SEARCHING INTO AMORPHOUS INFORMATION ARCHIVE[1]

Javed I. Khan and David Y. Y. Yun

Department of Electrical Engineering
University of Hawaii at Manoa
javed or dyun@wiliki.eng.hawaii.edu

 Published in the proceeding of the International Conference on Neural Information Processing, ICONIP'94, Seoul, October, 1994, pp739-749.

SEARCHING INTO AMORPHOUS INFORMATION ARCHIVE

Javed I. Khan and David Y. Y. Yun
Department of Electrical Engineering
University of Hawaii at Manoa
javed or dyun@wiliki.eng.hawaii.edu

Abstract

Recent approaches for understanding and archiving of image information are mostly model based. These approaches attempt to search for pre-defined concepts in image and to quantify the content information into a structured model. However, such approaches have demonstrated limited success in handling natural images, or drawings with multitude of interpretations. The lack of any easily distinguishable structure in the pixel representation of these images allows a large number of alternative, yet subjective, interpretations. It is nearly impossible for any modeler to use a finite language to express all such subjective interpretations. In this paper, we present an alternative content-based search mechanism. This approach does not require the stored images to be subjectively modeled into intermediate representations, but rather lets a preprocessing phase to automatically abstract the images into an organized state where distinguishability by any subset of pixels are maximized. Such an abstraction process allows an interrogator to use any collection of pixels as a sample search pattern. The interrogator's query originates from his expectation and mental interpretation. The interrogator uses an example image to expressed his interpretation to the search mechanism. This associative computing technique emulates a direct but fast search into the archive.

1. Introduction

1.1 Structured and Unstructured Information

One of the principal reason that images are difficult to manage is probably that the very reality of the world (manifested in an image) lacks any rigid structure in terms of well-defined concepts. Any image is merely a narrow projection of the real world. Any imposition of a "man-made" structure tends to confine the interpretation or understanding of the world through images.

Current database technology is strictly tabular, and only recently it is shifting towards object oriented approaches, which allows at least some flexibility of representation. However, when it comes to the management of image information, even object oriented approaches fall far short of the flexibility required to cope with the formlessness of image information. After two decades of research and development, some modest progress in structure oriented approaches for image management has been made [Chan92, ChKu81, ChFu81]. However, some innovative technique to deal with the formlessness of images and the amorphousness of pixels is much needed [Chan92, Jain93, GrMe89]. In this research we will show how a new paradigm of neural network technology can come to the aid of image information management to cope with such structurelessness.

Not that all images are structureless. S. K. Chang [Chan87] has classified pictorial databases into four classes by considering the types (logical/visual) of objects handled and the "visibility" (textual/visual) of the query language. However, information that can be considered "visual" can further be subdivided into two major classes: graphic images (such as an engineering drawing, iconic map, or hand writing) and natural images (such as sceneries, paintings, or Landsat images). The process involved with managing and understanding formless natural images are substantially different from (and more difficult than) that involved with crisp graphics.

Although both classes are sensed visually by human, processing them into useful information for image management may require different approaches. Pure graphics is more symbolic, refined and pre-abstracted [Kato92]. On the other hand natural images are unrefined and amorphous. As a consequence, concept recognition in images is much more difficult and requires sophisticated feature detection and assimilation.

Concepts and objects become less well-defined and less compact as we move from graphics to natural images. From the structural point of view, graphics information can, perhaps, be managed by the traditional database approaches. On the other hand, natural image information can not be handled within the structural framework of traditional database technology.

1.2 Image Information

Information about any image can be of three physical types: (a) contextual tags, (b) pixel content (raw image), and (c) derived and condensed symbolic model, as shown in Fig-1.

Context refers to the tag information that comes along with an image, (such as name, location, time, etc.). Contextual information is highly quantized and symbolic. Conventional database technology is mature and well suited to manage such well structured contextual information. Most of today's image management systems in commercial use operate primarily with databases in this symbolic and contextual form (example systems are PACs used in hospitals,

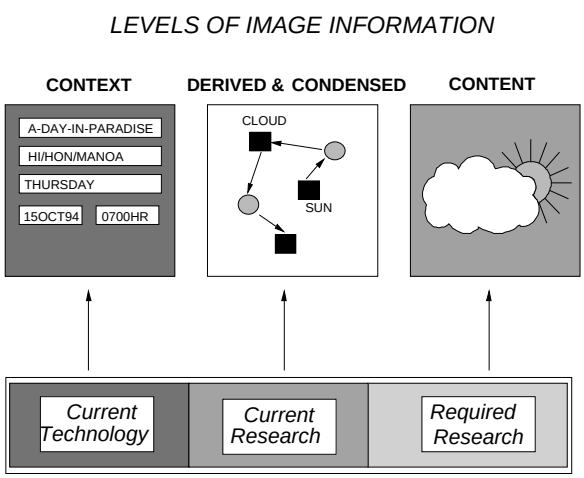

Fig-1 Physical Image Information

Multimedia systems, Macro Mind Director, AuthorWare, etc.). However, such contextual tags can not provide any information about the image contents.

