Event-based Multimedia Chronicling Systems

Pilho Kim
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
phkim@ece.gatech.edu

Ullas Gargi
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA, USA
gargiu@hpl.hp.com

Ramesh Jain
Bren School of Information and Computer Science
University of California, Irvine
Irvine, CA, USA
jain@ics.uci.edu

ABSTRACT
This paper presents a new framework for multimedia electronic chronicling systems. Its approach uses events as its driving force for heterogeneous information processing. Specifically, this new approach first separates symbols and data, then puts events between them to make a distinct connection. In addition, this approach provides spatio-temporal-semantic relations networks to map high-level semantic user queries into low-level queries that a machine can compute. The innovative user interfaces are designed for ease of use and are interactive to allow organization of information and a search capability for information retrieval. The results reported as a part of this paper show that we can achieve effective chronicling and high-quality user experiences by coupling together multimedia analysis, tagging and querying. We believe that our system can serve as a semantically accessible annotated multimedia log. The approach taken here may provide a foundation to provide further summaries of important events or access to events at the required level of granularity.

Categories and Subject Descriptors
H.1.2 [Models and Principles]: User/Machine Systems—Human information processing; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Search process; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Hyperlink navigation and maps

General Terms
Algorithms, Documentation, Experimentation, Human Factors, Management

Keywords
Chronicle, interaction, interfaces, information, mobile, publication, monitoring, tag, event, browse, search, retrieval, publishing, information sharing, context modeling, action capture, presentation systems

1. INTRODUCTION
The evolution of multimedia data, with their wealth of support for multimodal interactions, is definitely encouraging people to develop more ambitious systems. However, integration of these media raises concerns about the complexity and heuristic nature of this integration. Moreover, metadata or abstractions for each media are also diverging in terms of type, structure and the rule of associations.

As data itself is becoming a composite object as a consequence of the semantic enrichment of multimedia, it is becoming acute to develop new fundamental information storage system to handle the time-serial nature and enormous information size of multimedia data. Besides, in virtually every area of data managements, interoperation between applications requires that the underlying applications interoperate meaningfully. Data warehouses require the correct semantic merging of data from semantically diverse sources. As a consequence, it is becoming essential to integrate such multimedia data into organizational frameworks that emphasize semantic coherence.

A multimedia electronic chronicle or eChronicle is at the center of these newly emerging fields. Many variants of the eChronicle system [24, 29, 11, 31, 4, 15] have already appeared and are being used. Although much of this work has been directed toward the capture and organization of information at a restricted domain level, little emphasis has been put on tools for people using a unified perspective for interoperable purposes. The role of the eChronicle systems fall into three categories: (1) recording data using multiple sensors, (2) supporting rich tags for access and presentation of appropriate information, and (3) providing access to this data at multiple levels of granularity and abstractions. To develop and exemplify these components, several preliminary research efforts have been made in our group for event-based multimedia modeling methods [33], personal chronicling systems [20], and multimedia event tagging systems [19]. Based on this work, this paper raises two fundamental issues in multimedia chronicling: (1) the event-centric data organizational model and (2) the spatio-temporal and semantic relationship network.

The rest of this paper is organized as follows. Section 2 reviews related research works to identify the roles of the system for users. It specifically handles the challenging problems in multimedia and its analysis, and how to represent relations in between data properly. Section 3 introduces
our approaches for the data organization and their relation representation. This idea is formalized into the eChronicle relation network presented in Section 4. Section 5 explains research efforts extending our previous works on event capturing, media processing, and user tagging supporting. This is followed by experiments in Section 6 and conclusions in Section 7.

2. RELATED RESEARCH

In this section the functionalities of the chronicling system and formulate requirements will be articulated within the context of existing technologies developed in related research areas. A chronicle is by definition an extended account in prose or verse of historical events. It generally covers the entire recording of personal, organizational or social events. As it is getting easier to record disparate activities in different situations using different types of sensors, data are no longer simple alphanumeric values but occur in various types of composite media or multimedia. This evolution of multimedia data, with their wealth of support for multimodal interactions, is definitely encouraging people to develop more ambitious systems. However, integration of these media raises concerns about the complexity and heuristic nature of this integration. For example, many of the approaches for encoding and decoding become proprietary to each specific application. Besides, conventional electronic chronicling systems provide a query environment that in most cases is limited to the analysis of alphanumeric values of data and metadata. These approaches are inadequate to handle many challenges introduced in this section.

