*Volume:04 Issue:03 Pages:* 1640-1644 *(2012) ISSN : 0975-0290.*

# Study on Event Matching In Temporal Database Using AGT Approach

**N. Duraimutharasan [1], Dr. K. Sarukesi[2]**
[1] Department of Computer Science & Applications, Periyar Maniammai University Vallam, Thanjavur, India
Email: duraibose@pmu.edu
[2]Professor, Hindustan University, Chennai, India
Email: profsaru@gmail.com

**----------ABSTRACT----------**
At some instant, the user may want to acquire the event as image from the temporal database. Prior to the retrieval, the system has to be trained with the data based on the events. Firstly, the video file is converted into frames with the time range with a set of frames. Each frame i.e., image is subjected to image processing techniques to get the various features of that image such as RGB values, histograms, hue, saturation, intensity, time information, etc. These extracted features are stored in the temporal database. The image retrieval is based on the events and the user has to give the query in natural language query. Based on the natural language query, some set of images are retrieved from the database and these retrieved images are used as the query images. User can select any one of the image as the query image and give for the processing. Now the query image is subjected to the image processing technique and these techniques are used to extract the various features of query image. The proposed system is provided with the query image along with its features. Thus the SQL query is formed with NL query and features of the image. Then the image that matches exactly with the requirements is fetched from the database.
**Keywords** Histograms, Hue, Saturation, NL query

Date of Submission: November 03, 2012 Date of Acceptance: December 10, 2012

## I. INTRODUCTION

While building the Natural language interfaces, it is very challenging to provide the proper guidance to the system to link the user's view and the domain information that is structured in the database for further processing. The present day databases make use of different principle for representation; most of them are designed to suit with a variety of database management standards. The natural language query interface must have the capacity to handle any kind of queries. Most probably, the queries request the database to return the available predetermined data. In fact, the input English query is same in several cases, but the output may differ according to different system training. The output should epitomize the definite principles and structures of the database. One of the key tasks of the natural language query interface is to formulate the essential transformations and thus to isolate the end users from the minute details in the database. To give the isolation and also to link the gap between systems's structure for the data and the view of the user needs a blend of general information and the field-specific information. The system should comprise with the subject matter of the application model. This contains the information concerning the domain's object, the attributes and quality they have and their inter-relationship, the referring phrases and the words. It is highly significant that the system should have the knowledge about the links among the entities and the data in the database. A most important assertion in the system that if a natural language query interface is restricted in a well-organized manner, the information required for adjusting it to the new database and its related field can be obtained from the users who possess common skills on the database and computer system, whereas the users do not have any unique facts on the processes of natural language or about the natural language interface. While examining this theory, it is evident that the database might not be rebuilt or rearranged. By the theoretical perspective, it is forced to implement some wide-ranging solutions to problems in system design due to the reason that the data could not be restructured which may lessen the issues of the natural language processing. However, from the practical perspective, the difficulty and size of the existing database In this work, the events are represented both in the text and image format. If the user is satisfied with the text information about the event, then the process of matching the events in the temporal database is adequate. But this system gives an option for the user to visualize the event in the temporal database as an image. As said earlier, the video file contains the events in the form of images/frames. The AGT interface facilitates the user to give the in general terminology. The query sentence is parsed and checked with the lexicon and syntactical checker in order to extract the keywords from that query sentence. The keywords contain the details of the table name, field name, event time info, etc. Using the keywords, the events in the temporal database is traced and matching events are retrieved and displayed to the user in the text format along with the sample images of the events. In addition, if the user desires to view the exact image of that particular event, then he/she has to select a query

image. The selected query image is then processed to extract the features of that image. The regions and the mean value of the RGB color components are used for further retrieval procedures.

## II. PERSPECTIVE OF THE PROPOSED WORK

This work also focuses on the image processing techniques and content based image retrieval for event matching. An intelligent image retrieval system may be viewed as a computing platform with a friendly user interface that allows users to represent and retrieve images from a given database. In addition, a good retrieval system should provide several modules to perform automatic feature extraction and selection, similarity matching and user's interaction.

