Reprint from the Proceedings of the International Conference on
Intelligent Systems, ICIS'96, Nevada, June 96, pp. 187-191.

A Visuo-Logical Query Formalism for Content-based Image Retrieval with Holographic Associative Search Engine

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

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

ABSTRACT

This paper reports an associative approach for visuo-logical object oriented content-based query in image archives. The technique is based on a new associative search kernel called multidimensional holographic associative computing (MHAC) and a multivariate logic based inference mechanism which acts as a reasoning shell on the MHAC based search engine. This is one of the first query mechanism based of scalable associative search. It now allows user to define his/her own objects and index space and perform reasonably sophisticated query directly into images, without being mediated by intermediate symbolic model or encoder.

Key words: *associative memory, attention, visual query, fuzzy-inference, content-based retrieval.*

1. INTRODUCTION

Any database query mechanism is a combination of a pattern matching engine and an inference mechanism that allows the logical composition of searches to meet complex expectations of inquirer. It would not be inappropriate to state this relation in the following familiar form:

Query Mechanism = Search Engine + Inference Engine

In this paper we make observations about the nature of the traditional search and inference engines, specially with respect to their suitability in *content based image query* (CBIQ). Eventually, we explore a new search mechanism which has already demonstrated some very attractive properties for content based image retrieval [7] with dynamically definable indices. In this paper we particularly focus on developing an inference mechanism that can harness the power of this promising new search technique.

Many pioneering CBIQ approaches attempted to use traditional symbolic search mechanisms. However, now it is well experienced that these traditional search techniques face a number of serious difficulties while searching into image database [2,5,8]. First of all, any search in image is computationally massive. Compared to typical database record, image is representationally much more denser. In addition, image information lacks structure and consequently can not be conveniently divided and catalogued in any ordered fashion in predefined concept slots. Whereas symbolic search approaches rely critically on the concreteness of concept definitions (i.e. "slotability") and their ordering. Clearly, there is a need for more innovative search mechanism which is (i) characteristically adept in dealing with imprecise and structureless information, and (ii) is computationally much more scalability.

Very recently [7] has developed a new search technique which has both of these important characteristics desirable for CBIQ. It has also demonstrated encouraging capabilities to perform direct *content based image retrieval* (CBIR) in almost constant time from vast amount of images. This paper now investigates the architecture of a fuzzy logic based inference engine for this new search mechnaism which can support more advanced CBIQ by incorporating "reasoning" capability with visual objects. To our knowledge this is the first content-based image retrieval system which is founded on associative learning and retrieval [2,5,8].

The paper presents the design of this new CBIQ formalism. In the following section we briefly describe the relevant properties of MHAC, and in section 3 we present the proposed query formalism. In section 4 we present the inference mechanism that supports this formalism. Finally in section 5 we illustrate its operation through examples.

2. HOLOGRAPHIC SEARCH ENGINE

2.1 Strengths of Associative Computing:

Parallel and distributed models of artificial associative memory (AAM) have demonstrated the ability to graciously cope with the inherent imprecision of image information [1,4]. Compared to symbolic search mechanisms, they are adaptive, computationally efficient and their mode of computation is highly parallel and distributed. It has long been anticipated (almost from the days of the advent of neuro computing) that AAMs can be

Reprint from the Proceedings of the International Conference on Intelligent Systems, ICIS'96, Nevada, June 96, pp. 187-191.

potentially used for content-based retrieval of image information. However, no such system has actually been implemented yet.

One of the reasons for the lack of success is that current AAMs do not have the ability to focus on visual objects or more specifically on any "meaningful" subset of the pixels in the images denoting the visual object(s). During query, they emulate an indiscriminate statistical pixel-to-pixel distance evaluation (mean square error, entropy etc.) giving equal importance to all the pixels in the sample image, which makes current AAMs rather simplistic to be used for image retrieval.

The meta-control over the importance of pixels, or in other words localization of search scope is critical for any search based on object similarity both during encoding and recalling. A sample image can be interpreted in numerous ways based on the visual object(s) in it that the searcher wants to emphasize as a basis for similarity. Each interpretation may result in different answers. For example, in a sample CT image depicting an abdominal cross-section the searcher may choose either the "spinal column" or a "tumor" as his/her basis of similarity. But, current AAM models cannot accommodate such dynamic (post-learning) specification of focus. For the same reason, it also always converges to the closest match based on statistically dominant features [7]. However, the critical index objects used in image search as cues 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.