The focus of todays image information management research is to provide access to the content of the image. The current main-stream effort for content-based search has been directed towards cataloging the logical meaning from the image. The process of 'meaning' extraction is more accurately a process of derivation where the meaning is derived from our subjective knowledge. The extraction of 'meaning' from raw image is a formidable task. Regardless of how the condensed logical description of content is extracted, symbolic representations in various object oriented data structures are used for information storage. Derivatives of current object oriented database management techniques have been used for searching and reasoning in this symbolic and structured intermediate description of images. Example systems include QBIC [Niba93], IIDS [ChYD88], PICDMS [JoCa88], IDB [TrPr91].

As mentioned earlier, moderate success for graphic images can be achieved through such structured approach. However, as we begin to deal more and more with non-symbolic natural images, the derivation of object model becomes increasingly difficult. For these cases, techniques that allow direct content based search become more important.

Only a few of the relatively earlier attempts have dealt directly with the content of the raw image (IMAID-ARES and GRAIN) [ChFu80, ChFu81]. However, due to the lack of efficient (software or hardware) search mechanisms, mainstream attention shifted back to approaches based on condensed representation. In this paper we will demonstrate how an effective, direct content search technique can be developed through a new generation of neural computing technology

2. Previous Approaches

Fig-2 provides a generalized schematic of the existing approaches for the image information storage and query. All approaches have two distinct stages; encoding and decoding.

In encoding stage, each of the physical images is interpreted into an intermediate representation. The interpretation involves encoder's (also referred as interpreter) knowledge and the physical image itself. The objective of this stage is to develop a condensed, efficient and crisp description of the image content from which, all the queries of the future interrogators can be answered accurately and quickly. The interpretations from a collection of images are then stored in an archive.

In decoding stage, the mental expectation of the interrogator (also referred as user) is transformed into a template through interrogator's knowledge. Interrogator also generally blends a constraint language with representation of expectation to generate one or more templates. The search is performed by matching interrogator's template with the stored models. Generally there are multiple interrogators.

The principal research issues are (a) how completely information can be represented, reasoned and queried, and (b) how the human involvement can be reduced or eliminated from interpretation and reasoning stages. The first issue is related to the design, and the second issue is related to the extraction of model.

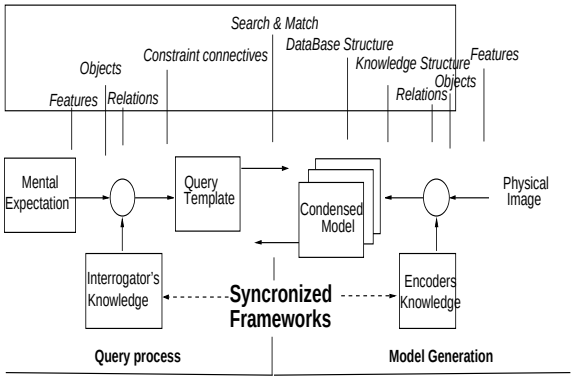

Fig-2 Model Based Approach

Current approaches can be classified into three types according to their emphasis on various aspects of the overall problem. We will name them as; (a) meaning oriented, (b) automation oriented, and (c) user profiling based approaches. Meaning oriented approach concentrates on the issue of representational completeness and depth. If necessary it assumes substantial and sophisticated human involvement. In contrary, automation oriented approach emphasizes the automation by substituting humans with computer programs. The third, and relatively recent approach, tries to use user's evaluation feedback to obtain a connection between his query and expectation. Below we briefly investigate each of these approaches more closely.

2.1 Meaning Oriented Approach

Any model generation requires knowledge at two levels. In the meta-level, interpreter must have a model about the extent and boundary of concepts that it is supposed to extract and store to satisfy possible queries. In the extraction level, It must know how to extract these concepts from the bit representation of image.

The encoding process involving human generally requires; (a) detection/annotation of concepts, (b) classification and organization of concepts, (c) construction of a knowledge structure to encode a scene knowledge, and (d) the archiving of collection of knowledge structures. Various approaches of this class can be divided into two main types; keyword/free-text based and semantic model based.

Keyword/free-text: Key word based approaches stores a set of key words with quantum descriptions of the scene [HiLe92]. For example to describe a football game scene, it may store words ball, leg, player, kicking, green, grassy, football field, game, etc. as key words. A slight variant is the free-text based approach which stores sentences. However, in both bases, search is based on plain key words. The positive sides are; (a) the system is easier to use for someone who is familiar with the range and type of keywords used by the encoder, (b) free-formatted text offers greater flexibility. There is no bound of concepts (and their level of abstraction) that can be stored and retrieved as long as both the encoder and query use the same keyword for them. The minus sides are; (a) needs agreement between the vocabulary of the encoder and user. Some system proposes judicious cross-indexing to alleviate the problem up to some extent. (b) relations connecting two specific concepts can not be handled.

Semantic Network: The semantic model based approaches attempts to capture and query not only the individual concepts, but also the relations that connect the concepts. As a first step it divided concepts into three classes, (a) objects, (b) entities and (b) attributes. It stores these entities in various forms of conceptual graphs (known as EAR models). Systems specialize on the nature of information it stores.

