FEATURE BASED VISUAL QUERY IN IMAGE ARCHIVE WITH HOLOGRAPHIC NETWORK

Javed I. Khan[1] **and D. Y. Y. Yun**

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

Abstract

This paper presents a mechanism to perform constant time content based search in an image archive with sample image. The technique described in this work is targeted to perform search in medical image archive where is particularly difficult to retrieve image based on symbolic description of shapes. Unlike conventional associative memories, this new model can perform search on the basis of scene objects or features present in the sample image. The system is based on a new associative memory model, which incorporates not only the pixel intensity information but also additional meta-knowledge about the pixel information.

1. Introduction

Content based associative search is expected be one of the critical component of intelligent image archival systems. Because, in contrast to conventional symbolic information, image information is low level, sparse and distributed and requires complex abstraction. Therefore, it is difficult to organize image information in the form of condensed key concepts (key words) and to support traditional relational search.

During the last three decades, extensive research in artificial neural networks has resulted in a number of artificial associative memory (AAM) models that support content based search mechanism, such as cross-bar associative memory, bi-directional associative memory (BAM), back propagation, ART [1,2,3], etc. These pioneering models have been successful to demonstrate the feasibility of time efficient (constant time) content based search in a highly distributed manner.

The general associative search is a complex cognitive phenomenon. These artificial models are nevertheless primitive in comparison. However, in this paper, we will address a very specific shortcoming of the existing artificial models. One of the key characteristics of cognitive matching process is that, apparently, we can focus on any specific feature or part of a scene, and consciously ignore some other, and use only the relevant part in the final matching. These critical index features are not necessarily statistically important compared to the physical dimension of the image. More importantly, we can dynamically shift the distribution of such cognitive importance during recollection. One of the critical shortcoming of the existing AAM models are their inability to support such flexibility during query.

Almost all of the current AAM networks are founded on the basic Mculloch and Pitts like cell [2,4], where, synoptic wights are learned during training, and remain static during recollection. As a consequence, the basic activation process during recollection treats all the input elements of information with unalterable weightage. The robustness of such network depends on the numerical balance between the "correct" versus "incorrect" part of total information. This results in a pre-weighted **statistical element-to-element matching**. There is no mechanism to regulate (or switch on/off) any input during query (even if we know that these are unreliable and will contribute error). As, a direct consequence, almost all of the existing artificial associative networks do not provide any mechanism to support dynamically shifting focus in the input 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. Each interpretation may result in different answers. Most of the AAM models converge only to the statistically closest match, without adjusting to the interpretation intended by the searcher. Few like ART [1] can provide multiple answers. However, the answers are ordered according to pure statistical closeness, but have no relevance to the cognitive focus.

In this paper, we present an associative search mechanism, which can overcome the above critical limitation of the existing AAMs. This associative memory with focus is based on a new notion of information. Unlike any artificial neural network, we consider each element of information as a bi-modal pair, which has (i) *content* and (ii) *meta-weight* components. As we will demonstrate, the resulting model can support dynamically shifting view-points (or interpretations) during query and still associatively retrieve appropriate frames from archive in constant time.

[1] The author is also with the computer science and engineering department at Bangladesh University of Engineering and Technology (BUET).

In the following section we first briefly present the computing paradigm on which the search mechanism is based. In section 3 we present an architecture of a prototype 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

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, we include 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:

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

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

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.

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

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.

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

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 proof of basic associative memeory characterstics of this model explaining how (1), (2), and (3) tagather can correctly retrieve original stored response despite superimposition of the associations in (2) is explained in [5,6].

2.4 Focus capability

Now, we show the unique characteristics of this model, which allows the complete reconstruction of the response pattern from a dynamically variable (during query) small (less than 30%) segment of any stimulus.

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

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

$$= [R_{principal}^e] + [R_{crosstalk}^e] \qquad \ldots .(4)$$

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*.

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

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

$$= \frac{1}{c}\sum_{k}^{n} \lambda_k e^{\hat{i}\left(\theta^e_k - \theta^t_k\right)} r^{t}_{j} \qquad \ldots(5)$$

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.

2.5 Focus characteristics

In this section we show 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:

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

Where, $\theta^{e-t}_{k-l} = \theta^e_k - \theta^t_k - \theta^e_l + \theta^t_l$, is the cross difference between the elements in l^{th} and k^{th} positions of the stored and query stimulus patterns. 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, $0 \leq |\,\phi^{e-t}_{k-l}\,| \leq 2d$.

For close match, ($d^{(e,t)} \rightarrow 0$):

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

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 focus (λ) density strength.

3. System design

In this section we will present a content based search mechanism into image database. 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.