In addition, to deal with imperfect information like images, any "best match" search engine should also provide a quality of match feedback for every retrieval. Such feedback is not ornamental but vital if the search engine has to be interactively integrated with any high level inference mechanism. Any previous AAM lacks such "assessment" ability.

The holographic memory proposed by [7] overcomes both of the above inabilities. What makes it more attractive for CBIQ is that it also retains the usual advantages of associative computing, namely adaptability, efficiency, ability to cope with imprecision, scalable and parallel mode of computation.

2.2 Holographic Associative Memory:

This new memory model is based on a formalism which assumes the trust in each piece of transacted information as inherently non-conformal (thus enabling it to handle selective and dynamically changeable attention). In addition to the basic measurements, the formalism includes the meta-information about the status of each given piece of measurements as an integral part of its representation and computation. A conventional AAM computes only with the former. Internally, the information is represented as multidimensional complex numbers (MCN) spanned in a hyperspherical space. The phase set of MCN represents the measurements while the magnitude represents the meta-information. The computational model is physically a conceptual generalization of optical holographic principles [3] and computationally an instance of associative memory. The details of this model can be found in [7].

The search scheme is as following. A large number of image frames are first "folded" into the correlation memory substrate of MHAC, called as the holograph using a generalized multidimensional differential Hebbian learning algorithm. The learning algorithm associates each of the learned images with a label called *Response Label Pattern* (RLP). A search or decoding is performed with an example image I and a mask M on it specifying the attention distribution on this image (which defines the feature object). For each retrieval with a [I,M] pair it returns the pair [RLP_T,MNC_T], which respectively represents the RLP of the closest matching image T, and a measure of the closeness.

The computational advantage is gained from its property that this entire search process is a single step operation involving a convolution of the sample image with the holograph. It can mediate a best match search with dynamic cue in a database of p images of size n with computational complexity O($n.logp$). Like traditional approaches, it does not need to search individual images (which has complexity O(np)). For large p, this is a saving of exponential order.

3. THE INFERENCE FORMALISM

At the inference level, most of current approches use traditional bi-valued logic based inference system. However, a bi-valued symbolic logic appears brittle while managing image information. The reason is that although the bi-valued logic provides a means of doing "logical algebra" with symbols, but the symbols themselves are assumed to be hard defined. Conceptually, a search involves the following steps, (a) a set of basis concepts (i.e. symbols), (b) specification of the concept in a query format, (c) search for the concept in image archive, (d) measurement of concept strength, and (e) verification of the measurements with expectation stated in the query. In the context of image information, each of these steps is inherently subjective both in the sense of "concreteness" of definitions and "precision" of measurements, thus the symbols are “soft”. Bi-variate logic based inference engines provide no convenient mechanism to alter or adjust these underlying assumptions of "concreteness" and "precision". A reasoning capability suitable for imperfect information [6,10] is needed for quering image information.

We demonstrate an experimental fuzzy-logic [10] based symbolic inference mechanism to interface with the holographic associative search engine. The proposed engine at each stage of its inference process use a range value (rather than a fixed value) of "concreteness" and "precision" of the involved symbols and concepts. In addition, it also allows dynamic specification of these ranges.

The formalism supports two types of query compounding, namely (a) logical compounding, and (b) transform compounding. A logically compounded query is similar to the compounding used in conventional databases which integrate logical connective AND, OR and NOT. The mechanism of transform compounding allows users to perform searches involving variations based on transformations, and is an important part of image search. It corresponds to active processing of templates before match. Both, of these compounding can potentially require multiple template matching for a single query.

Reprint from the Proceedings of the International Conference on Intelligent Systems, ICIS'96, Nevada, June 96, pp. 187-191.

3.1 Logical Compounding

The specification language allows logically compound queries to be chained together. The query language is composed of (i) objects, (ii) logical connectives, and (iii) MNC linguistic quantifier (MNC-LQ).

Format: The query formalism is of the following format:

COM-QUERY: (MNC-LQ.OBJECT). LOGICAL-CONNECTIVE. (MNC-LQ. OBJECT)

Example: The following is an example of logically compounded query. It defines its complex query objective OBJ-NINJA as:

OBJ-NINJA: (HIGH.NINJA-HEAD).OR. (HIGH.NINJA-CHEST)

This example searches for object NINJA in the archive. The above command specifies that the presence of compound object NINJA can be confirmed by the strong presence of either, NINJA-HEAD or NINJA-CHEST. The quantifier HIGH is a linguistic quantifier of the required detection strengths of NINJA-HEAD and NINJA-CHEST objects to confirm the presence of NINJA.