Such as, some systems encode the physical composition of the scene objects. It describes a scene in terms of component objects. Component objects are in turn described with their finer components. For example, a human figure can be decomposed into head, body, feet, hands, etc. Head can be further decomposed into eye, nose, etc. Generally tree graphs are used as data structure where each arc represents "part-of" or "composed-of" kind of relations between the objects it connects to represent such hierarchical compositional relations [HoHs92].

Some model stores abstraction levels of feature concepts in a hierarchy. For example, Jun Yamane and Masao Sakuchi's system [YaSa93] hierarchy represents the keywords at different levels of abstraction. They connect "football" with its alternate abstract descriptions, such as "white-obj", "white-mass", etc. through a hierarchy. The objective is to connect machine detectable features with more subjective objects. Here the arcs represent "an-instance-of" kind of relations. A number of other researchers have tried to create abstraction hierarchies of features [BeZi92, IrOX92].

The need to improve geographical information systems , such as satellite imagery, map data, etc. has sparked considerable research in encoding spatial relations among the objects in a scene [Chan87,LeHS90]. Quad tree, B-tree, R-tree, 2D-string are some of the popular representation structures proposed for encoding spatial relations. In these approaches, the space is generally divided into a 2D grid.

In Quad tree representation the 2D image space is recursively decomposed from the top into quadrants, Thus, each node of the tree is made up of four children. Each of these four links bears specific meaning (such as "on-the-first-quadrant-of") expressing the child's position relative to the parent. S. K. Chang [Chan87] on the other hand, has proposed a quite different approach where he represents the grids and their symbolic contents with two symbolic strings. The first string is obtained by scanning the grids horizontally from top to bottom. Objects in each line in equated and lines are ordered. A similar string is constructed by scanning the grids vertically from left to right.

Another group of researcher concentrated on the representation of events. The target of these systems is generally to cover broad range of knowledge and subsequent query [HiLe92]. Museum databases of history and arts are just one example of such databases. An interesting example is the Birbeck system developed by Hibler et. al [HiLe92] developed at UK. From a given text sentence, it recognizes noun, adjective and verb respectively as the entity, attribute and relation in their EAR model. A human is need to interpret a scene through English like sentences. He puts capital first letter for all nouns (Horse, Book) and puts all verbs in present continuous tense (running, playing, etc.) so that parser can easily recognize concept types and construct the semantic net.

The EMIR meta-model by Gilles Halin and N Moubaddin from France [HaMo92] is example of a more complex and broader knowledge representation formalism.

2.2 Automation Oriented Approach

A considerable research effort has been focused for automatic extraction of condensed representation. To economize the extraction, almost all attempts decompose the overall concept space into fewer basis ones and emphasize the automatic detection of these basis concepts (also referred as features). In contrast to knowledge based approaches, it uses mathematically (or algorithmically) quantifiable features. Once, these basis features are detected by searching images through filters, detected features are assimilated with the help of pre-encoded composite object models to detect higher level objects which have some kind of semantic meaning to humans.

Various methods can be distinguished on the type of feature they use. These features can be classified into two main types (a) global features, and (b) local features. Global feature based approaches utilize properties which are derived from the entire object. Geometrical features are very popular for representing shapes. Area, perimeter, a set of rectangular or triangular cover, moments [Jaga91, Hou92], etc., are few of the geometric global features those have been used to encode shape. On the other hand, local features are composed of only some important segments of object, such as line, object contour, points of maximum curvature change, [GrJi92], dots on minimal rectangle [YaSa94, etc.

Local features can withstand partial loss of object components, generally search is fast, but are susceptible to major errors in special situations. Some researchers have used statistical features such as patch histogram [Swai93]. As a typical example Hou et. al. at Siemens [Hou92] proposed first order polar moments to describe shift and rotation invariant compositional representation of objects.

Instead of using such 'meaningful' mathematical features, some research has been directed towards constructing rather covert but 'efficient' features like Forrier coefficients, fractal coefficients, or principal components. The objective is optimize some performance objective. Such as maximize distinguishability among the images, or minimize the representation space through orthogonalization of the feature space, etc. Generally these features are adaptive to the particular set of images under consideration.

Features are assimilated to detect complex objects and concepts. Computers require an object model knowledge to find out what combination of which features will make an object. If humans are to encode these object models, then it becomes easier if the basic features have some conceptual 'meaning' (even when they are mathematical).

Neural networks have been used by many researchers in automatic feature detection. Neural networks have the advantage of being able to recognize features [CJHD93], as well as object model knowledge from example. Since, modeling image features it self is a tedious task, therefore, the trainability of neural network greatly reduces the task on both feature and object modeling. Generally, rote learning networks have been used in this approach.

On the other hand, autonomous learning networks have been used to construct efficient features, such as approximation of principal component analysis [RiSt93]. However, these neural network based approaches are distinct from our approach. Because, these also actually attempt to formulate an intermediate model representation of the actual image content.