### A. The Neuro-Genetic Approach

Due to the capabilities of neural nets for pattern memory, generalization and adaptation, some promising results on learning similarity metrics using neural nets for CBIR systems have been developed. Indeed, the concept of learning similarity is closely related to visual feature classification. Recently, it has been reformulated as a problem of pseudo metric approximation using neural nets. In some existing neural nets based CBIR systems, the weights of neural nets are obtained through two phases: off-line training followed by on-line updating. These two steps correspond to the processes of pattern memory and neural similarity metric adaptation. Although some problems in this scheme still remain open, for example, the impact of the subjective RF on retrieval performance, the reported results indicate the usefulness of the RF techniques. Neural nets, as a powerful modeling tool, have demonstrated its good potential for image retrieval tasks. It has been successfully applied in intelligent image retrieval systems, especially for feature recognition and learning similarity measure. A modified cost function for "error back-propagation" training algorithm is presented to implement the RF, where feedback signal comprises the neural net actual output and the query. In all CBIR systems there are two opposite demands: the accuracy of the retrieving procedure and the retrieving speed. Better accuracy is obtained when using larger feature vectors but then the retrieving process is time consuming, particularly when working with large database. This work adopts a content-based image retrieval (CBIR) system using user's assisted neural network inference system. Although the choice and/or optimization of a feature vector is of great importance, the main attention in our research was addressed to design of neural network decision stage driven by user's relevance feedback as an efficient tool for image retrieval system. Though the similarity measure may be missed due to a variety of reasons, such as inadequate training time, unsuitable features or content of the images used or generalization ability of the neural network may be constrained. This makes to fuse this neural network with the genetic algorithm for better results. The mean value of the color image is used as color feature. In addition, we have also considered an active contour as the region feature. Genetic algorithm is used to reduce the semantic gap between retrieval results and user expectations. In a CBIR system, the retrieval of images has been done by similarity comparison between the query image and all candidate images in the database. To evaluate the similarity between two images, the simplest way is to calculate the distance between the feature vectors representing the two images. Color, shape and texture are three low-level features widely used for image retrieval. In this work, the region-features (shape) and mean values (color) are used. It has been already proved that the approach based on the genetic algorithm can provide better results for pattern recognition than classical techniques, such as Support Vector Machines and Neural Networks. Also Figure1 explains about the overall architecture of the Proposed Neuro - Genetic system with its overall functions.

### B. Components and Functions of EMAGT System

The components and their functions in the EMAGT system are outlined below **AGT interface** – AGT interface is the component by which the user interacts with the system. It allows the user to give the query in general form of English for the event to be matched and retrieved. The user can give any kind of sentences with the temporal aspects such as time information, which helps the system to find the event that occurred at that specific time duration. If the system finds the exact matching for the event, then it is fetched out and displayed to the user.

**Parser** –The AGT queries, which are given as English sentences, are parsed using Parser. It comprises the lexical and syntactical analyzers. Parser is the component in the EMAGT system which is used for describing words in query sentence grammatically to the system. The lexical and syntactical aspects of the words are checked.

**Lexical Checker–**The lexical analyzer is a repository of the information about each word that is essential for morphological, syntactic and semantic analysis. There are two kinds of lexical items: closed class and open class. Closed classes (e.g., pronouns, conjunctions and determiners) contain only a finite, usually small number of lexical items. Typically, these words have complex and specialized grammatical functions, as well as [at least some] fixed meanings that are domain-independent. They are likely to occur with high frequency in queries to almost any database. Open classes (e.g., nouns, verbs, adjectives) are much larger and the meaning of their items tends to vary according to the given database and domain. Therefore, most closed-class words are built into the initial lexicon, while open-class words are acquired for each new domain. However, there are a number of open class words – for example, words corresponding to concepts in the initial

*Volume:04 Issue:03 Pages:* 1640-1644 *(2012) ISSN : 0975-0290.*

conceptual schema and those for common measure units, such as meter and pound that are so widely applicable to so many database domains that they are included in the initial lexicon as well. The lexicon entries acquired by this system include those for the names of fields, field values, and file subjects. It provides adjectives and verbs that correspond to various relations and properties represented in the database; they can also supply synonyms for words already acquired. Associated with each lexical entry is syntactic and semantic information for each of its senses.

