VISUAL QUERY IN MEDICAL IMAGE ARCHIVE[1]

Javed I. Khan, D. Y. Y. Yun and H. M. Garcia

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

Abstract

This paper presents a technique for searching image archive with visual examples. The technique described in this work is specifically targeted for searching into anatomical image archive where it is difficult to obtain symbolic description of the objects of interest. This technique is based on a new computing method called as multidimensional holographic associative computing (MHAC). MHAC introduces object based search in associative computing. Unlike conventional associative computing methods, this new method can perform search on the basis of scene objects or features present in a sample image. It also retains the usual advantages of comventional assiciative computing. It is computationally efficient and the mode of computation is highly parallel and distributed. This paper presents the design, query mechanism and performance of an experimental ***medical image archive (MEDIA)*** *search system based on this new attention modulated computing technique. MEDIA search system can retrieve from cue objects as small as 15-10% of the frame size of the sample query image, which far exceeds the capability of previous associative memories.*

1. Introduction

PACS and Content Based Retrieval: One of the fundamental objective of the rapidly evolving PACS (Picture Archiving and Communication systems) research is to facilitate the access and manipulation of medical images by various medical applications. Many of these required techniques are not supported by current image database technology. One such essential capability that is necessary for PACS applications is the content-based retrieval of images. Content-based retrieval (or pictorial query-by-example) can be defined as the process where images are identified from a part of it. This part generally refers to an object or a group of objects that is useful to users' applications. These retrievals are based on the similarity between the specified objects. For instance, a doctor may want to find a similar case to confirm a diagnosis at hand. He can request PACS system for finding all the patient cases that have tumor of a certain shape. As it would be difficult to describe the shape and the "content" of the tumor, the doctor can present the image "example" which contains the tumor and expect the system to find all the images that contain such a tumor. As another instance, a doctor can request the system to find kidneys that are not of the normal shape like the one in an image example.

Existing Techniques: The need to search medical image archive by content has long been recognized [KKKP90], particular for pathology, medical research and education purposes. Many pseudo-content-based retrieval techniques have been developed. Techniques used for database and document bases in traditional Information Retrieval (IR) field [BeRS88, Salt89] have been adopted for image retrieval where the retrieval of images are mediated via the retrieval of predefined attributes or textual description (in natural language or keywords) of images [CPLJ94]. Such techniques have limitations. A picture is worth a thousand words. Due to the vast expressiveness of pictures, it is difficult to contain all the possible descriptions of objects and their (anatomical, spatial) inter-relationships inside an image by a finite vocabulary of attributes or keywords [KhYu94b]. Some images are very difficult to comprehend or interpret. Moreover in such approaches key words, index terms or attrib-

1 Published in the proceedings of the 4th International Conference on Image Management and Communications, IMAC'95, Honolulu, HI, August 1995.

utes in user queries will have to be pre-defined. Such approach necessitates extensive domain-specific modeling to be effective. On the contrary, content-based retrieval is based on direct visual similarity, which eliminates the problems of intermediate modeling.

Associative Memory for Content Based Search: Various pattern matching algorithms exists to find exact match in a pattern string [Sedg91]. However, these procedural search methods can not perform content based search on images efficiently because of the characteristically huge volume of image data[2]. The artificial associative memory (AAM) models, evolved from the recent advancement of neural network research offer a promising alternative which can overcome the inefficiency of these procedural search methods [HiAn85, CaBu90]. These AAM computation models are also inherently distributed and parallel. However, current AAMs can not support object based search. This is because they lack the ability to concentrate their focus (during similarity extraction) on any subset of the pixels in the sample image. These models perform a pre-weighted **statistical pixel-to-pixel matching** with all the query pixels. There is no mechanism to regulate the scope of search to the pixels denoting a particular object in the example frame. Dynamically shifting focus in specifically relevant in visual query inside image archive. A single image supplied as a sample during search can be interpreted in numerous ways by the searcher based on the object perception. Each interpretation may result in different answers based on the particular visual object(s) in the query scene emphasized by the searcher as a basis for similarity. For example, it is up to the searcher to decide whether the object of interest in a sample CT image depicting an abdominal cross-section is the entire spinal column or a tumor. Most of the AAM models converge only to the statistically closest match based on the entire sample scene, without adjusting to the object region intended by the searcher. Few AAM models, such as ART [CGMR92] can provide multiple answers. However, the answers are ordered according to pure statistical closeness, but have no relevance to the cognitive focus. In addition, the requirement that cues to be statistically significant[3] in the query frame is also unrealistic for querying into image archive. The critical index features used in image search as a cue are always based on their cognitive importance irrespective of their statistical dominance. Such a cue is quite often only a fraction of the entire image.