2.3 User Profile Based Approach

Oommen and Fothergill from Canada [OoFo93] describes a method for image examination and retrieval, which basically eliminates the process of image annotation by human or machine encoder, but indirectly shifts the task of modeling to the users. The technique is to adaptively group/classify images to the objects in query based on user's subjective evaluation of success and failure of the responses. The approach is explained in Fig-3. Initially images are randomly assigned to a fixed number of groups. During the first query, one image from each group (generally the one at the center of cluster) is presented to the user. The user then provides a reward/penalty response for each of these pictures. The clustering algorithm, then moves the correct responses to the state of maximum certainty (generally in the same cluster) and moves the incorrect responses to a state of minimum certainty (generally to a different cluster). Gradually the system, establishes connection between user's query and expectation. This is one of the first approach that performs matching on the basis of subjective evaluation.

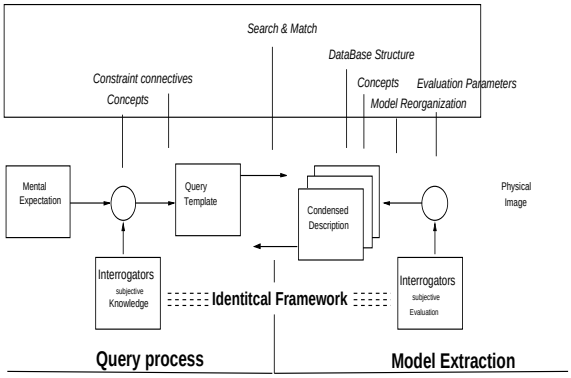

Fig-3 Subjective Modeling Approach

3. The Problem of Synchronization

The approaches based on intermediate representation of image information suffers from a fundamental difficulty of synchronization between the encoder's and the inquirer's knowledge. The difficulty of synchronization exists at both levels of knowledge.

Meta-Model Level: In this first level, the encoder has to correctly guess the expectations of the users and construct a fairly general framework (meta-model) from which he can satisfy users' expectations. The difficulty is that there may not be any such framework of finite dimension to describe visual information. Consequently, any pre-modelling runs the risk of loosing some information because of the asyncronous emphesis of particular framework. The lost information may be relevant just from a different subjective perspective. For example, in an effort to describe a historical scene the encoder may meticulously try to describe all the events portrayed. But, a future query may be on the spatial location of a character. In additional, many visual information can not be represented using symbols, keywords, or even numerals. For example, there is no convenient language to describe shape, or texture.

Model Extraction Level: Even if we assume that such a framework of representation exists, in the second level, we run into the problem of subjective evaluation of the encoder during model extraction. Same situation can be interpreted by multiple keywords. The multiple interpretability is not

only associated with the objects but also with relations that connects them. A scene can be analyzed from numerous equally valid subjective viewpoints, resulting in numerous equally valid structured models, even within the bounds of a finite well-defined language. In general, judging the visual contents of an image itself is an imprecise task.

In meaning oriented approach, when human annotates a scene, the problem is the synchronization between the user's subjective knowledge and that of the inquirers. On the other hand, in automation oriented approach, it becomes the problem of synchronization between the user's subjective knowledge and the encoders mathematical model of the world (or more precisely the subjective model of the programmer who developed the encoder).

Mathematical features: The mathematical description of basis features has the advantage, that they can be precisely quantified. However, these are very restrictive in their capability to model complex objects. It is very difficult to quantify complex concepts like "hill", "grassy field", etc., from a finite set of pre-defined mathematical basis features. The geometry of same object (even a strictly mathematical tetrahedron object) can be wildly different when viewed from different angles. On the other hand, the mathematical description of one object can easily match the mathematical description of the other. For example the circular geometry of moon is very close the circular geometry of coin.

User profiling: The approach is in fact a reaction to the synchronization problem that is deeply associated with the first two approaches. It attempts to make the interrogator also perform the functions of the encoder (computationally it may be equally tedious like the meaning-oriented approach).

It never physically accesses the raw image information. Therefore, the model of information content that the interrogator helped to construct is entirely subjective. This intermediate model representation is the subjective opinion of a collection of previous users. If there is consistency among the subjective knowledge of past users, then this approach can overcome the synchronization problem. On the other hand, if the previous users are using different vocabularies and subjective interpretations, then this aproach also suffers from missyncronization.

4. Our Approach

In this research we present a complementary approach that is conceptually equivalent to the model based search technique, but removes any intermediate interpretation/representation of the scene. Rather, the user himself translates his/her subjective expectation directly into pixel representation and performs search into the archive. Thus it eliminates the ubiquitous problem of synchronization between the subjective interpretation of the user and that of the encoder. Fig-4 shows the schematics of the approach. The approach allows direct visual search into the pixel database for features that can be represented by a finite amount of pixel subset.

Traditionally it is believed that such a direct search is computationally very expensive. We show how a space and time efficient organization of the raw image pixels through a new neural like technology can emulate (rather actually doing it) such a direct search, without being computationally expensive like the actual search.

Technically, it does convert the image information into an intermediate representation. But the objective of this transformation is to maximize mathematical distinguishability over any subset of the images and to order the multi-dimensional images to increase the search speed.

