Designing Experiential Environments for Management of Personal Multimedia

Rahul Singh
Department of Computer Science
San Francisco State University, San Francisco, CA 94132
rsingh@cs.sfsu.edu

Rachel Knickmeyer, Punit Gupta, Ramesh Jain
Experiential Systems Group
Georgia Institute of Technology, Atlanta, GA 30332
{rknickm, punit}@cc.gatech.edu, jain@ece.gatech.edu

ABSTRACT

With the increasing ubiquity of sensors and computational resources, it is becoming easier and increasingly common for people to electronically record, photographs, text, audio, and video gathered over their lifetime. Assimilating and taking advantage of such data requires recognition of its multimedia nature, development of data models that can represent semantics across different media, representation of complex relationships in the data (such as spatio-temporal, causal, or evolutionary), and finally, development of paradigms to mediate user-media interactions. There is currently a paucity of theoretical frameworks and implementations that allow management of diverse and rich multimedia data collections in context of the aforementioned requirements. This paper presents our research in designing an experiential Multimedia Electronic Chronicle system that addresses many of these issues in the concrete context of personal multimedia information. Central to our approach is the characterization and organization of media using the concept of an "event" for unified modeling and indexing. The event-based unified multimedia model underlies the experiential user interface, which supports direct interactions with the data within a unified presentation-exploration-query environment. In this environment, explicit facilities to model space and time aid in exploration and querying as well as in representation and reasoning with dynamic relationships in the data. Experimental and comparative studies demonstrate the promise of this research.

Categories and Subject Descriptors
H.2.5 [**Heterogeneous Databases**], H.2.8 [**Database Applications**], H.3.3 [**Information Search and Retrieval**], H.5.1 [**Multimedia Information Systems**], H.5.2 [**User Interfaces**]

General Terms
Management, Design, Human Factors

Keywords
Experiential Computing, Personal Media, Unified Multimedia Modeling, UI Metaphors

1. INTRODUCTION
With advances in processing, storage, and sensor technologies over the last few decades, electronic media of different types is increasingly being used to capture and store disparate activities in different situations. Examples vary from news videos, to proceedings of a conference, to family photographs and videos. As opposed to the (increasing) ease with which such data can be collected, the problems related to its management, presentation, and understanding the information it contains are becoming increasingly complex. In the specific context of personal information, this trend has significantly accelerated in the recent past with the introduction of affordable digital cameras, portable audio recorders, and cellular phones capable of supporting, capturing, and storing information as text (e-mails and instant messages), images, videos, and sound clips. The two fundamental challenges that are encountered in the design of systems for management for such information include:

- Capturing semantics of the information that may be spread across different media.
- Supporting efficient and effective interactions between users and the information.

The first of these challenges is related to the issue of developing multimedia data models that can capture the semantic correlations between various media. It may be noted, that a wide variety of data models are available for representing multimedia data such as images, video, audio, or text. Also, systems such as [5] have considered the problem of personal information management in settings that simultaneously involve different types of media. However, the distinction of the formulation we consider from such approaches lies in the observation that different types of media (when used in a coordinated manner) describe different aspects of the same informational entity. Therefore, in order to adequately account for the semantics it is essential to model the different media in a unified manner.

In order to comprehend the second challenge, we begin by noting that unlike alphanumeric data, the semantics associated with complex media (such as images or video) is non-unique and emergent; media is endowed with meaning by placing it in context of other similar media and through user interactions [13]. Therefore, facilitating exploratory, user-centric capabilities rather than pure syntactic query-retrieval is one of the important requirements of interfaces that support user-media interactions. The user centricity is important also because media rich data are (necessarily) rich in sensorial terms and need to be presented so that users can directly experience them. The transition from settings dominated by alphanumeric data and precise query-retrieval to those involving rich heterogeneous media and an emphasis on exploration, rather than syntactic querying has been the focus of study in [7], where the term "experiential environment" was used to denote systems and interfaces that take

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'04, October 10–16, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-893-8/04/0010...$5.00.

advantage of the human-machine synergy and allow users to use their senses and directly interact with the data. Such systems were characterized to: (1) be direct, in that they do not use complex metaphors or commands, (2) support same query and presentation spaces, (3) maintain user state and context, (4) present information independent of (but not excluding) media type and data sources, and (5) promote perceptual analysis and exploration.

In this paper, we present the results of our ongoing investigations in developing an experiential system for managing and interacting with personal information. Our research prototype, called *eVITAe* (for electronic vitae) is meant to chronicle the life of its users through various media collected by them over their lifetime. In this, it shares the goals of [5]. Our distinction from [5] and other prior research lies in the emphasis of our formulation on the two fundamental challenges enumerated earlier.