2.1 Multimedia Analysis and Associated Metadata Repository

Storage, retrieval and processing of relevant media are key challenges to the success of a data model. As mobile devices and sensors become more pervasive, more and more aspects of an individual’s life are being captured in richer detail. Several different ways have been developed to tackle specific problems of the retrieval process: feature extraction methods for multimedia data, query languages, problem specific similarity measures and interactive user interfaces (cf. current approaches on the multimedia retrieval can be found at one place [22], and [39] is referred as the survey on the content-based multimedia retrieval). A common problem found specifically in multimedia analysis is that this analysis simply is incapable of achieving the synergistic effect of using multifarious signals. For instance, Figure 1 shows a sample result of a group meeting video. At the time it was captured, one speaker introduced himself. Imagine that we want to detect his “introduction” event that is represented by four signal processing results. Then we soon realize that without strict limitations on environments the detection cannot be made simply by creating balance between each analysis output [4, 37]. Even if a system succeeds in detecting the event, those running parameters to detect his introduction are generally useless for videos captured in a different recording environment. In fact, those multifarious signals and processing results are correlated, not only by signal strength and features, but also by spatially, temporally and semantically [34, 27]. So the real issue in this multimodal analysis approach is to capture those essential relations and to store and retrieve them.

As shown in the above example, the organizational principles used for the storage of multimedia and associated metadata or abstractions should be carefully designed. This oversight will be addressed by examining time, space and semantics as the possible avenues for unification of this information since those are becoming first class concepts in any information system to support the targeted applications effectively.

Temporal Databases Temporal data and its related research areas have been studied for more than twenty years. James Allen was one of the foremost researchers in this field. His works [2] provided the foundation for many succeeding temporal relation research efforts (cf. a cumulative bibliography [6], and survey [38]). He defined thirteen possible interval-based temporal relationships that are contained in Table 3.

Allen’s thirteen relationships assume for any interval $t$ that the lesser end-point is denoted by $t_-$ and the greater by $t_+$, and $t_{p}$ is for time-point and $t_{l}$ for time-interval. These temporal relations can represent the continuity and discontinuity features of events. For instance, a user may map a data registration event to an infinitesimal time-point and map moving-object tracking events to a time-interval.

Spatial Databases Spatial concepts are also being integrated into the database data model as a geometrical element [30, 13]. Table 3 shows the spatial elements defined by OpenGIS consortium assuming $g_1$ and $g_2$ as a spatial object [26]. These spatial relations are good enough to represent both the geographical locations associated with data and the spatial relations between detected objects in multimedia data.

Spatio-Temporal Databases Because spatial and temporal concepts are the primary aspects of information, research efforts are emerging that merge both concepts into one. Some researchers employ the tuple model in which all spatio-temporal features are aligned as one feature vector in tables [5, 14] for computing speed and utilizing relational database systems. Some others [7] store spatio-temporal relations into E-R entity diagrams by using Unified Modeling Language (UML). This will make it easier to extend relationship sets and represent the hierarchical relations between data. However, these gains will come at the expense of computing speed and implementation complexity. So in this paper, we propose a new information storage architecture introduced in Section 3.

Research progress made in spatio-temporal research areas is very promising. However, to ensure clarity about the ori-
gion of temporal and spatial information, variations in the data stream should be also detected and recorded as events. Besides, spatio-temporal concepts are too incomplete alone to represent a fact as fully and richly as they can in conjunction with all related data. As a further extension, the semantics of data will be handled in the next sub-section.

**Semantics of Data** Data itself is becoming a composite object as a consequence of the semantic enrichment of multimedia. Therefore it is more challenging to clearly represent the relationships between data and semantics [3, 8]. Let us consider one specific example.

Example 1. Select the meeting report submitted by Jimi Hendrix after the group meeting held on March 1st, 2005.

This query is in much abbreviated form from the viewpoint of information retrieval. It actually requires high-level semantic understanding because (1) this query does not specify the type of data but merely asks for the meeting report; (2) the name of the creator, Jimi Hendrix, is specified, but it let up to the system to understand that this is a person’s name; and (3) there may be two people who have the same name, which may cause semantic confusion.

Several approaches are ongoing to address such problems as those described above and enable data to be shared semantically and reused across application, enterprise, and community boundaries. For instance, the Semantic Web\(^1\) provides a common framework through a collaborative effort led by W3C that also involves participation from a large number of researchers and industrial partners. The Semantic Web is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. The main thrust of the work led by W3C for semantically shareable data has been focused on dictionary supports such as RDF Vocabulary Description Language. Based on RDF Vocabulary Description Language, the Web Ontology Language, OWL\(^2\), has been developed as a vocabulary extension of RDF (the Resource Description Framework) and is derived from the DAML+OIL Web ontology language.