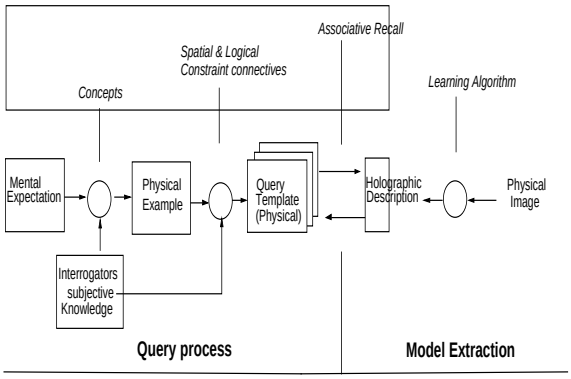

Fig-5 Content Search Approach

The user expresses his expectation through an example image. This approach entitles the user to have his/her own structured interpretation (which may be subjective) of the image. User uses this subjective interpretation to efficiently communicate his/her expectation about the relation between the objects and concepts he/she perceives to exist in the example image to the search mechanism.

In traditional approach, a structure is imposed and subsequently hard encoded in the intermediate representation. In contrast, in our approach such structure is used only at the query interface of the decoding stage. Thus, the difference of subjective interpretation between various users (or the very subjectivity of such structures) does not create any problem in our approach.

Technically, this approach is based on a new representation of basic information and a computing mechanism that handles this new representation. In many ways, this new computing paradigm resembles the neural computing, however, with some fundamental distinctions. The theoretical aspects of this new computing paradigm is explained in [KhYu94,Suth90].

4.1 Information Representation

At the heat of this new technique lies a novel notion of information. A stimulus pattern is a suit of elements $S = \{s_1, s_2, \ldots . s_n\}$. Such as, an image pattern can be considered as a sequence of pixels. Unlike conventional notion, 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 element 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:

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

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

4.2 Encoding

In the encoding process, we use methods analogous to artificial associative memories [CaGM92,CaBu90,Gabo69,Hint85]. 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.

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

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.

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

4.3 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:

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

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

4.4 Focus Capability

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

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

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

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), and for reasonably symmetrical distribution of the multi-dimentional vector elements, this remains well below unity, and thus, the net response closely follows the principal-component [KhYu94].

The j^{th} component of the retrieved response (the retrieval of its other components are also identical and independent) is shown by equation (5). For the sake of notational simplicity we have also assumed *d=2*.

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

$$= \frac{i}{c}\left[\lambda_1 e^{\hat{i}\theta_1^e}, \lambda_2 e^{\hat{i}\theta_2^e}, \ldots . \lambda_n e^{\hat{i}\theta_n^e} \right] \begin{bmatrix} 1.e^{\hat{i}\theta_1^t} \\ 1.e^{\hat{i}\theta_2^t} \\ . \\ . \\ 1.e^{\hat{i}\theta_n^t} \end{bmatrix} r_j^t$$

$$= \frac{1}{c}\sum_k^n \lambda_k e^{\hat{i}\left(\theta_k^e - \theta_k^t\right)} r_j^t \quad \ldots(5)$$

Equation(5) shows that each of the elements in the query stimulus (θ_k^e) tries to cancel the phase component of the corresponding encoded stimulus element (θ_k^t) by forcing $\theta_k^e - \theta_k^t \Rightarrow 0$. Thus, each tries to reconstruct the associated r_j^t 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 our model 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.

Here it would be appropriate to clarify the critical distinction of this computational model from conventional neural network models. 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. Weights becomes fixed after learning. Therefore, each piece of stimulus element becomes essential in the overall reconstruction. Whether it will converge correctly or not depends on the statistical balance between the correct versus incorrect components among all the pixels in the pattern.

In contrast, the proposed **vector product rule** of synoptic efficacy is a form of weighted average. Each term is not essential to the overall reconstruction. This critical distinction allows our model to dynamically adjust the subset of pixels those should contribute in the matching. [KhYu94] shows more detail mathematics behind the focus characteristics.

This new characteristics makes it possible to use any part of the content of the pattern as cue and still recover the associated response correctly. In contrast, any conventional networks (so far it is based on scalar product rule) requires statistically at least 50% in the elements to be present correctly.

In database scenario, the objects of cognitive focus are generally derived from the subjective model conceived by the user. Quite often such objects are statistically weak. Not only that, such focus varies from query to query on various pixel subsets. The focus capability of this new associative computing paradigm allows the retrieval mechanism to establish an associative search only on the basis of any user selecteble subset of pixel elements. The particular sub-set of pixels, which a user will use can be determined completely by the user depending on his own subjective interpretation of the scene.

As expected, this new computational paradigm has similar computational advantages like conventional neural networks. In fact, as shown in equation (3), a search into thousands of stored images through this technique requires a single complex matrix multiplication. A conceptually comparable pixel wise search by conventional methods will require massive amount of computations in the process of performing linear search into each of the images. In addition, the computations are highly parallel and distributed.

5. System Design

Now we will present a direct content based image database search mechanism to perform query-by-example using this new technique. Fig-5 shows the schematic diagram 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.

In archive, images are compressed before storage. The query mechanism is independent of this storage sub-system. The later two subsystems will be explained in details.