A key issue in designing experiential systems is developing a generic framework for managing different types of media in a unified manner. This is essential not only to present information independent of media types, but also to ensure that semantically correlated media does not get separated due to type-specific constraints or management strategies. This issue relates the two key challenges that lie at the core of our formulation.

In our research, we have used an event-based approach to model varied media in a general, semantically unified manner. An event in our approach is an observed physical reality that functions as the fundamental information generator. Using this as a conceptual model, a formal data model is developed (see [14] for details) that facilitates unified modeling of multimedia and helps in representing various dynamic relationships present in the data. In the ensuing sections, we present details on how we build on the aforementioned event model to develop an experiential system for personal multimedia information management.

2. SYSTEM DESCRIPTION

The primary components that are essential for supporting the aforementioned characteristics of an experiential system include an *event-entry subsystem, event-indexed media storage,* and an *event-based media query and exploration environment*. In this paper we briefly outline the first two components and present a detailed description of the third one since it constitutes the experiential environment, which is the focus of this paper.

The first of the above components is an *event-entry subsystem*, the role of which is to populate the event-model, based on the available media and associated meta-information. The primary ways in which this goal can be achieved include [15]: (1) manual entry, (2) automated entry by reading pertinent information from data formats or data capture devices such as image headers and GPS enabled cameras [12], (3) entry by reading meta information from a digital calendar or from surrounding information, and (5) by media analysis and association [10]. Currently, we have implemented a manual entry interface and have developed functionality that supports automated reading of information from files and file headers. The second important functional component in the system is *event-indexed media storage*. This is essentially a physical implementation of the event-based data model. In our implementation, each specific media (such as photographs, video, audio, text, or e-mails) are stored in specialized tables in a relational implementation. Another table is created to store events, along with related information such as different attributes such as event name, location, start and end time, participants, and link(s) to the media. Relationships between various events are also captured here. Time is modeled by following the standard solar calendar and location (space) is modeled based on latitude-longitude. Lookup tables are created to map names of geographical importance (for example, cities) to their latitude-longitude coordinates.

Figure 1: Snapshot of the system. Spatial characteristics of the data are displayed by using a map (in the top right box). The timeline and the *EventCanvas* are on the top left. The temporal and multimedia characteristics of the events are displayed here. Media constituting the support for a given event are shown at the lower left in the *EventDetailCanvas*. Specific event attributes, such as participants, can be displayed in the *EventAttributCanvas* at the bottom right. Note that since time and space are considered as the primary attributes in the event-model, they are depicted and used as the two most significant dimensions for data exploration.

These tables help users to interact with spatial information using an intuitive (nominal) nomenclature rather than having to work with latitude-longitude information. Queries on space and/or time are addressed in terms of containment within an interval (for time) or a rectangular region (for space). Both the temporal and spatial model support *semantic zooming*, allowing the users to examine information at varying resolutions of time and space.

A snapshot of the *event-based media query and exploration environment* is shown and explained in Figure 1. Options for zooming, filtering, extraction, viewing relations, and details-on-demand have been provided in order to help users "explore" information space. The system employs direct manipulation techniques [6] to facilitate natural user-system interactions. In this paradigm, users can directly perform different kinds of operations on items (events) of interest. Furthermore, combining the query and presentation space alleviates the cognitive load from the user's part unlike traditional query environments.

The visual representation of events in this system is done through the use of a hierarchical structure called *event plane* consisting of none, one, or more rectangles embedded within each other stretching across time as shown in Figure 2. Each such rectangle

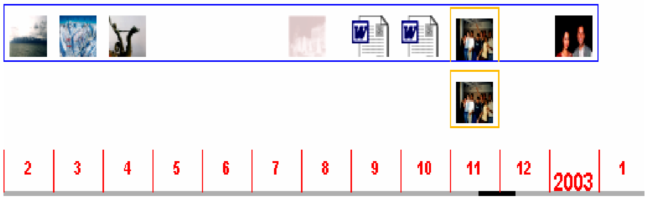

Figure 2: Figure showing an example of an event plane and event hierarchy. The event represented by the top rectangle contains another event. The sub-event is displayed as a nested event plane inside the parent event. Various media constituting the support of each event are also shown through thumbnails or icons.

represents an event and contains within itself the thumbnails or icons of its media support. The event planes serve multiple purposes in addition to visually representing events and event hierarchies: First, they provide a temporal characterization of events. Dynamic relationships between events such as point-point relations (*before, after, simultaneous*), point-interval relations (*before, during, after*), and interval-interval relations (*before, equal, meet, overlap, during, start, finish*, and the corresponding symmetric relationships) can be visually represented and discerned. Such relations or their derivatives have long been used for modeling dynamic characteristics of data [16] and in multimedia data models [9], including the event-based formalism used by us [14]. Additionally, event planes also capture the inherently sparse (in terms of media support) nature of events, with the media thumbnails or icons denoting the temporal durations for which media is available. This is essential, for example, to correctly represent in a semantically unified fashion, all media from a "Trip to Europe" that may have lasted over multiple consecutive days, while the associated media related to it (such as photographs) have discrete timestamp information associated with them and can not be semantically interpolated over the entire duration of the event.