Interestingly, a similar approach is found in the knowledge engineering area. (Do “meeting people” require “social interaction”?) in the form of OMCSNet [32] Knowledge Base (KB) represents a common sense approach that “meeting people” requires “social interaction” as an action. They indexed their existing KB (Knowledge Base) to be searchable and shareable via WordNet [12], which is known as the most popular semantic resource in computational linguistics. Technically, these developers first parse the sentence to identify grammatically each word as a noun, verb, adjective or adverb and so on. Then they index each word with the sense number in WordNet, (e.g., Do “meeting.n#1 people.n#1” require.v#1 “social.adj#1 interaction.n.1”). For instance, “meeting.n#1” in WordNet means the first sense of the noun “meeting”. They are doing this so that they can share their KB with other KB-related applications without semantic confusion.

As the Semantic Web has shown with its use of ontologies and OMCSNet has demonstrated with WordNet, a common and shared dictionary is essential to avoid semantic confusion. Ontology-based approaches are good at extending the dictionary of the application, but these approaches have limited ability to be shared beyond the group of interest. WordNet-based approaches are generally good at representing the fact because of its enriched definition of commonly used and shared words, but such approaches may lag in the inclusion of new words because of the need for frequent updates of content. In both cases Besides, a user plays an important role in both cases in linking data to its correct semantic symbols. To accommodate the ability of users to create such linkages, any system that does information chronicling should support rich tags for access and presentation of appropriate information, and the interface for creation of such tags should be easy to use [20].

### 2.2 Uncertainties in Data and User Queries

Implementation and usability should be given high priority if a system is to be practical. Foremost among such considerations should be recognition of the uncertainties that inevitably exist in data representation and in users’ queries. Users will be frequently unable to express their interests at an appropriate level of accuracy. This inability arises partly from the uncertainties and imprecision inherent in language itself and partly from the limitations of human cognition. This inability that often afflicts users makes it important for the system to have a way to represent uncertainties in data, a capability that must exist for all sorts of sets of data relations. For instance in temporal relationships, some work employed the concept of granularities for temporal models [10, 40] to mask these uncertainties. Granularities used in this way may be relative or absolute so that they can indicate data within some absolute value range or, alternatively, present relative variations from a given data value.

### 3. EVENT-CENTRIC RELATIONSHIPS

To develop more generalized system that overcomes inherited heterogeneity in multimedia, a system should satisfy two requirements. First, the system should suggest a unified method to store data and its related symbols, while simultaneously considering its heterogeneity. Second, generic relationship representation methods like spatio-temporal and semantic relationships should be richly supported to represent the fact correctly. For this, we introduce a new event-based information processing framework.

#### 3.1 Problem Statements and Terminologies

In multimedia, one symbol is not enough to represent the semantics of data. The size of data is becoming larger, and the format is gaining in complexity. In multimedia, spatio-temporal or semantic variations exist in data. Hence, data may be mapped to multiple symbols. Conversely, one symbol may be mapped to multiple data. In conclusion, a new object, which can represent spatio-temporal or semantic variations, should exist in between data and symbols to make the connection distinct. Our approach defines this as an event. Technically, an event within an eChronicle is a log of the activity detected by event detectors running for each type of data channel. To overcome the heterogeneity of media while providing a unified method to represent their related information, an event-centric relationship (ECR) model is presented in this paper. Specifically, we provide fundamental logistics to develop a repository for multimedia and its related information storage. Our ECR

\(^1\)http://www.w3.org/2001/sw/

\(^2\)http://www.w3.org/TR/owl-ref/
model breaks a conventional entity-relationship models that are prevalent in relational databases or XML approaches.

In general, data and symbols are tightly bundled together. For instance in relational databases, alphanumeric values are tabulated to each matched text column name. In XML, every text value is accompanied by the text element name. This paper departs from stereotyped architecture by introducing events between data and symbols.

Alphanumeric values, once considered as data at the early stage of computing, now exist inside multimedia as one subset of data. To handle these composite data, a new concept, c-data, is introduced. C-data represents the concatenated data set linked with the index of data in different types.

A symbol of an eChronicle is the semantic representation of data. It is the name of the data property. An eChronicle uses a unique approach to store a symbol and its associated data separately, but both are connected by events to make the origin of information clear. The column name of data in relational databases and the element name in XML are matched with this symbol concept. A symbol object has a word sense index mapped to the electronic word reference, WordNet[12], to make the semantic sense of a symbol clear and shareable. It may represent a tag extracted automatically by an event detector or provided manually by users. Also, it can be linked with other symbols using predefined relations. Relationships between symbols are limited so that they will be computable. Essential relationships are space, time, and semantic relations as introduced in Section 3.2.