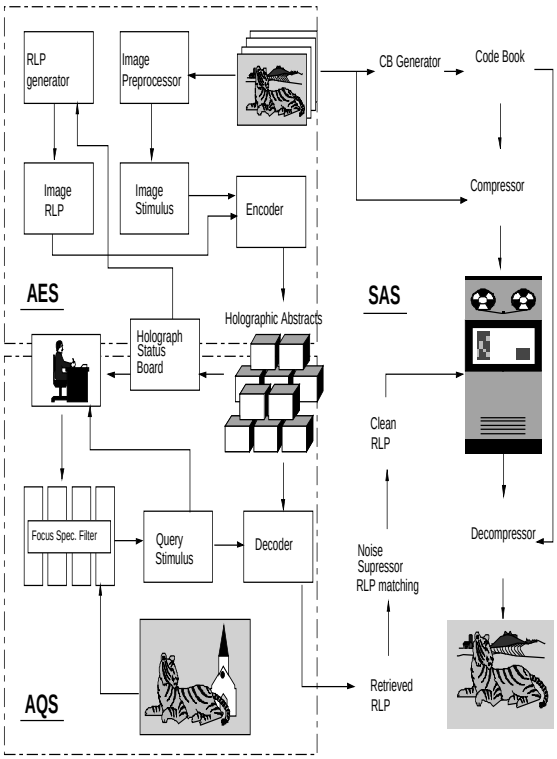

Fig-6 System Architecture

5.1 Encoding Subsystem

No human involvement is required in the encoding subsystem. Each of the stored images is first associated with one unique response label pattern (RLP). RLPs are internal system indices for the archive sub-system.

The first stage of image encoding is auto-adaptive segmenter unit (ASU), which segments the images 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 set, provided the conservation of net belongingness. The belongingness values generated a membership mask (MM) for each of the subimages. Each of the segmented sub-images can be considered as external indices to the image. 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. 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 uses a differential encoding algorithm. In this approach before encoding a new association, it is first applied to the system, and only the difference between the current response and expected response is encoded by equation (1). Holographic abstract stores all the associations.

5.2 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 description (VPD) in the example image. The view point description language has three types of specifiers; (a) element selection masks (ESM), (b) composite object connectives (OCC), and (c) spatial relation connectives (SRC).

The user is allowed to construct his own subjective interpretation (model) of a scene. Fig-6(b) shows a typical structure perceived by a subject when he was shown XQ1 (Fig-6(a)). This particular interpretation happens to emphasize only the compositional aspect of the scene. A different user may have a different interpretation of the same scene.

object	red	green	blue	x	y	density
Nbody	180,30	150,30	20,0	141,112	92,48	0.056
Nhead	80,0	90,0	60,1	137,112	110,97	0.015
Nrarm	60,0	100,20	80,0	111,94	93,63	0.013
Nlarm	60,0	100,20	80,0	159,141	97,66	0.019
Neyes	200,50	90,0	90,0	135,116	117,110	0.005
Nrleg1	60,0	100,20	80,0	122,104	60,38	0.012
Nlleg1	70,0	100,20	80,0	150,134	64,46	0.012
Nrleg2	60,0	100,20	100,0	117,104	37,27	0.004
Nlleg2	60,0	100,20	100,0	159,139	41,14	0.010
Nrknee	180,30	50,0	100,10	112,102	41,34	0.002
Nlknee	180,30	50,0	100,10	153,141	47,39	0.003
Cboy	255,80	180,80	120,20	144,122	41,24	0.008
Ccar	210,50	30,0	100,10	148,98	28,3	0.041
Sbody	255,0	255,0	10,0	105,0	55,0	0.164
Grass	200,0	220,100	170,40	159,0	22,0	0.039

Table-1 Subjective Objects of scene XQ1

Table-1 shows the ESM descriptions of primitive objects used this model. Each of the ESM specifies a window in space (x,y for 2D) and a window in color (r,g,b). The pixel locations satisfying these filter ranges contribute to the focus defining elementary objects. The regions of focus on elementary objects can be logically combined to obtain focus fields for composite objects. The focus mask of composite object is created by OCC, which is actually fuzzy set operators (fuzzy union, intersection and negation). The concept "NINJA" is an example of a composite object envisioned by the subject. It has been created by logical combination of basis objects from Table-1.

The user can also specify a region of search (ROS) for objects corresponding to a spatial space where the objects are expected. The spatial expectations expressed through SRC can be relative or absolute. (SRC also requires a search resolution number, which determines the accuracy of spatial search). The user can also express multiple alternate expectations with OR connectives of OCC. Fig-6(c) describes a typical query. It shows three alternate focus fields. Pan-C also has a ROC (shown by the filled rectangle).

Once, the expectation is specified by the user, an optimizer translates the specifications into one or more view point templates (VPTs). Each template is then associatively decoded for closest match. The results are then again logically assimilated to produce an answer corresponding to the desired search.

Given the view-point template (VPT), 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 * and thus independent of the number of stored patterns.

Each raw RLP is passed through a noise suppressor unit (NSU) to generate a RLP from the stored RLP set. The noise suppressor measures the distance of the generated response from the stored RLPs. Each RLP element is a complex number. The stored sRLPs 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.