The timeline based presentation of events supports zooming into a particular time interval to find more details and zooming out to see the overall distribution of events on the temporal axis. Support for local zooming (zooming within a specific interval) allows focusing on a specific period. Further, the semantic fisheye-view technique [8] is used to highlight the objects of current focus in the timeline while the user moves the slider across the timeline and zooms into a particular time interval. User can also select multiple intervals in the timeline, thereby creating multiple foci of the fish eye view.

Support for spatial (location-based) interaction with the data is provided through a location-aware map implemented using an open source JavaBeans package called OpenMap, developed by BBN technologies [2]. OpenMap is essentially a collection of Java Swing components that understand geographic coordinates. The spatial display supports options to zoom down to a particular location and options to zoom out. Furthermore, panning of the entire space is also supported. Selection of a region of interest (for querying or exploration of all events that have occurred there), is done by drawing a rectangle. Locations whose latitude-longitude fall inside the query rectangle are then identified to obtain the corresponding events.

The design of the interface has been done to support emergent and exploratory user interactions with heterogeneous multimedia data. The use of the event model, allows information to be presented to users in manners that are independent of media and data sources. In the interface various views of the data (such as those across time, space, or other attributes) are tightly linked to each other, so that interactions in terms of any one of them are instantaneously reflected in all the views (for example, selecting an event by time leads it to be highlighted in the spatial view). This is essential for maintaining context and helping in formulation of insights.

3. EXPERIMENTAL RESULTS

To evaluate the *eVITAe* system, two experiments were performed. These explored the role of space, and the quantitative retrieval capabilities of the event model. In general, the role of space in applications such as these has not been sufficiently explored, while the role of time in organization of media has been well studied [5, 11]. We sought to show that spatial information is of equal, if not greater importance, than time for retrieval tasks. The quantitative comparison shows the power of an event based model as opposed to a variety of well studied media management approaches such as *PhotoMesa* [1], *FileHand Search* [3], *File Viewer*, [4], and *Windows Explorer*.

A total of 9 participants completed a set of retrieval tasks in each of the above systems as well as on a version of *eVITAe* with the temporal interface disabled, and again with the spatial interface disabled. Conditions included one task involving image-only (single media) retrieval and two tasks requiring the retrieval of multimedia information.

3.1 Results

The spatial only condition resulted in faster data retrieval as shown in Figure 3. When combined with temporal querying the speed of interaction increases even further.

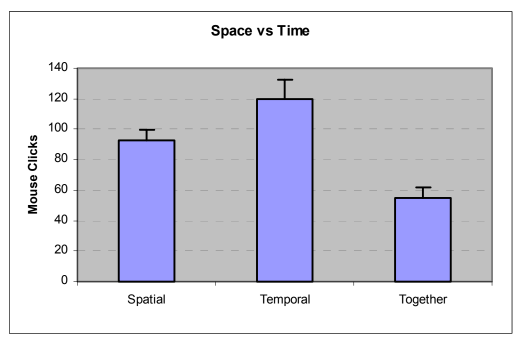

Figure 3: Spatial vs. Temporal Queries

Many users noted that they preferred using the map as it was easier to recall where things occurred then the specific or relative time. Most users in the combined task would start by narrowing results down by space, and then used the timeline to browse the remaining events. This approach shows clear advantages over both time and space alone.

In the second experiment (see Figure 4 and Figure 5), we tracked mouse clicks and keystrokes in two general conditions: syntactic data retrieval for photographs and retrieval of multimedia data. *eVITAe* performed better than Windows on both task types, and on par with Photomesa for the image retrieval task. When *eVITAe* was tested against windows as well as the two other systems for multimedia retrieval tasks, it preformed noticeably better across the two tasks. Photomesa was not included in this analysis as it can only support single media (image) retrieval.