In summary,

- C-data may be mapped to multiple symbols.
- Conversely, one symbol may be mapped to multiple c-data.
- An event is in between to make a connection distinct.

A symbol, S, and c-data, D, may be linked in many ways via events, E. But temporally there can be only one relation at a specific transaction time, $t_j$, between $E(t_j)$ and $S_i$, and between $E(t_j)$ and $D_k$ because of the finite time feature of the event in the eChronicle system.

$$I f \left( S_i \leftrightarrow \left\{ E_1, E_2, \ldots, E_N \right\} \leftrightarrow D_k \right),$$

$$\exists ( S_i \leftrightarrow E_j(t_j) ) \leftrightarrow D_k \right)$$

where,

$$t_j = timestamp(E_j).$$

3.2 Fundamental Event-based Information Storage Framework

Suppose we have a data set, $D = \{ d_1, d_2, \ldots, d_X \}$, where containing X associated tuples for real data. In the same way, we have a symbol set, $S = \{ s_1, s_2, \ldots, s_Y \}$ and an event set, $E = \{ e_1, e_2, \ldots, e_Z \}$. These three sets can be linked with each other based on following four definitions. Following definitions are depicted in Figure 2 as a linked graph.

Definition 1. A basic event-data-symbol (EDS) set is represented as $i = \{ e_i, i_d, i_s \}$, where $i_d$ and $i_s$ is an index to each linked data and a symbol, and $i$ is a temporal index of a specific event. $\mathcal{E}[i_d, i_s] = \{ E_{i_d, i_s} \}$ is a multiple EDS set, which is composed of multiple events linked to one data-symbol set.

A basic EDS set is composed of symbol, data and their connecting event. It may be generated by an event detector that may create a composite EDS set linking each EDS set with a connector, $CI_i$. For instance, when some event happens and an event detector detects it, it may have several types of abstractions or metadata of what it detects. Those event sets are linked together via a connector.

Definition 2. A composite EDS set, $C_i = \{ T, CI_i \}$, where $T = \{ t_1, t_2, \ldots, t_2 \}$ and $CI_i$ is a scalar data index to a connector property.

Besides, when an event detector registers what they detected, each symbol and data pair is queried to prevent duplication. If there exists a same pair, then it creates a linked set as depicted in Figure 2. Therefore, one symbol-data pair can be linked to many events as depicted as multiple sets and these connections create search paths in between composite sets.

Figure 2: Event-based Information Relation Framework.
Table 3 and in Table 1 as ways to gain the maximal benefits from existing work and also introduce new concepts as necessary. These relationships initially originated from Allen’s works [2] for temporal relationships and from OpenGIS [26] for spatial relationships. For the sake of simplicity and compatibility, the same notation is used if the relationships from the two different sources are semantically compatible. Table 3 assumes $x_p$ and $y_p$ for any infinitesimal time point, $x_1$ and $y_1$ for any time interval, and assumes $g_1$ and $g_2$ for the geographic object.

In addition, the semantic relationships are proposed in Table 1. These come from the various semantic relations defined in WordNet. WordNet groups English words into sets of synonyms called synsets, provides short definitions, and records the various semantic relations between these *Synonyms*. For instance, when a user wants to find meeting related information, she may use the Synonyms operator for the term, “meeting,” to find all semantically related information (e.g., gathering, assemblage, social affair are linked with meeting as synonyms in WordNet). It is also possible to search incorrectly spelled words using *Morph* operator and to search words with a part of word using *Partial* operator.

<table>
<thead>
<tr>
<th>Table 1: Semantic WordNet Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Synset Relations</strong></td>
</tr>
<tr>
<td>Equal</td>
</tr>
<tr>
<td>Synonyms</td>
</tr>
<tr>
<td>Morph</td>
</tr>
<tr>
<td>Hypernyms</td>
</tr>
<tr>
<td>Hyponyms</td>
</tr>
<tr>
<td>VerbFrame</td>
</tr>
<tr>
<td>Antonyms</td>
</tr>
<tr>
<td>Holonyms</td>
</tr>
<tr>
<td>Causeto</td>
</tr>
<tr>
<td>Pertaining</td>
</tr>
<tr>
<td>Attributes</td>
</tr>
<tr>
<td>DerivedForms</td>
</tr>
<tr>
<td>Domain</td>
</tr>
<tr>
<td>Familiarity</td>
</tr>
<tr>
<td>VerbFrames</td>
</tr>
<tr>
<td>Coordinate</td>
</tr>
<tr>
<td>Terms</td>
</tr>
<tr>
<td>Partial</td>
</tr>
</tbody>
</table>

It is natural for humans to consider the granularity of information. For instance, users commonly ask a question to know in which day, month, or year something happened. In the same way, they may ask in which city, state, or country it happened. For this, granularities for each type of relationship are proposed in Table 4. Notations used in Table 4 and ERN are in Table 2. Granularities for each relation type are designed based on our experience with the preliminary eChronicle prototype. For instance, IP-to-Map distance relationships in spatial granularities are employed to identify the geographical location of mobile users. By using this, it is possible to identify the location, although in lower resolutions based on 451,000 records of cities in the world have been collected. These records include the longitude and latitude of each city and country specific information such as the capital and the population.