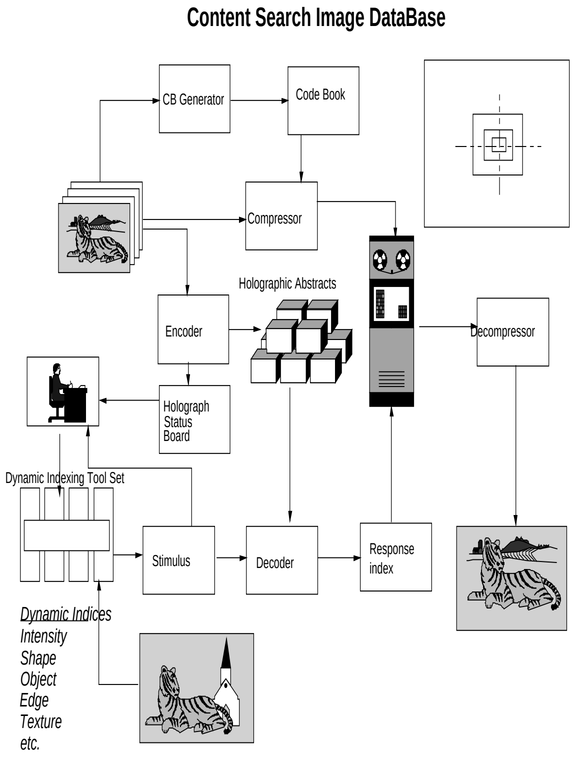


Fig-1 System Architecture

3.1 Encoding subsystem

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.

First, the auto-adaptive segmenter unit (ASU), segments the image into an analog set of subimages. Each pixel has a net belonging value of 1. Pixels are allowed to be the member of more than one sets, provided the conservation of net belongingness. The belongingness values generated a membership mask (MM) for each of the subimages.

The objective of auto-adaptive segmenter is to guess the segmentation patterns that may be generated during dynamic query as closely as possible, however, without any human intervention. Each of the segmented sub-images can be considered as external indices to the image. A multi-median threshold based algorithm is used to perform this segmentation. Each of these subimages is then transformed into a sub-image stimulus pattern(SSP). The phases of complex stimulus elements are generated from the pixel color values, and the magnitude values are generated from the membership mask MM. Each SSP is then associated with the assigned RLP of corresponding image. For each association, the encoder unit encodes the difference between the generated response and expected response by eq (1). Holographic abstract stores all the associations.

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.

This raw RLP is passed through a noise supressor unit (NSU) to generate a RLP from the stored RLP set. The noise supressor measures the distance of the generated response from the stored 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 suppresor 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 20 240x120 color images was abstracted into a holograph. Fig-2 shows the training characteristics of the encoder. It took only about 50 iterations to converge.

To illustrate the focus characteristics, we show how the network can perform the retrieval when some objects on the query template are indicated to the system as of principal focus. Fig-3 shows an example of a typical sample image. Fig-4 shows three possible view points of matching. These are few of the possible dynamic indices in this query image. Pan-A focuses on the DOG-SIMBA. Pan-B focuses on the FRED-ON-CAR, and Pan-C focuses on the NINJA. The sample image was not present in the holograph. However, during decoding, Fig-5(a),(b) and (c) were respectively pulled out by the system from the holographic memory as closest. 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-1 lists the corresponding performances of some typical 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 10-20% of the total image. The performance of most conventional network sharply decreases when it falls below 50% of the frame because of flat statistical matching[7].

Object	Density	SNR (db)	Correct Match
A-PATCH-OF-BKGRD	.108	9.73	1st (A1)
POND	.208	24.37	1st (A1)
DOG-SIMBA	.193	19.10	1st (A4)
NINJA	.144	16.93	1st (A6)
FRED-ON-CAR	.039	16.43	1st (A5)
A-PATCH-OF-JUNGLE	.09	10.65	1st (A7)

Table-1 Object based query

5. Conclusion

Here we have presented the result of a small system with 20 images. However, the capacity of this network is very encouraging. Given reasonable symmetry in the distribution of the color values, virtually 1000/2000 images can be abstracted into a single holograph. In fact, it is possible to show that virtually unlimited number of patterns can be stored by higher order encoding.

A separate but related problem is the automatic detection of the focus field. Currently, we are investigating an interactive semi-automatic focus detection mechanism on the basis of lambda reflex provided by each search. The same reflex can also be used to perform transation, scale and rotation varying search. Finally, the authors would like to thank Mr. Yu Jun for helping with the SGI system.

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.

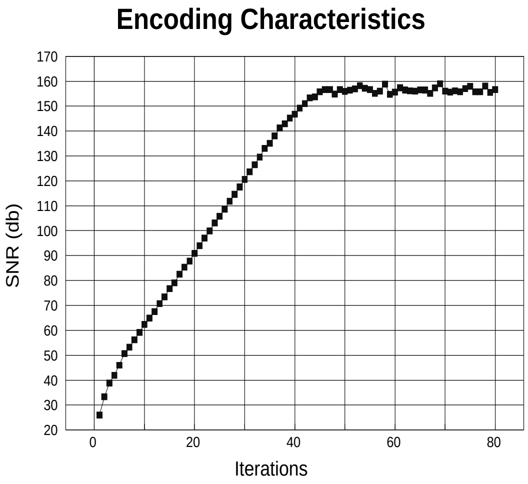
[3] Gabor, D., "Associative Holographic Memories", *IBM Journal of Research and Development*, 1969, I3, p156-159.