3.2 Transform Compounding

Transforms are used when the query image differs from the expected image in some procedurally definable sense (such as the sought object may be approximately half the size than the sample). It is also used when searching in based on invariancy with respect to some procedurally definable feature (such as size, translation, size). Translation in space, rotation, all are examples of transforms. Users can define the transforms.

Each of the transforms has a set of parameters to quantify the abstract concept it defines. The general formalism of transform compounding accepts a specification of a range of these transform parameters on a given object. Each query is composed of (i) objects, (ii) transforms (XFORM(parameter-range)), (iii) transform-relations (XFORM-RELATION), and (iv) transform-relation-quantifiers (XFORM-REL-LQ or LQ).

Format: The format is specified below (here p,q and r are indices to the keywords):

COMP-OBJ: (OBJECT1(XFORMp(range-parameters)). XFORMp-RELq-LQr.XFORMp-RELq. (OBJECT1(XFORMp(range-parameters))

Example: Below, an example of transform query is provided (the prepositions are just for the ease of reading and can be ignored):

HUMAN:(NOSE(in LOCATION(middle of the picture)) JUST.ABOVE.(MOUTH(in LOCATION(lower half of the picture))

In this example, LOCATION is a transform (SIZE, CONTRAST, etc. are also other possible transforms). The range-parameters of LOCATION is generally a region in the picture frame, where the object NOSE is expected to exist. ABOVE is a relation concept which is associated with LOCATION (LEFT, DISTANCE, are possible relations which can be associated with the LOCATION transform). JUST is a linguistic quantifier associated with the ABOVE concept (FAR is another example of such quantifier).

3.3 The Language (CQL)

The grammar for the Compound Query Language (CQL) is given below. The basic types are (i) base objects [I,M] OBJECT, (ii) tranforms XFORM, (ii) transform relations XFORM-RELATION, and (iv) linguistics quantifier LQ.

$$
\begin{aligned}
\mathit{object} &\leftarrow \mathit{OBJECT}\\
&\leftarrow \mathit{object}(\mathit{transform})\\
&\leftarrow \mathit{object}.\wedge.\mathit{object}|\mathit{object}.\vee.\mathit{object}|\neg.\mathit{object}\\
&\leftarrow \mathit{object}.\mathit{relation}.\mathit{object}\\
\mathit{relation} &\leftarrow \mathit{LQ}.\mathit{XFORM}-\mathit{RELATION}\\
&\leftarrow \mathit{relation}.\wedge.\mathit{relation}|\mathit{relation}.\vee.\mathit{relation}|\neg.\mathit{relation}\\
\mathit{transform} &\leftarrow \mathit{XFORM}(\mathit{parameter}-\mathit{range})\\
&\leftarrow \mathit{transform}.\wedge.\mathit{transform}||\neg.\mathit{transform}
\end{aligned}
$$

Objects can be compounded directly or through compound relations. However the direct logical compounding of objects can be interpreted as a special case of transform-relational compounding with a default transform-relation purporting the sense of existence with a linear evaluator function s=MNC. In the syntax, the LQ and transform-range can be omitted, implying a pre-declared corresponding default value for LQ.

4. INFERENCE ENGINE

Once a query is requested, each CQF query statement is decomposed into four components: (i) canonical object specifications [I,M], (ii) transform specifications, (iii) decoder specifications, and (iv) quantifier specifications. After decomposition, it is processed through following four stages; (1) search pattern identification (2) generation of templates, (3) matching of templates, and (4) assimilation of match. Fig-1 shows these stages.

In the first stage, the object specifications in the command are parsed and component objects which are to be searched independently are isolated. A pattern generator generates these canonical patterns from the object base. The objects can be either elementary or complex.

Reprint from the Proceedings of the International Conference on
Intelligent Systems, ICIS'96, Nevada, June 96, pp. 187-191.