The other aspect of granularities is the semantic variance or distance between a keyword and a target word. Typically, several matches could be found in a semantic search using WordNet. A computer may need to find the best one or at least find words within some specified depth of distance. So we provide the *wrdd* relative distance in Eq. 2 by employing the Jiang and Conrath equation [18]. Their work was proven to be the best within existing WordNet semantic distance measurers [8]. Eq. 2 computes the notion of information content in the form of the conditional probability of encountering an instance of a child synset, given the occurrence of a parent synset.

$$
\text{wrdd} (c_1, c_2) = 2 \log (p (lso (c_1, c_2))) + \log (p(c_2))
$$

$c_1, c_2$ : synsets

$$
p(x) : \text{probability of encountering } x \text{ in a specific corpus}
\text{lso}(x, y) : \text{lowest super} – \text{ordinate}.
$$

<table>
<thead>
<tr>
<th>Table 2: ERN notation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Notation</strong></td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>MIME</td>
</tr>
<tr>
<td>SENSOR</td>
</tr>
<tr>
<td>$\leftrightarrow$</td>
</tr>
<tr>
<td>$\cup$</td>
</tr>
<tr>
<td>$\cap$</td>
</tr>
<tr>
<td>${}$</td>
</tr>
<tr>
<td>${}$</td>
</tr>
<tr>
<td>${}{}$</td>
</tr>
</tbody>
</table>

### 4.1 Case Studies of The eChronicle Relationship Network

As a demonstration of ERN, several example queries are parsed in the form of the eChronicle relations network presentation. It follows the notation specified in Table 2. The primitive user interface in Figure 8 is provided to help users to compose the query languages semiautomatically.

**Query 1.** Select registered audio records captured on March 1st, 2005.

$$
E \{ tga = (“2005-03-01”) \} \leftrightarrow D \{ \text{MIME (“audio”)}. \}
$$

A event set linked with audio MIME type data is filtered by a temporal relational operator, Equal(=) , in a day resolution (tga) on March 1st, 2005.

**Query 2.** Select detected human faces captured on March

Ted Pedersen developed extensive libraries for the public to compute WordNet similarities: http://search.cpan.org/ tpederse/WordNet-Similarity/
It first finds symbols matched with any of the synonym set (Sy) of each human and face symbol. Then, it finds linked events registered on (=) May 1st, 2005, in a day resolution (tga) by an event detector located at Atlanta within (d) a city distance (ipmg\_city). Each event should be linked with at least one of each two synonym symbol sets. Output result sets are filtered only for events linked with the image MIME type data.

Query 3. Select the meeting report submitted by Jimi Hendrix after the group meeting held on March 1st, 2005.

$$\begin{align*}
S \{ \text{“hendrix”} = \text{Sy ("firstname")}, \\
S \{ \text{“lastname”} = \text{Sy ("lastname")}, \\
S \{ \text{“report”} = \text{Sy ("report")}, \\
E \{ \text{tga} \geq ("2005-03-01")).
\end{align*}$$

$$\begin{align*}
D \{ \text{MIME ("video")}, \\
S \{ \text{“hendrix”} = \text{Sy ("lastname")}, \\
S \{ \text{“report”} = \text{Sy ("report")}, \\
E \{ \text{tga} \geq ("2005-03-01")).
\end{align*}$$

This query requires two data sets. The first is of the meeting video. The second is the time stamp of Jimi Hendrix’s talks. A second set can be acquired by specifying the event linked with his name and talk synonym symbols and timestamp type data. Using these two data sets, users can see the video, selecting only what they want to see. The query result output is composed of multiple different types of data. To play this composite data set in a synchronized way, technologies like Synchronized Multimedia Integration Language (SMIL) may be employed.

Several distinct features make ECR Logistics inherently very different from conventional data storages. First, ECR permits the data reusing as depicted as a linked set in Figure 2. Moreover, our data referencing mechanism in ECR for reusing data actually make a big difference in information chronicling. For instance in personal chronicling, a user may attach new data about herself without distorting existing database architecture but just add an event and data to a specific symbol, in this case her name.