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

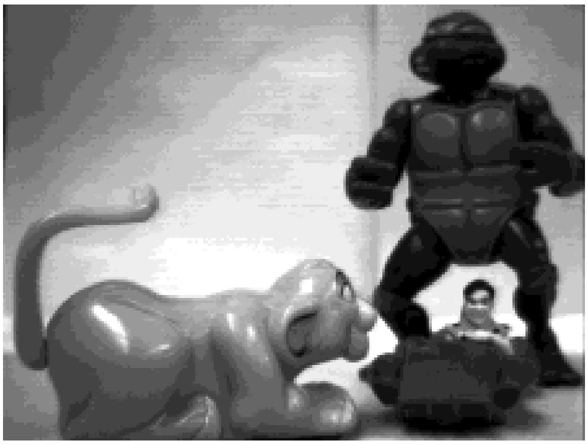
[5] J. I. Khan and D. Y. Y. Yun, "Chaotic Vectors and a Proposal for Multidimensional Complex Associative Network", *Proceedings of SPIE/IS&T Symposium on Electronic Imaging Science & Technology '94, Conference 2185*, San Jose, CA, February 1994.

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

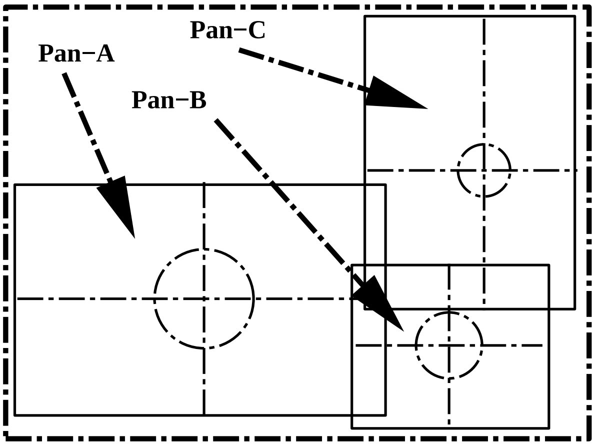
[7] Tai, Heng Min and T. L. Jong, "Information Storage in High-order Neural Networks", *J. of Franklin Institute*, V.327, no.1, pp16-32,1990.



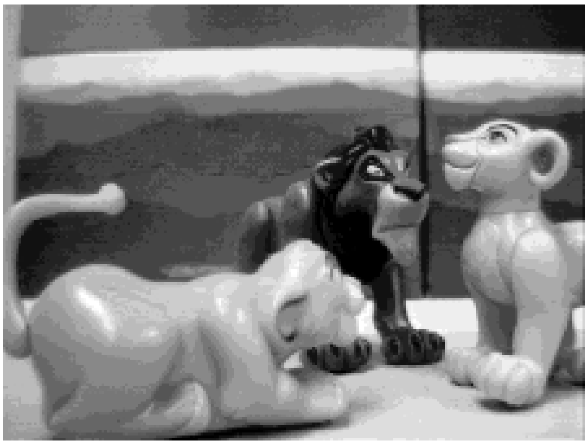
Fig−2 Performance



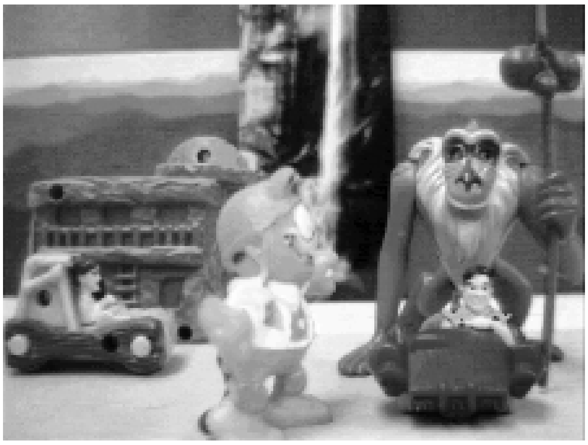
Fig−3 Query Image



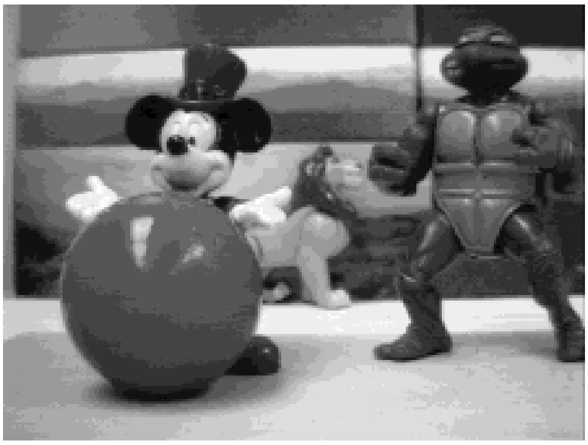
Fig−4 Object focus fields



Fig−5(a) Retrived from Pan−A



Fig−5(b) Retrieved from Pan−B



Fig−5(c) Retrieved from Pan−C