6. Experiment Results

Below we show the experimental result of a prototype system implemented on a Silicon Graphics Onyx platform. A set of 20 240x120 color images was abstracted into a holograph.

Object	Density	SNR (db)	Correct Match
A-PATCH-OF-BKGRD	.108	9.73	1st (A1)
POND	.208	24.37	1st (A1)
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-2 Object Based Query

Fig-6(a) shows an example of a typical sample image. Fig-6(c) shows three possible alternate view points of matching. These are few of the possible dynamic indices in this query image. Pan-A focuses on the SIMBA. Pan-B focuses on the FRED-ON-CAR, and Pan-C focuses on the NINJA. The sample image was not encoded in the holograph. However, the decoder process pulled out, Fig-6(d),(e) and (f) from the holographic memory as closest. As evident, although none of these stored pictures have statistical similarity with the query image, yet each matched closely on the basis of respective cognitive objects. Table-2 lists the performances for some typical queries. The 2nd column in each table shows the density of the focus window (w) of each of the used object features.

7. Conclusion

7.1 Space and Time Efficiency

In this research we have indicated how a truly content based search mechanism based can be instituted based on a new computing paradigm. The advantages of traditional model based approaches are generally the search speed and condensed representation space. Although these two approaches can not be compared in terms of their capabilities, yet it can be argued that our new approach is not inefficient either in terms of space or in terms of speed.

Space Efficiency: Table-3 shows the space situation for few typical archive dimensions. The space factor represents the information compression due to holographic condensed representation. Clearly, the holographic abstract takes a nominal[2] space.

frame size	# of frames	memory loading	raw storage	RLP size	holograph storage	space factor
128x128	512	.014	25MB	8	1.5MB	.062
320x240	2048	.0088	472MB	10	9.2MB	.019
1024x1024	10,000	.0030	30GB	12	170MB	.005

Table-3 Holographic Space

Time efficiency: Irrespective of the number of frames stored in the holograph abstract, it requires one complex matrix multiplication. However, for a given retrieval accuracy (in the table we allowed error to be as big as 50% of the dynamic range with ample margin. Experiments indicate that the error can be as low as 5-10% of the dynamic range even when the memory load factor reaches .25 to .5) the size of the RLP, and thus the size of the holograph increases only logarithmically with the number of patterns. Thus, the holographic search is of logarithmic order in terms of the number of frames stored. This is equivalent to a conventional search into an ordered set. In fact, the holographic encoding can be considered as a multidimensional ordering process. In comparison, conventional image database requires linear time for comparable search. Such logarithmic time associative search into a massive image space makes it feasible to allow direct spatial search with series of rotated/shifted templates. This is not realistic with conventional image database, where the images are stored as unordered pile of pixels.

7.2 Bounds of Representation

A profound nevertheless interesting issue pertaining to our new approach as well as for any image database is that what subset of concepts can be searched automatically?. One of the critical requirement for our approach is that the index concepts, used by the user, must be expressible in terms of finite set of pixel subsets before he can search for it. Clearly, it will not be possible to come up with a general representation of concepts such as "hill", "ocean" ever.

We believe that this is probably a fundamental limitation for fully automatic content based search mechanisms. In automatic model based approach, the initial model extraction requires a filter program. In the absence of such pixel-format representations, there also cannot be such a program.

For those concepts, which are representable, both, model based as well as our approach can work. The relative efficiency depends on the frequency with which a specific representation is recalled. If the set of such concepts is smaller or at least finite, then most probably it is profitable to run these concept filters at the encoding time and build up a concise model based representation from computational efficiency point of view. On the other hand, if the representation is more subjective and non-reusable, then it will be more efficient to compute the concept at the query time as in our approach.

7.3 Recommendations

Collective experience over a period of two decades evinces that most likely there is no panacea to solve the complex problem of image information management. Various approaches have been developed in these two decades to address various special cases, of this extremely complex problem, all with their relative strength and weakness.

In this research we just present another approach for searching into image information, which has its advantage when the content it self does not show any unambiguously distinguishable structure. However, most probably, any intelligent image database has to be collection of many complementary search mechanisms. As a whole, the content based search is not only a database problem, but also a problem of understanding our own mechanism of perception. Finally we would like to make the following recommendation as a direction towards future research:

1.**Fusion of pictorial and logical reasoning:** The information contained in a picture is extremely complex. Some of it is better communicable in graphical language (such as spatial relations), some may be in command language (such as logical constraints) and some may be pictorially (such as a tree). An effective image management system should be able to harmoniously fuse various representation languages.

2.**Fusion of content and context-based search:** The tag, condensed representation as well as direct content search, all have their advantages as well as disadvantages. A good database should be able to provide alternate and complementary search strategy to coherently reason into all three segments of image information.

---

2 The problem in direct comparison is that it is difficult to estimate exactly how many symbolic keywords are sufficient to describe a scene. Is a picture worth a thousand words?

3.**Modeling of user expectations:** The understanding of users subjective expectation is both critical as well as difficult task. Automated techniques are need to reduce the cost of modelling user's expectation. The idea of user profiling can be incorporated into other approaches too.