5. SYSTEM IMPLEMENTATION

Different types of multifarious data are generated in our daily lives. The captured data may be related to many user actions and may originate from many applications on a variety of devices. This section introduces our research efforts to handle these data on event capturing, media processing,
and support of user tagging. The prototype of the eChronicle system consisted of three components: (1) event monitoring, (2) tagging, and (3) querying. Brief descriptions of each component will follow. For details, please refer event-based multimedia modeling methods [16, 17, 33], multimedia event tagging systems [19] and personal chronicling systems [20].

### 5.1 Event Monitoring and Logging

Although many personal information management systems only support manual tagging by the user, our chronicling system has the goal of monitoring every communications channel between users and computers. The goal, then, is to log all events in any sort of communication. Our eChronicle system was initially implemented with limited event monitoring tools for each type of data communication in Table 5. The current tools are mainly composed of software sensors to help a user detect events automatically.

To support a variety of other types of devices and communications, reconfigurable event processing architecture that uses XML is supported (please see [20]). By using this model, users can do the easy addition and reconfiguration of event responses and tagging interfaces without programming. For instance, when new data is placed at the clipboard data channel, a user can configure the system to first save the new data to the database and then send a command to the application message channel that will invoke the user interface to place an appropriate tag on this new entry.

### 5.2 User Tagging Supports

While the event monitoring tools capture all user activities on an ongoing basis, a user will have the ability to tag manually any interesting event. User tagging can result in valuable metadata that is not available through any other source. To make it easy for users to generate tags, an innovative tagging interface is developed. It tracks the creation, activation, and destruction of any application software and checks its attributes as a preliminary to attaching a tagging button (At-Button) as shown in Figure 3. The At-Button was first introduced at [20] as the T-Button and enhanced specifically to work with newly developed smart tagging interfaces and to handle multimedia data. Technically, At-button installs system and applications hooks for each Windows. Then it analyzes an application class to further capture application-specific messages. For instance, it can detect Web browser activities like link forwarding, moving back and new Window popping up. It works as multithreaded to work for every top-level Windows.

When the At-button is pressed, it will do two jobs as depicted in Figure 5: common operation and application-specific context analysis. Each extracted data will be placed on the data communication channel, which is like a clipboard, for local media processing or on a network for distributed media analysis. Then it will do the next action as necessary.

### Table 4: Granularity supports for each relation type

<table>
<thead>
<tr>
<th>Category</th>
<th>Granularities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
</tr>
<tr>
<td>Absolute Distance</td>
<td>tga</td>
</tr>
<tr>
<td>(e.g., tga=[“2000-03-09 10:xx:xx”], find events on 2000-03-09 10 A.M.)</td>
<td></td>
</tr>
<tr>
<td>(e.g., tga=[“2000-03-09 10:xx:xx”] or tga=[“2000-03-09”], find events on 2000-03-09 10 A.M.)</td>
<td></td>
</tr>
<tr>
<td>Relative Distance</td>
<td>tgr, seconds, tgr, minutes, tgr, hours, tgr, days, tgr, months, tgr, years</td>
</tr>
<tr>
<td>(e.g., tgr, seconds[10,≤ (“2000-03-09 10:32:15”)], find close events within 10 seconds before happen on 2000-03-09 10 A.M.)</td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td></td>
</tr>
<tr>
<td>IP-to-Map Distance</td>
<td>ipmg, city, ipmg, state, ipmg, country, ipmg, isp</td>
</tr>
<tr>
<td>(e.g., ipmg, city=[“Atlanta”], find events at Atlanta)</td>
<td></td>
</tr>
<tr>
<td>Latitude &amp; Longitude Distance</td>
<td>lld</td>
</tr>
<tr>
<td>(e.g., ipmg, city[0.01,=(84.3888460, 33.7525040)], find events located within 0.01 radius)</td>
<td></td>
</tr>
<tr>
<td>(e.g., ipmg, address[0.01,=(“USA”, “GA”, “ATLANTA”, “30341”)] find events located within 0.01 radius)</td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td></td>
</tr>
<tr>
<td>Relative Distance</td>
<td>wrd</td>
</tr>
<tr>
<td>(e.g., wrd[5, Sy(“meeting”), find events similar with meeting within 5 synset distancet)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 5: eChronicle system specifications.