Holographic Attentive Memory: Very recently [Khan95,KhYu94] has proposed a new associative computing mechanism called as multidimensional holographic associative computing (MHAC), which can overcome the above critical limitation of the existing methods. This associative model with focus is based on a new notion of information. Unlike any artificial neural network, it considers each element of information as a bi-modal pair, which has (i) *content* and (ii) *meta-weight* components. The resulting model can support dynamically shifting view-points (or interpretations) during query and still associatively retrieve appropriate frames from archive in constant time. Thus it creates the opportunity of performing attention modulated retrieval in associative computing, which is analogous to the object based retrieval in symbolic search. It also retains the highly parallel and distributed mode of conventional associative computing.

We have developed an experimental image archive for medical images called MEDIA based on this new computing mechanism. To our knowledge, this is one of the first image archive based on distributed and associative mode of computing. This paper presents the design and performance of this archive system.

2 There is also no convenient technique to apply pre-precessing (such as sorting or FSM construction [Sedg91]) and arrange images into an ordered set to gain search efficiency. This is because the search space for images is multidimensional.

3 In Conventional AAMs the object cue is required to be at least 50% of the query frame for correct retrieval [TaJo90]

In this paper, the following section first presents a brief overview of the MHAC computing paradigm. Section 3 then presents the architecture of the prototype medical image archive, which allows content based search on the basis of various objects from in the sample image. Section 4 finally presents few example and performance result from an implemented prototype archive.

2. Holographic Associative Memory

This section briefly presents the representation, computing model, and explaining the focus ability of MHAC. Details can be found in [KhYu94, Khan95].

2.1 Information Representation

A stimulus pattern is a suit of elements $S = \{s_1, s_2, \ldots s_n\}$. Unlike conventional AAM, which express and processed each of these pieces as a scalar valued real number, MHAC includes the *meta-knowledge* about each of its pieces as part of the basic notion of information. Thus, each piece of information is modelled as a bi-modal pair.

$$s_k = (\lambda_k, \{\alpha_1^k, \alpha_2^k, ..\alpha_d^k\}) \Rightarrow \lambda_k e^{\left(\sum_j^{d-1} \hat{i}_j \theta_j^k\right)}$$

Where, α's make a set of basic information elements and λ represents the meta-knowledge associated with this set. Multidimensional complex numbers are used as operational representation to map the bi-modal information. Each α_i^k is mapped onto a phase element θ_i^k in the range of $\pi \geq \theta \geq \pi$ through a suitable transformation, and λ_k becomes its magnitude.

Where, each $s(\lambda_k, \theta_1^k, \theta_2^k, \ldots \theta_{d-1}^k)$ is a d-dimensional vector. Each of the θ_j^k is the spherical projection (or phase component) of the vector along the dimension $\hat{i}_j$. Thus, a stimulus and a response are represented as:

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

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

Such complex valued representation is not at all a new concept. Coherent light based representation used in optical holography [Gabo69] is one such example. Also, Sutherland in his pioneering work used 2-D complex numbers [Suth90] to construct one of the first truly artificial holographic associative memory. However, no previous attempt was made to investigate the attention aspect of such represetnation.

2.2 Encoding

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

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

The associations derived from a set of stimuli and a set of corresponding responses are superimposed on a super matrix X of same dimension referred as Holograph.

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

2.3 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:

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

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

The basic associative memory characteristics of this model explaining how (1), (2), and (3) together can correctly retrieve original stored response despite superimposition of the associations in (2) is given in [KhYu94, Khan95].