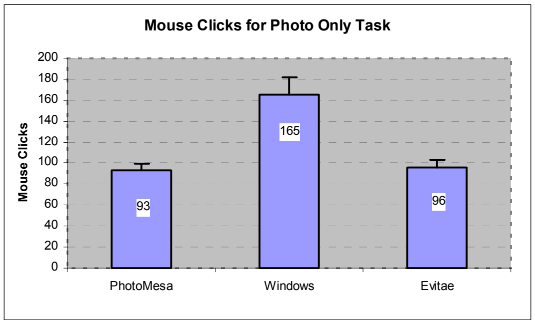

Figure 4: Complexity of Image Retrieval Task

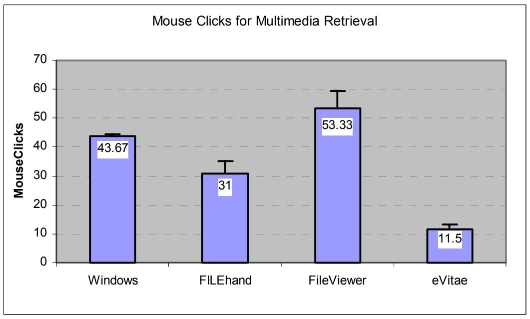

Figure 5: Complexity of Multimedia Retrieval Task

To illustrate this further, one of the tasks involved scanning the dataset for a specific publication and its associated images. Participants typically found the images quite quickly, picking them out by visual cues alone, but finding the publication proved more difficult. Many participants clicked and opened documents before determining that they were not the correct ones. In *eVITAe* on the other hand, due to the use of the unified multimedia model, images were grouped directly with their associated data, and thus participants quickly narrowed down the choices, and were able to determine the correct publication in fewer clicks than they could with the other programs. Furthermore, the ability to locate the document by space and time made the task much easier, as participants seemed to recall more readily when and where the publication was from, but not necessarily its exact title or even the correct file type.

4. CONCLUSIONS

In this paper we have outlined our research on designing experiential environments for user multimedia interactions. Conceptually, such systems support human-machine synergy by presenting a holistic picture that is inclusive but independent of media and data sources. The system described here achieves these goals by using an event-based unified multimedia data model. A unified presentation-query-exploration interface facilitates direct interactions between the users and the data. Further, mechanisms such as spatio-temporal zooming, semantic fisheye-views, and tight coupling between different views of the data are supported to maintain user state and context. The results presented in this paper are specifically obtained in the context of the problem of personal multimedia management. However, they make no fundamentally constraining assumptions and therefore are expected be extensible to a wide class of problems where user-multimedia interactions play a key role.

5. REFERENCES

[1] Bederson, B. “Quantum Treemaps and Bubblemaps for a Zoomable Image Browser”, in ACM Conference on User Interface and Software Technology, 2001

[2] http://openmap.bbn.com

[3] http://www.filehand.com,

[4] http://www.accessoryware.com/fileview.htm

[5] Gemmell, J., Bell, G., Lueder, R., Drucker, S. and Wong, C. “MyLifeBits: fulfilling the Memex vision”. ACM Multimedia, pp. 235-238, ACM, 2002.

[6] Hutchins E.L., J. D. Hollan, D. A. Norman. "Direct Manipulation Interfaces," User Centered System Design. Lawrence Erlbaum Associates. 1986

[7] Jain, R. “Experiential Computing”, Communications of the ACM, Vol. 46, No. 7, July 2003

[8] Janecek, P. and Pu, P. A Framework for Designing Fisheye Views to Support Multiple Semantic Contexts. in International Conference on Advanced Visual Interfaces (AVI '02), ACM Press, 2002

[9] Little, T. and A. Ghafoor, “Interval-Based Conceptual Models for Time-Dependent Multimedia Data”, IEEE Trans. On Knowledge and Data Engineering, Vol. 4, No. 5, 1993

[10] Loui, A. and A. Savakis, “Automated Event Clustering and Quality Screening of Consumer Pictures for Digital Albuming”, IEEE Trans. On Multimedia, Vol. 5, No. 3, 2000

[11] Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B. “LifeLines: visualizing personal histories”, in Proceedings of CHI'96, pp. 221-227, Vancouver, BC, Canada, April 14-18, 1996

[12] http://www.dcviews.com/press/Ricoh_RDC-i700G.htm.

[13] Santini S., A. Gupta, and R. Jain, “Emergent Semantics Through Interaction in Image Databases”, IEEE Trans. On Knowledge and Data Engineering, Vol. 13, No. 3, 2001

[14] Singh R., Li Z., Kim P., Pack D, and Jain R, “Event-Based Modeling and Processing of Digital Media”, First SIGMOD Workshop on Computer Vision Meets Databases, 2004

[15] Toyama K., Logan R., Roseway A., and Anandan P., “Geographic Location Tags on Digital Images”, ACM Multimedia, pp. 156-166, 2003

[16] Widerhold G., S. Jajodia, and W. Litwin, “Dealing with Granularity of Time in Temporal Databases”, CAiSE91, pp. 124-140, 1991.