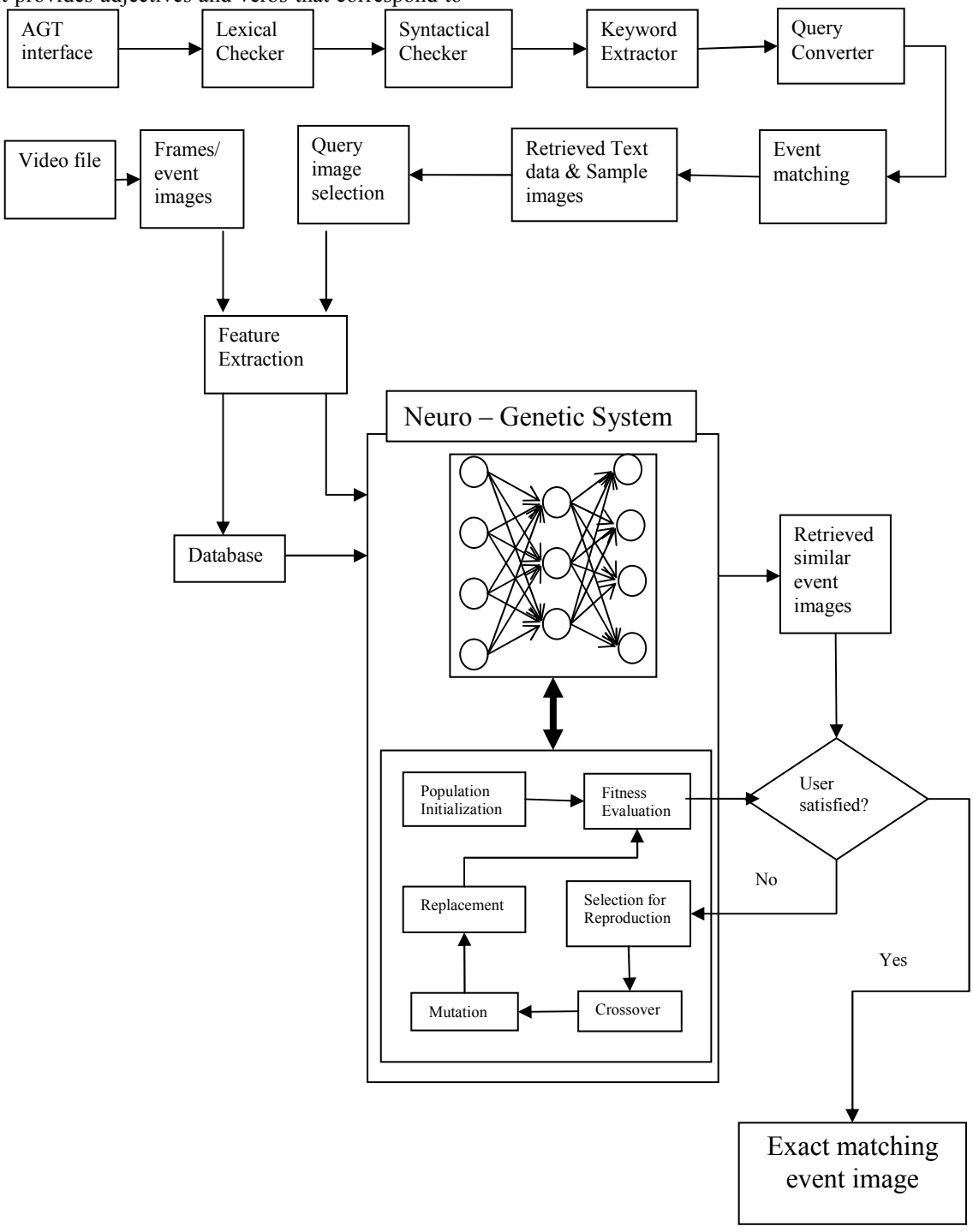


**Fig 1. Proposed Neuro - Genetic System Architecture**

**Syntactical Checker–** Syntactic information consists of a primary category (e.g., noun, verb, adjective), a subcategory (e.g., count, unit, mass for nouns; kinds of objects for verbs), and morphological information (e.g., irregular

*Volume:04 Issue:03 Pages:* 1640-1644 *(2012) ISSN : 0975-0290.*

plurals and comparatives). Semantic information depends on the syntactic category. The entry for each noun includes the sort(s) or individual(s) in the conceptual schema to which that noun can refer. Entries for adjectives and verbs include the conceptual predicates to which they refer and information as to how the various syntactic constituents of a sentence map onto arguments of the predicate. Scalar adjectives (e.g., "high") also include an indication of direction on the scale that is plus or minus. Thus the syntactical analyzer checks for the grammatical correctness for English language.