2.4 Focus capability

Now, the unique characteristics of this model, which allows the complete reconstruction of the response pattern from a dynamically variable (during query) small (less than 10%) segment of any stimulus is discussed.

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

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

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

Where, S^t is considered the candidate match. From (4) it can be deduced that 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 $[R^t]$.

The cross talk component behaves as a summation of randomly oriented vectors. Up to an acceptable number of associations (P), this remains well below unity, and thus, the net response closely follows the principal-component.

Let us consider the retrieval of the j^{th} component of the response (the retrieval of its other components are also identical and independent). We consider only the principal component. For the sake of notational simplicity we also assume *d=2*.

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

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

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

Equation (5) shows that each of the elements in the query stimulus (θ^e_k) tries to cancel the phase component of the corresponding encoded stimulus element (θ^t_k) by forcing $\theta^e_k - \theta^t_k \Rightarrow 0$. Thus, each tries to reconstruct the associated r^t_j on its own. The accuracy of each reconstruction depends on the closeness of these two elements. It is possible to visualize that the resultant response is a weighted average of the reconstructions done by all these individual query stimulus elements, where the weight terms are λ_k. This, mathematical construction of MHAC plays the key role in selective focus. By appropriately choosing the λ_k values, it is possible to dynamically set the importance of each query stimulus component without effecting the independent reconstruction efforts by the others. By setting $\lambda_k = 0$ it is possible to completely shut off the k^{th} stimulus element. If we have meta-knowledge that the k^{th} element is incorrect, then we can effectively block it from contributing errors in the weighted sum.

Almost all of the conventional artificial neural networks use the classical **scalar product rule** of synoptic efficacy, where the reconstruction is performed as a linear weighted sum. Where, weights are fixed during learning. Therefore, each piece of stimulus element becomes essential in the overall reconstruction. In contrast, the proposed **vector product rule** of synoptic efficacy is a form of weighted average. Thus, each term is not essential to the overall reconstruction. This critical distinction allows MHAC to dynamically adjust focus depending on the input condition.

The *signal-to-noise ratio* in the retrieved response of this model is given by:

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

Where,

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

w intuitively refers to the 'porosity' of the window frame or the overall focus (λ) density strength.

3. System Design

This section presents the content based search mechanism image archive. Fig-1 presents the architecture of the system. The system can be decomposed into three major sub-systems, namely (a) image archive (IA), (b) holographic encoding and (c) dynamic indexed query.

The actual image archive is independent from the query mechanism. Generally, images are compressed (lossy or lossless) before storage. The query mechanism does not interfere with this storage sub-system. We will describe the later two subsystems in details.

3.1 Encoding Process

Each of the stored image is first associated with one unique response label pattern (RLP). RLPs serves as an internal index for the archive sub-system. RLPs are generated using reverse Grey code to ensure maximum inter-distance between them.

The encoding process of MEDIA system involves (i) pre-processing and stimulus pattern generation (ii) Assignment of *response label patterns* (RLP), and (ii) training. These steps are explained below.

3.2 Decoding subsystem

In this sub-system, the example image is supplied by the human user. With dynamic indexing tool-set, the searcher creates a view-point mask (VM) in the example image. Given the view-point mask (VM), and the example image, the subsystem generates the query stimulus. The decoder unit uses this stimulus to search into its collection of holographic abstracts and generates a response label (RLP). The computation follows (3). The computation time is of $O(mn)$ and thus independent of the number of stored patterns.

