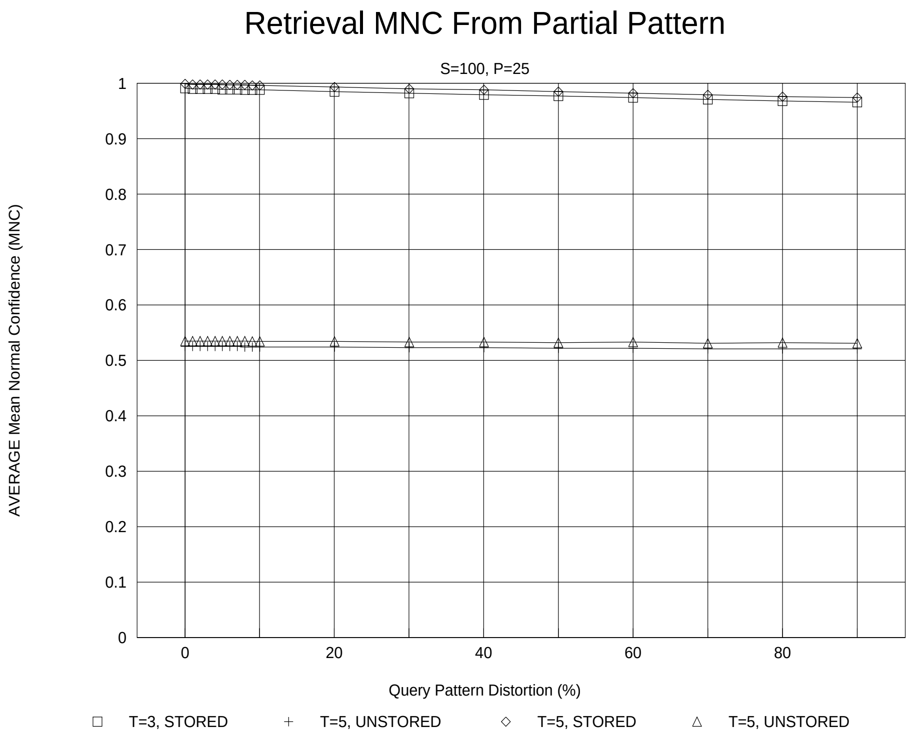


Fig-5

6. REFERENCES

[1] 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.

[2] Caudill, M. & C. Butler, *Naturally Intelligent Systems*, MIT Press, 1990.

[3] Gabor, D., "Associative Holographic Memories", *IBM Journal of Research and Development*, 1969, I3, p156-159.

[4] Hinton, G.E., J. A. Anderson, *Parallel Models of Associative Memory*, Lawrance Erlbaum, NJ, 1985.

[5] J. I. Khan 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.

[6] Keshavan, H. R., J. Barnett et. Al. , "Introduction to the Special Section on probabilistic Reasoning", *IEEE Tran. On Pattern Analysis and Machine Intelligence*, v15, no.3, p193-195, March, 1993.

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