**Keyword Extractor** – as the name suggests, is mainly used for extract the keywords from the parsed sentence. These keywords such as noun, nouns other than the field names, table name, final field name, temporal event, conditions to be satisfied are given to the system for event matching in temporal database.

**Query Converter** – Query Converter is the part of the EMAGT system to translate the general query keywords into a logical query for the temporal database. The AGT query does not match with the SQL-like query. Thus there is a need for a query converter. The major role of the query converter is to convert the AGT query into a clear database-oriented query. As the temporal database is considered in this work, the converted query is an exact temporal database query with all its constraints and rules.

**Event Matcher** – Events are stored both in textual and image form in database. Our main objective of the work is to match the events stored in the temporal database. This can be achieved by using the event matcher. The matching of the events is done by appropriate mechanism.

## III. PROPOSED SOLUTION

The overall retrieval framework can be divided into two different approaches based on whether or not it considers the use of validation sets in the similarity discovery process. The use of *validation sets* aims to avoid the effect of overtraining (over fitting). Overtraining can occur when the learned or evolved model fits the particulars of the training data overly well and consequently does not generalize to new unseen examples. The performance of the system relies on the use of a validation set to identify appropriate individuals to be used on the test set. The performance of individuals in the training set (stored in the set *S*) and in the validation set (stored in the set *F*), and selects the individual that satisfies the condition.

### *C. Proposed Algorithm Overview*

(1) Let T be a training set
(2) Let V be a validation set
(3) Let S be a set of pairs (i, $fitness_i$), where i and $fitness_i$ are an individual and its fitness, respectively.
Let g be a generation
Let Ng be the number of generations
(4) S ← ∅
(5) P ← Initial random population of individuals
(6) For each generation g of Ng generations do
6.1 For each individual i∈ P do
6.1.1. $fitness_i$ ← fitness(i, T)
6.2. Record the top Ntop similarity trees and their fitness values in Sg
6.3. S ← S ∪ Sg
6.4. Create a new population P by:
6.4.1. Reproduction
6.4.2. Crossover
6.4.3. Mutation
(7) F ← ∅
(8) For each individual i ∈ S do
8.1. F ← F ∪ {(i, fitness (i, V))};
(9) BestIndividual ← SelectionMethod (F, S)
(10) Apply the "best individual" on a test set of (query) images

## IV. RESULTS AND DISCUSSION

In this section different simulation results are been demonstrated to show the efficiency of the proposed Approach. The Proposed approach is developed with object oriented programming. Also the simulation is considered to have different event for different video file. The training set is done by different video files with event capacity is 'n' numbers. The performance of the proposed shown with simulation graphs based on video files events. For evaluation, we apply our approach to a collection of time-series events for temporal data mining, the EMAGT visual tracking database, and the time series data set cover both training set and validation set. First present the temporal data representations used for time-series and various types of events which are taken from the EMAGT database for processing have different range of values.