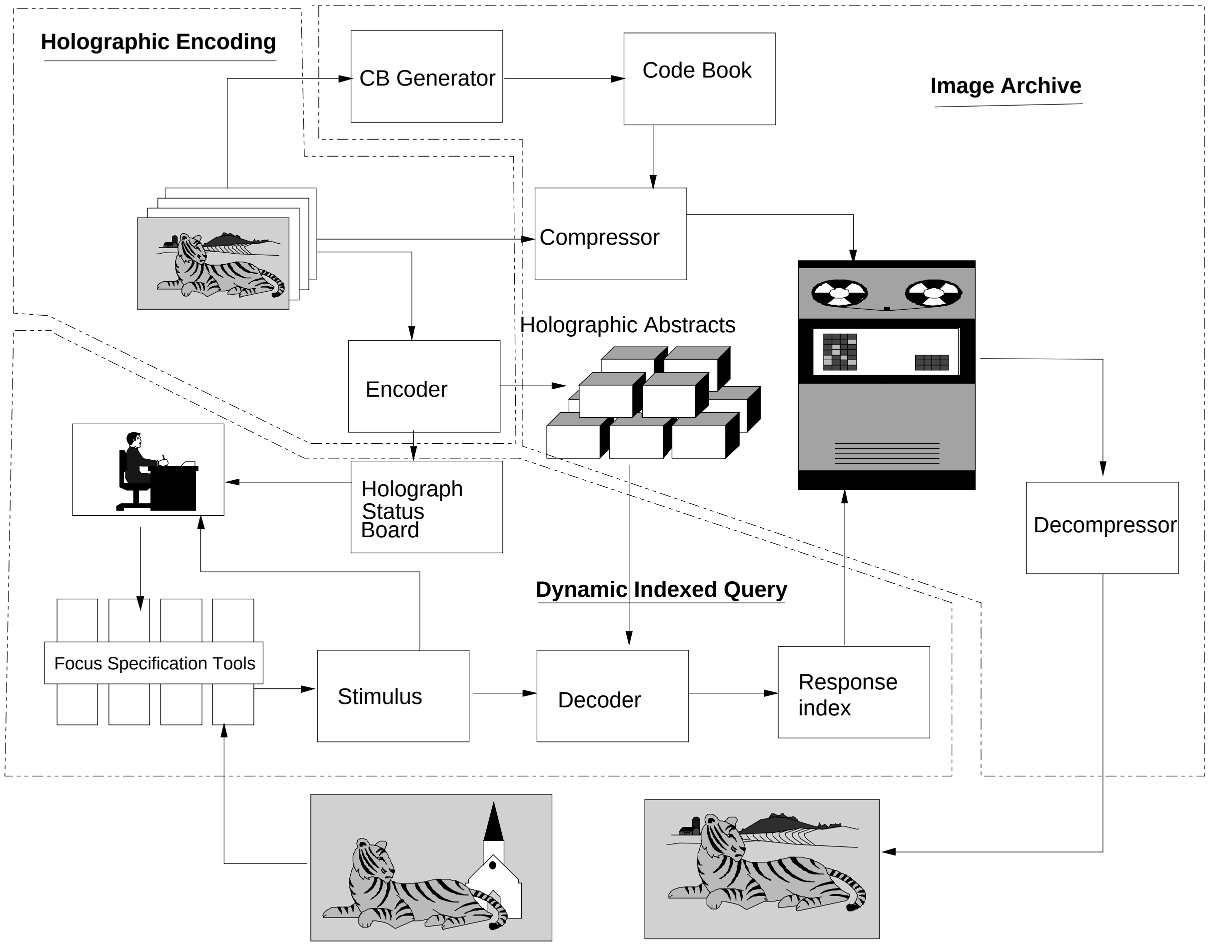

Fig-1 System Architecture

This raw RLP is passed through a noise suppressor unit (NSU) to a obtain a winner RLP from the stored RLP set. The noise suppressor measures the distance of the generated response from the previously encoded RLPs. Each RLP element is a complex number. The stored RLPs are generally assigned a magnitude of 1. On the other hand, the generated RLP magnitude provides a measure of confidence of the system on the accuracy of the generated element. Noise suppressor performs an output confidence weighted matching to converge to the closest stored RLP. This RLP is then passed to the archive sub-system to retrieve the actual image.

4. Experiment Result

Below we show the performance of a prototype system implemented on a Silicon Graphics Onyx platform. A set of 64 256x256 CT-Scan and MRI frames was abstracted into the MEDIA holograph. Fig-2 shows the performance of the encoder during the training cycles. MEDIA search system took only about 6-7 iterations to attain more that 50 db average retrieval accuracy. This is significantly faster compared to other AAM convergence speed. Fig-3 shows the performance for each of the 64 patterns at the end of 30*th* cycles. It shows the retrieval accuracy for each of the stored images. Fig-4 shows some examples from the images encoded and stored in MEDIA.