<table>
<thead>
<tr>
<th>Category</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Channel</td>
<td>console, application message, clipboard, keyboard, mouse</td>
</tr>
<tr>
<td>User Interface</td>
<td>smart tagging console, symbolizer, event-condition-action rule modeler</td>
</tr>
<tr>
<td>User Activity Monitor</td>
<td>smart clipboard watcher, application status watcher, document status watcher, IP location watcher, keyboard/mouse watcher</td>
</tr>
<tr>
<td>Media Analysis</td>
<td>face detector, face recognition, region segmentation, contour processing</td>
</tr>
<tr>
<td>Video</td>
<td>motion detection, shot detection, object tracking</td>
</tr>
<tr>
<td>Audio</td>
<td>end-point detector, speech detector, speech-non-speech-noise classifier, speech recognition</td>
</tr>
<tr>
<td>Document</td>
<td>text/image extractor for various applications, HTML/XML parser, regular expression supports</td>
</tr>
</tbody>
</table>

...
based on ECA models. To support user tagging, Smart Tagging Console (STC) will pop up shown in Figure 4, if necessary, as shown in Figure 6. This will be populated with a significant amount of metadata automatically extracted from the application to interact with the user. STC is semi-transparent not to block other applications from a user view. Also, it only capture mouse messages where its input Window is activated so that it makes minimal interference with existing applications.

By integrating ECA models and STC, users can configure eChronicle to do intelligent event detections and achieve high-level user interactions. Figure 6 shows an example of what occurs when a user presses the At-button while she is reading news on the Web. In this example, the At-button captures a screen and places it on the clipboard. Because the eChronicle system is monitoring the clipboard as a data channel, it checks ECA rules whenever a new image is placed on the clipboard. In this example, it will automatically run the face detector (Of course, many other things can be done at the same time, but this serves for an example). When the eChronicle system finds a face that differs from those registered, it then invokes STC to obtain user tags. In this example, a user specified the name of the captured face as “Osama Bin Laden.” User tags are parsed into words and then matched, using the Symbolizer, with the correct sense in WordNet.

In the same way, smart interactive tagging for multimedia applications now can be achieved as shown in Figure 7. A user captured the online video meeting by pressing the At-Button. Based on the predefined ECA models, our system compared a detected face with face references and then identified the person. One thing that should be noted is that as chronicling progresses, too many faces may be presented for comparison, just as occurs in this example. In our approach, a user can limit face references spatially, temporally or semantically by using an eChronicle relationship network as specified in Section 4. A similar example is at Eq. 9.

5.3 Browse, Search and Retrieval

The early version of the eChronicle system was developed using the Lotus Notes client for personal browsing and searching of a user’s archived events-based data [20]. A new user interface shown in Figure 8 is to support users for composing the eChronicle network presentation, which was introduced in Section 4. The new query interface consists of three parts to specify each symbol, event and data set. When a user selects options and values for searching, this interface will automatically generate queries using the eChronicle relationship network. Also, users can query multiple sets by linking each set with others as shown in the bottom of the picture. For temporal search, users can select one of two options: transaction time and valid time. In ERN presentations, a transaction time represent a time log of an event, while valid time is an associated timestamp data. So if a user selected a transaction option, it finds events occurred within specified time ranges. If she selected a valid option, then it finds events which are associated with timestamp data within specified time ranges. For an output, we currently provide three limited methods to view results: GRID, HTML and XML.
6. EXPERIMENTS

In this section, early partial results using the proposed eChronicle system are shown. We used two sets of data for the experiment: (1) author’s personal archives captured using the eChronicle system, and (2) twelve videos that our group has archived for the past three years as shown in Figure 9. Experiments were done for three cases: (1) complex querying of personal archives, (2) enhancing Automatic Speech Recognition (ASR) using captured domain environment information, and (3) improving face recognition results by limiting search space spatially and temporally.

The first experiment is about how a complex query example can be parsed in the eChronicle system. Because few systems are available for performance comparisons, the query results are compared with those available through Google Desktop (GD). We found that GD implemented very similar techniques with our desktop file monitoring system, which we develop last year [19]. In our case, eC captures file access message and if the type of a file is registered in eC, then it attempts to extract associated texts and metadata automatically. For a comparison of the search performance, we used author’s personal chronicles collected by using eChronicle and by GD. It is assumed that a user needs to prepare a meeting report. So she wants to find all meeting related materials with assumptions that she knows when and where it happened and other related information. The experiment results are tabulated in Table 6. In an actual experiment, the query set “meeting hp” did not work well with GD because many related files do not have names of “hp” but have like “mets_xx.xxx” (cf. METS is the name of the project). So we added one more query for GD, “mets hp,” which contains the exact project name. By using GD, all related files

![Figure 7: Interactive face recognition and tagging example.](image)

![Figure 8: The eChronicle relation network query interface.](image)

![Figure 9: Various types of the organizational life videos used for the eChronicle system performance test.](image)