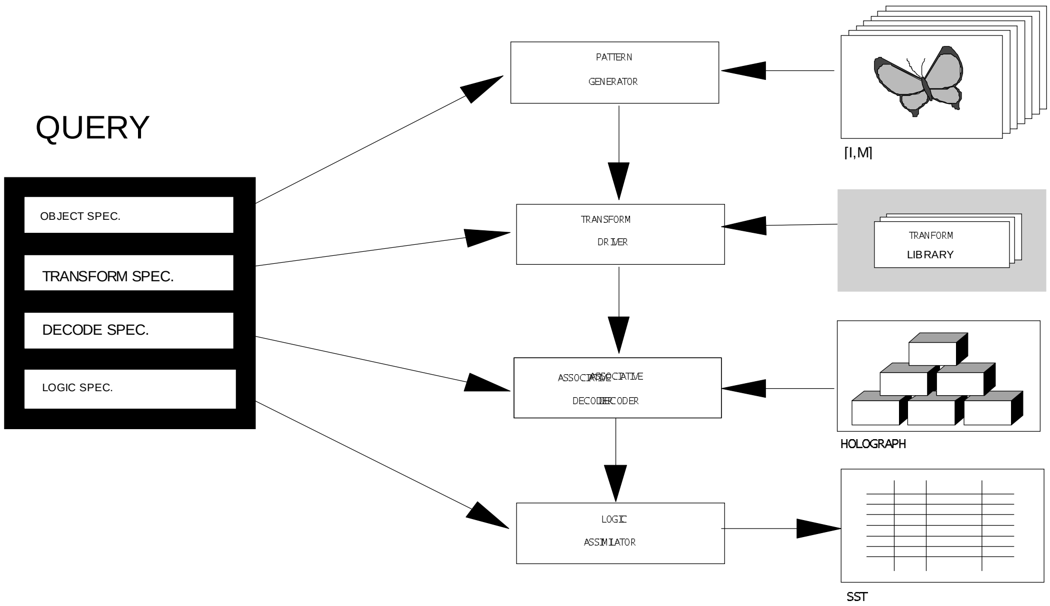

Fig-1 Inference Engine

In the second stage, each canonical pattern is replicated into a set of templates which are parametrically translated from the canonical pattern. The parametric translations are generated using pre-declared transforms from transform library. The transform driver performs parametric translations according to the range specification provided by the search controller.

The next step is to perform associative decoding. The associative decoder logs its result in a table called *Search Summary Table* (SST). It prepares one SST for each canonical search pattern. The search summary table contains (i) the suggested matching RLPs, (iii) corresponding MNC, and (iii) the parameter values at which the match has been registered.

The final stage is the logical assimilation of the information obtained in the SSTs. In this stage, first, the transform-relation evaluator functions are invoked to compute the strength (*s*) of the transform relations.

Each of the transformed-relation concepts is linguistically quantified into several quantifier levels or fuzzy sets (such as,HIGH, MODERATE, LOW). For, a given *s*, the membership strength of each quantified level can be determined by the fuzzy sets membership functions. These functions return the strength of the quantified-transform-relations. One function is required for each linguistic quantifier level for each relation.

The individual membership strengths are gradually assimilated to satisfy the logical constraints specified in the command using fuzzy inference rules. Standard fuzzy transformation functions and operations are used to compute the confidence values for each possible intermediate answer. The inference results are presented in the assimilated SST sorted in accordance of the computed MNCs.

5. EXAMPLE

Now a visuo-logical query will be illustrated to explain the integrated inference and search process. Below is a sample question:

OBJECT3:(HIGH. OBJECT1(in LOCATION2D $(x_1-x_2)(y_1-y_2)$)). MODERATELY. LEFT-OF. (HIGH. OBJECT2(in LOCATION2D $(x_1-x_2)(y_1-y_2)$)

This example involves transforms as well as fuzzy spatial reasoning. The inference engine first decomposes it into two independent searches, one involving OBJECT1 and the other involving OBJECT2. For both of these searches, the transform driver generates a set of transformed templates of the supplied canonical patterns by using the transform LOCATION2D() from the transform library. (LOCATION2D(xmax,xmin,ymax,ymin) is a transform with two parameters and can be used as a basis for spatial reasoning in images of 2D scene). These templates are then passed to the search engine. The associative decoder then performs holographic decoding and finally returns the results to the logical layer in the search summary table (SST). SST contains the RLP's and corresponding MNC's for each match, and the transform parameter value for which these matches were found (i.e. the locations in this case).

The search results in two SSTs, which include the location of match in the image (figures 2(a) and 2(b)). Each of the SSTs is mapped with fuzzy set functions to evaluate the MNC for the concepts HIGH.OBJECT1 and HIGH.OBJECT2. A linear function is assumed in this case which leaves the MNC values intact.