Fig-2 Encoding Accuracy

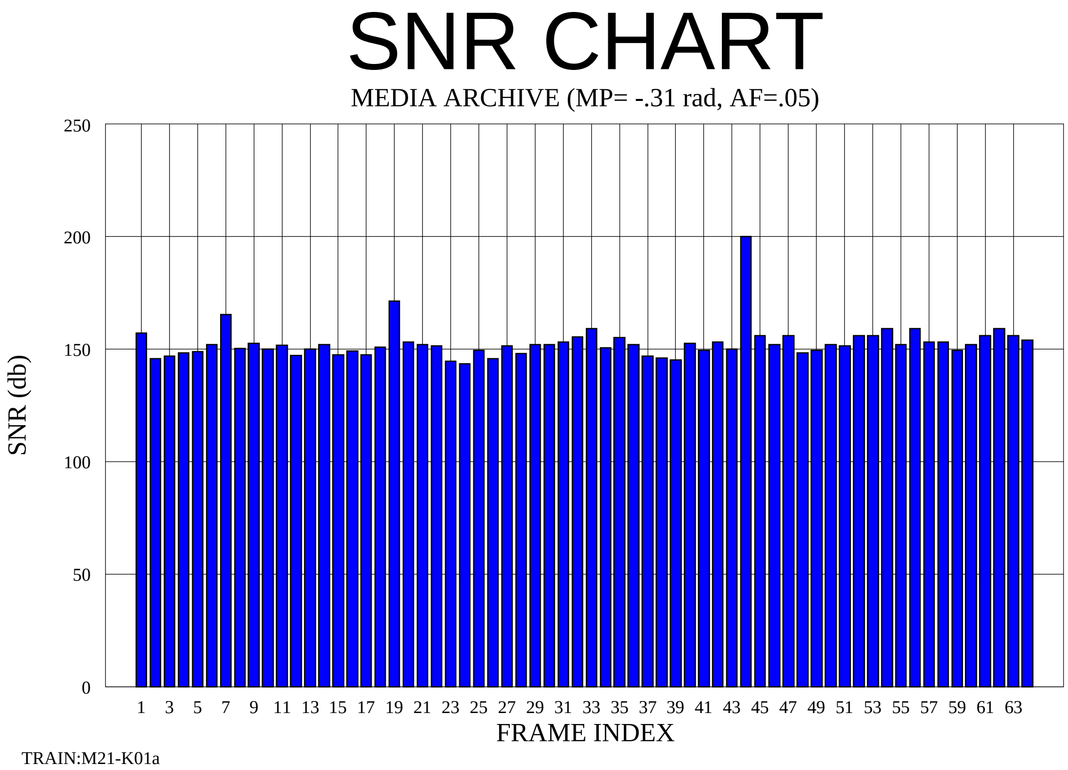

Fig-3 Decoding Accuracy

To illustrate the focus characteristics, it is shown how MEDIA search system can perform the retrieval when some objects on the query template are indicated to the system as of principal focus. Fig-5(a) shows the example of a typical sample image used during query. Two view-points are shown each of which can be a possible index object [Ande83]. Table-1 shows the view-point mask (VM) specifications of these index objects. Each of these masks are rectangular windows. Each rectangular focus window is specified the xmax, xmin, ymax and ymin. This table also shows the focus strengths (rho) of these VMs. These are few of the possible dynamic indices in this query image.