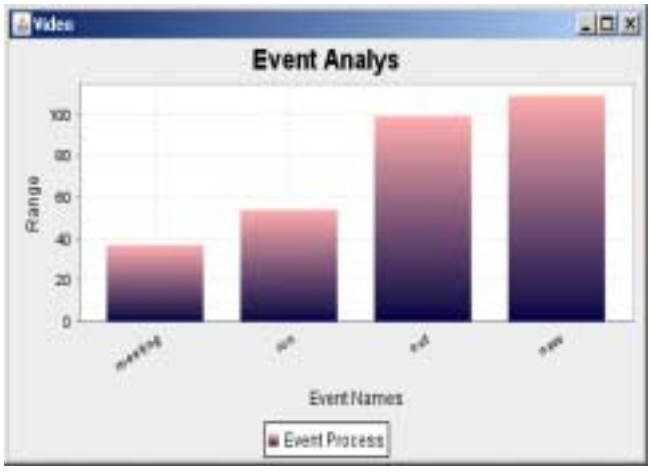


Fig. 2. a. Different Event vs. Time-Series Fig. 2. b. New Event vs. Mismatch Event explains about the various event performances which is of a New Events vs. the Mismatch event. The experiments are conducted by implementing the idea in Java 1.6 and setting the values for different event tasks.

*Volume:04 Issue:03 Pages:* 1640-1644 *(2012) ISSN : 0975-0290.*


## V. CONCLUSION

The main purpose of this work is to construct a framework for event matching in temporal database. Also there were many existing works that are done on the event matching process in several databases. In fact, the framework for event matching in temporal database had been proposed with new few works. Bearing in mind the issues in the existing system, this work focuses on most sophisticated region on the event matching. With the emergence of new technologies, multimedia has gained more popularity among the users. So there comes the need to find new ways for retrieving those data. The retrieval procedure can be improved with the use of neural-genetic network and human's assistance as described. This new framework is flexible and powerful for the design of effective event matching. The effectiveness results demonstrate that the framework can find better similarity functions than the ones obtained from the individual descriptors. Our experiments also demonstrate that the neuro-genetic framework yields better results than any other existing systems. Experimental results have shown that the performance of the proposed approach is appreciable.

## REFEENCES

[1]. N. S. Vassilieva, \Content Based Image Retrieval Methods", Programming and Computer Software Vol. 35, No. 3, 158 { 180 (2009).

[2]. Y. Liu, D. Zhang, G. Lu, W. Y. Ma, \A Survey of Content-Based Image Retrieval with High Level Semantics", Pattern Recognition 40, 262 { 282 (2007).

[3]. R. C. Veitkamp, M. Tanase, \Content | Based Image Retrieval Systems: A Survey", Technical report, UU-CS-2000-34, University of Utrecht (2000).

[4]. S. Antani, R. Kasturi, R. Jain, \A Survey of the Use of Pattern Recognition Methods for Abstraction, Indexing and Retrieval", Pattern Recognition 1, 945 { 965 (2002).

[5]. X. S. Zhou, T. S. Huang, \Relevance Feedback in Content-Based Image Retrieval: Some Recent Advances", Information Science 48, 124 { 137 (2002).

[6]. T. C. Lu, C. C. Chang, \Color Image Retrieval Technique Based on Color Features and Image Bitmap", Information Processing and Management 43, 461 { 472 (2007).

[7]. C. C. Chiang, Yi. P. Hung, H. Yang, G. C. Lee, \Region based image retrieval using color size features of watershed regions" Journal of Visual Communication and Image Representation 20, 167 { 177 (2009).

[8]. Chunming Li, Chenyang Xu , Changfeng Gui, Martin D. Fox, \Level Set Evolution Without Re initialization: A New Variational Formulation", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).

[9]. D. E. Goldberg, \Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, Reading (1989).

[10]. J. Holland, \Adaptation in Natural and Arti_cial System", The University of Michigan Press, MI (1975).

[11]. J. H. Su, W. J. Huang, P. S. Yu, V. S. Tseng, \E_cient Relevance Feedback for Content Based Image Retrieval by Mining User Navigation Patterns", IEEE Transactions on Knowledge and Data Engineering Vol. 23, No. 3, 360 { 372 (2011).

[12]. Huazhang WANG, \Content Based Image Retrieval Using Combination of Color and Texture Feature", Journal of Computational Information Systems Vol. 5, No. 3, 1349 { 1357 (2009).

## Authors Biography

**N. Duraimutharasan** is currently working as an Assistant Professor, Dept. of Computer & Applications, Periyar Maniammai University since august 2010 and currently pursuing Research in Temporal databases in Hindustan University, Chennai.

**Dr. K. Sarukesi** is currently working as a Professor, Dept. of Computer Science & Engineering in Hindustan University, Chennai.