<table>
<thead>
<tr>
<th>Category</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>eC: Use the centralized eChronicle database server. GD: Collect results from the user’s laptop and two desktops each one at office and one at home.</td>
</tr>
</tbody>
</table>
| Query    | eC: 

\[
S \{[= Sy(“meeting”)]\}, \\
S\{[“HP” = Sy(“company”)]\} \\
\iff E\{tgr\_months[2, (“2005-05”)]\}, \\
E\{ipmg\_city = (“Atlanta”)]\} \\
\iff D\{MIME(“doc”, “ppt” “message”))\} |

| Results  | eC: 4 (“doc”), 4 (“ppt”), 2 (“message”). GD: Sum of three results: 21 (“doc”), 16 (“ppt”), 9 (“message”) and 65 others (files, web histories or chats) not related with what the user wants to find. |
can be found just as the eChronicle system does. However, this can be achieved only after repetitive manual filtering that involves steps such as time reordering, manual selecting and remote file copying to one place. Actually, direct comparison of the results from the two systems is not fair because GD has inherent limitations in semantic and spatial searching. However, it should be noted that the convenience and query precision of the eChronicle system were superior to GD. The query result is shown in Figure 10 using the GRID output format.

As mentioned in Section 2.1, multimedia analysis typically depends heavily on an ongoing environment. For instance, the result of ASR varies depending on the quality of recorded speech. So it is typical to apply preprocessing and noise cancelation on the front end, and sometimes the speech classifier threshold requires adjustment. In a next experiment, we show that it is possible to transfer a proper spatio-temporally related environment configuration by using ERN for ASR.

Each video shown in Figure 9 is processed using the media analysis engine in Table 5. In the case of ASR, CMU Sphinx-4 is employed since it is highly reconfigurable by providing a way to control running processes externally. In this experiment, ASR outputs are filtered only for the highest confidence levels. Also, input running configurations for Sphinx-4 are automatically provided by using ERN in Eq. 8. As a result, the confidence level of ASR outputs applied to twelve meeting videos is improved by 22.3%, with three sets of spatio-temporally and sensor specifically fine-tuned parameters. Because so many variables are related with ASR, performance can be enhanced with more careful tuning.

However, it is clear that experiment shows that users now can process multimedia, depending on its captured environment, just by specifying spatio-temporal environment specifications in a unified way as shown in Eq. 8. As a consequence, any fine-tuned media analysis parameters are now reusable and shareable via eChronicle systems.

\[
S \{ ["Sphinx" = Sy ("software")], \\
S \{ [= Sy ("configuration")], \\
\leftarrow \\
E \{ tga > ("2002-01-01") \}, \\
E \{ ipmg asp = Pa ("Atlanta") \}, \\
E \{ Sensor = Pa ("M900") \} \\
\leftarrow \\
D \{ MIME ("XML") \}.
\]

The last example handles a major problem in face recognition that the quality of an image varies depending on the environment in which it is captured and also on the device with which it is captured. In our approach, a training set for face recognition can be adjusted for spatially or temporally limited sets by using ERN as shown in Eq. 9.

\[
S \{ ["Jimi" = Sy ("Hendrix")], \\
S \{ [= Sy ("face")], \\
\leftarrow \\
E \{ tgr years [1, < NOW ()] \}, \\
E \{ Sensor = Pa ("QuickCam") \} \\
\leftarrow \\
D \{ MIME ("image") \}.
\]

7. CONCLUSIONS

An eChronicle can serve as a semantically accessible annotated multimedia log. Our systems could be as proactive as the application dictates to help users in their activities. In addition, to trace the threads of an individual’s life, a gigantic amalgamation of personal information is necessary to

---

Figure 10: Meeting related information query result sample.

This example limits samples captured within the past year using a specific camera. Face recognition results using the trained face model constructed from the input set in Figure 11 shows that it becomes very robust without becoming confused in distinguishing between Asians of similar appearance. Simple Haar-like features and a cascade of boosted tree classifiers are used for face detection. The output is transferred to the face recognition engine, which uses the embedded Hidden Markov Model (HMM). Both tools are distributed in OpenCV and are modified to get the input data and training set from the eChronicle system.

Registered face images

Filtered face images captured using Logitech cam and within last one year

Figure 11: Face training example using temporally and sensor specifically extracted face sets.

---

see exactly how a relationship or event developed. For this, our system proposed two important research challenges: (1) the event-centric data linkage model, and (2) ERN to represent spatio, temporal or semantic variations. This may provide a foundation to provide further summaries of important events as well as access to events at the required level of granularity.

8. ACKNOWLEDGMENTS

We appreciate the time and resources made available to us for this project by Neerja Raman of Hewlett-Packard Research Laboratory.

9. REFERENCES