The authors would like to thank Yu Jun and Sheng Liu for their wholehearted and diligent assistance in tracking many relevant works mentioned in this paper. The authors would also like to thank Charlie Lee, who helped us with his subjective model of Ninja.

8. References

[CaGM92] 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.

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

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

[Hint85] Hinton, G.E., J. A. Anderson, *Parallel Models of Associative Memory*, Lawrance Erlbaum, NJ, 1985.

[KhYu94] 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.

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

[BeZi92] D.Benson, G.Zick, "Spatial and Symbolic Queries for 3-D Image Data", *SPIE, V1662 Image Storage and Retrieval Systems*, pp134, 1992.

[ChFu80] N.S.Chang, K.S.Fu, "Query-by-Pictorial-Example", *IEEE Trans. on Software Engineering*, Vol SE-6, N6, pp519, November 1980.

[ChFu81] N.S.Chang, K.S.Fu, "Picture Query Languages for Pictorial Data-Base Systems", *Computer*, pp23, November 1981.

[Chan87] S.K.Chang, "Visual Languages: A Tutorial and Survey", *IEEE Software*, pp29, January 1987.

[Chan92] S.K.Chang, "Image Information Systems: Where Do We Go From Here?", *IEEE Trans. on Knowledge and Data Engineering*, V4, N5, pp431, October 1992.

[ChYD88] S.K. Chang, C.W.Yan, D. Dimitroff, et al., "An Intelligent Image Database System", *IEEE Trans. on Software Engineering*, V14, N5, pp681, May 1988.

[CKLY93] T. Cheng, J. I. Khan, H. Liu, & D. Y. Y. Yun, "A Symbol Recognition System", *Proc. of Int. conference on Document analysis and recognition*, Japan, pp918-921, October 1993.

[GrJi92] W.I.Grosky, Z.Jiang, "A Hierarchical Approach to Feature Indexing", *SPIE, V1662 Image Storage and Retrieval Systems*, pp9, 1992.

[HaMo92] G.Halin, N.Mouaddib, "An Object Oriented Approach to Design a Content-Based Image Retrieval Model", *SPIE, V1662 Image Storage and Retrieval Systems*, pp100, 1992.

[HiLe92] J.D.Hibler, C.H.C.Leung, et al., "A System for Content-Based Storage and Retrieval in an Image Database", *SPIE, V1662 Image Storage and Retrieval Systems*, pp80, 1992.

[HoHs92] T.Y.Hou, A.Hsu, et al., "A Content-Based Indexing Technique Using Relative Geometry Features", *SPIE, V1662 Image Storage and Retrieval Systems*, pp59, 1992.

[IrOX92] M.A. Ireton, J.P. Oakley, C.S. Xydeas, "An Hierarchical Classification Method and Its Application in Shape Representation", *SPIE, V1662 Image Storage and Retrieval Systems*, pp154, 1992.

[Jaga91] H. V. Jagadish, "A Retireval technique for Similar Shapes", *Proc. of ACM SIGMOD Int. conference on the management od data*, pp208-217, Denver, May 1991.

[Jain93] R. Jain (editor), "Workshop Report: NSF Workshop on Visual Information Management Systems", *SPIE, V1908*, pp198, 1993.

[JoCa88] T. Joseph, A.F. Cardenas, "PICQUERY: A High Level Query Language for Pictorial Database Management", *IEEE Transactions on Software Engineering*, V14, N5, pp630, May 1988.

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

[NiBa93] W. Niblack, R. Barber, et al., "The QBIC Project: Querying Images By Content Using Color, Texture, and Shape", *SPIE, V1908*, pp173, 1993.

[OoFo93] B.J.Oommen, C.Fothergill, "Fast Learning Automaton-Based Image Examination and Retrieval", *The Computer Journal*, V36, N6, pp542, 1993.

[RiSt93] R.Rickman, J. Stonham, "Similarity Retrieval from Image Databases - Neural Networks Can Deliver", *SPIE, V1908*, pp85, 1993.

[ChKu81] S.K.Chang, T.L.Kunii, "Pictorial Data-Base Systems", *Computer*, pp13, November 1981.

[Swai93] M.J.Swain, "Interactive Indexing into Image Databases", *SPIE, V1908*, pp95, 1993.

[GrMe89] W.I.Grosky, R.Mehrotra, "Image Database Management", *Computer*, pp7, December 1989.

[TuPr91] A. Turtur, F. Prampolini, et al., "IDB: An image database system", *IBM J. RES. DEVELOP.*, V35, N1/2, pp88, January/March 1991.

[LeHs90] S.Y.Lee, F.J.Hsu, "2D C-String: A New Spatial Knowledge Representation for Image Database Systems", *Pattern Recognition*, V23, N10, pp1077, 1990.

[YaSa93] J. Yamane, M. Sakauchi, "A Construction of a New Image Database System which Realizes Fully Automated Image Keyword Extraction", *IEICE Trans. Inf. & Syst.*, Vol E76-D, N10, pp1216, October 1993.

[YaSa94] Y.Yaginuma, M.Sakauchi, "Proposal for an Intermediary Shape Representation Using Dots on Minimal Rectangles", *Systems and Computers in Japan*, V24, N9, pp52, 1993.