MASK#	Object	xmax	xmin	ymax	ymin	rho
4	Kidney	230	-135	165	-067	.145
6	Vertebrae	119	-045	145	-031	.131

Table-1 View-point Masks (VM) for Objects of Focus

MASK#	Object	rho	Match#	SNR (db)
4	Kidney	.145	4.4 (A26)	31.22
6	Vertebrae	.131	6.1 (A25)	27.55

Table-2 Results of Query

The encoding process searches in the holograph with the masked sample stimulus patterns at various spatial locations of the holograph. During decoding, The match numbered 4.4, 6.1 of Fig-5(b) were respectively pulled out by the system from the MEDIA as closest ones. The accuracy of the retrieved RLPs are listed in Table-2. As evident, although none of these stored pictures have statistical similarity with the query image, but each match closely on the basis of respective cognitive objects. Table-2 lists the corresponding performances of some these queries. The 2nd column in each table shows the density of the focus window (w) of each of the used object feature. As evident, the typical features or objects, which are used by humans as indices quite often fall below 10-15% of the total image. The performances of most other conventional AAMs sharply decrease when it falls below 50% of the frame because of flat statistical matching [TaJo90]. This example demonstrates the unique ability of the MHAC to retrieve images based on focus objects. It also clearly demonstrates the potential of MHAC as a central matching machine for a truly content based associative image retrieval technique.

5. Conclusion

Here we have presented the result of the experimental associative medical image archive (MEDIA), with 64 images. However, the capacity of this network is very encouraging. Given reasonable symmetry in the distribution of the intensity values, virtually 1000/2000 images can be encoded into a single holograph. It has also been shown that virtually unlimited number of images can be stored by higher order encoding [Sedg91].

A separate but related problem is the automatic detection of the focus field. Another unique aspect of this memory is the lambda reflex which provides an indication of the quality of match. It is possible to design an interactive semi-automatic focus detection mechanism on the basis of this lambda reflex in combination with various other edge detection and segmentation techniques. The same reflex can also be used to perform translation, scale and rotation varying search.

PACS, or intelligent management of image information as a whole involves the understanding of the process of cognitive perception. It is one of the most complex research problem lying at the fringe of current research in computing technology. Perhaps no single retrieval technique is sufficient to cope with the diversity and sophistication required by the PACS related applications. This technique of direct content based search can be augmented with other conventional index image retrieval techniques in a PACS. The combination of computational efficiency with object based focus ability makes MEDIA search system a serious contender as a technique for fast content based search. Also, the inherent parallel and distributed nature of assciative computing makes its potentially scalable.

Finally the authors would like to acknowledge the gracious support of Dr. Ahmed Jamal, Dr. Jay Cook and Mr. Jim Francious. A part of this research has been funded by ACTS and Supercomputing in Remote, Co-operative Medical Triage Support and Radiation Treatment Planning project of ARPA under research grant DABT 63-93-C-0056.

6. References

[Ande83] Anderson, J. E., *Grant's Atlas of Anatomy*, 8th Ed., Williams & Willikins, Bultimore, 1983.

[BeRS88] Bertino, E., Rabitti, F., and Simon, G. "Query processing in a multimedia document system". *ACM Transaction on Office Information Systems*, 6, 1, January 1988, pp1-41.

[CaBu90] Caudill, M. & C. Butler, *Naturally Intelligent Systems*, MIT Press, 1990.

[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.

[CPLJ94] Chua, T.-S., Pung, H.-K., Lu, G.-J. and Jong, H.-S. "A concept-based image retrieval system". *Proceedings of 27th Hawaiian International Conference on System Sciences*, Maui, Hawaii, Jan. 4-6, 1994.

[Gabo69] Gabor, D., "Associative Holographic Memories", *IBM Journal of Research and Development*, 1969, I3, p156-159.

[HiAn85] Hinton, G.E., J. A. Anderson, *Parallel Models of Associative Memory,* Lawrance Erlbaum, NJ, 1985.

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

[KhYu94] Khan, J. 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.

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

[KKKP90] Kofakis, P., Karmirantzos, A., Kavaklis, Y., Petrakis, E., Orphanoudakis, S. "Image Archiving by Content: An Object-Oriented Approach", *Proceedings SPIE Medical Imaging Conference IV*, February 4-9, Newport Beach, CA, 1990.

[Salt89] Salton, G., *Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer*, Addison-Wesley, Reading, MA, 1989

[Sedg91] Sedgewick, R., *Algorithms*, Addison-Wesley, Reading, MA, 1990.

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

[TaJo90] Tai, Heng Min and T. L. Jong, "Information Storage in High-order Neural Networks", *J. Of Franklin Institute*, V.327, no.1, pp16-32,1990.