In this command, LEFT-OF is a transform-relation defined with the parameters of transform LOCATION(). It invokes a

Reprint from the Proceedings of the International Conference on Intelligent Systems, ICIS'96, Nevada, June 96, pp. 187-191.

LEFT-OF() strength evaluator function, which operates on the LOCATION parameter (x,y) columns of the SSTs of OBJECT1 and OBJECT2. The user is entitled to have any interpretations of the linguistic concept LEFT-OF().

The evaluation suggests that RLPs 1012 and 2017 are the possible matches with respective normalized strengths .568 and .110. The next step is to compute the strength (y) of the quantified-relation MODERATELY. LEFT-OF. This is evaluated by the fuzzy set function associated with the linguistic quantifier MODERATELY as defined in (Fig-2(c)).

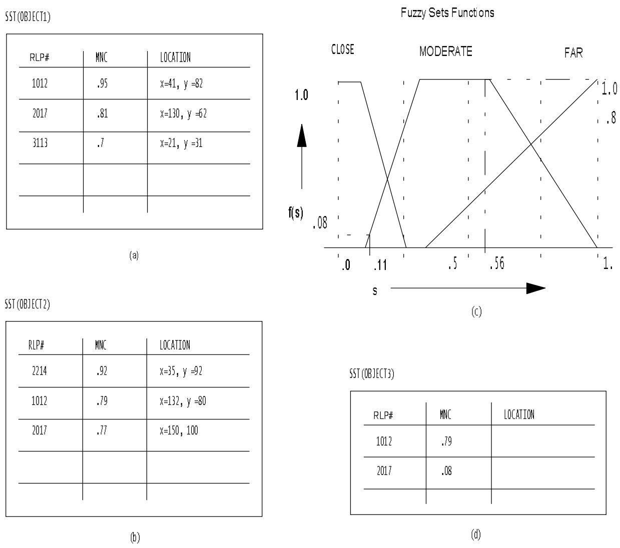

Fig-2 Example of Inference

Finally, the three concepts HIGH.OBJECT1, NEAR.DISTANT-FROM and HIGH.OBJECT2, are combined together by min-rule to obtain the MNC of OBJECT3 (Fig-2(d)).

6. CONCLUSIONS

The new approach demonstrated in this paper allows users to launch complex content-based queries into an image repository by incorporating "reasoning" capability into the search process with visual objects. The principal strength of this approach is that it empowers the end user. Such empowerment interm helps coping with the uncertainty, ambiguity and imprecision that has been found to be inherent part of pioneering CBIQ approaches that use some form of intermediate symbolic model to mediate search.

At the definition level, the user can dynamic specify the index objects ([I,M]). In the inference level, CQF formalism provides the user a flexible environment to set up his/her own basis for the reasoning process through the following control points: (a) Transform (XFORM) generation routines, (b) Transform-relation (XFORM-REL) strength evaluation routines, and (d) Transform-relation-quantifier (LQ) membership functions. Together, these make the "assumptions" underlying the interpretation and inference of outer tier (CQL) "programmable". Finally, at the innermost tire, the efficient pattern matching capability needed to support such empowerment is provided by MHAC. This search method is also highly parallel and scalable, and even potentially optically realizable [7,9].

In general, an associative visual search approach like MHAC will be more effective when the images tend to be natural (non-graphical), the objects are difficult to describe or model, the image volume is enormous and examples with visual similarity at the object level are available.

This research has been funded by ACTS and Supercomputing in Remote, Co-operative Medical Triage Support and Radiation Treatment Planning project of DARPA under research grant DABT 63-93-C-0056.

REFERENCES

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

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

[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] Iyenger, S.S., R. L. Kashyap, "Guest Editors' Introduction on Image databases", IEEE Trans. on Software Eng., v.14, n.5, pp608-609, May 1988.

[6] Keshavan, H. R., J. Barnett et. Al., "Introduction to the Special Section on probabilistic Reasoning", IEEE Tran. On Pattern Analysis and Machine Intelligence, v15, no.3, March, 1993, pp193-195.

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

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

[9] Kumar, B. V. K. V & P. K. Wong, "Optical Associative Memories", Artificial Neural Networks and Statistical Pattern Recognition, I. K. Sethi and A. K. Jain (Editors), Elsevier Science Publishers, pp219-241, 1991.

[10] Zadeh, L. A., Fuzzy Sets, Information and Control, v.8, 1965, pp338-358.