DESIGNING INTERACTIONS IN EVENT-BASED UNIFIED MANAGEMENT OF PERSONAL MULTIMEDIA INFORMATION

Juan Camilo Pinzón[1], Rahul Singh[1], Wolfgang Taube[2], Jensen Galan[1]

[1]Department of Computer Science, San Francisco State University, San Francisco, CA 94132
[2]Department of Digital Media, Fachhochschule Furtwangen, Furtwangen, Germany, 78120
juanpin@sfsu.edu, rsingh@cs.sfsu.edu, taube@fh-furtwangen.de, jrgalan@sfsu.edu

ABSTRACT

Advancements in sensor technologies have made it easier and increasingly common to capture information using multiple media. This is especially true for personal multimedia information. Effective assimilation of such information requires recognizing the semantic correlations between media and the ability to model and interact with them in a unified manner. This paper presents our research in designing capabilities to support user-data interactions in context of the aforementioned issues. Central to our approach is characterization and modeling of media using the notion of an "event". Building on this idea, we propose the design of operations as well as visualizations that not only allow event generation and manipulation, but also the ability to interact with various semantically important characteristics of the underlying information. Experimental and comparative evaluations demonstrate the efficacy and promise of the approach.

1. INTRODUCTION

As computational resources and sensor technologies become ubiquitous, it is becoming increasingly common for people to electronically record, photographs, text, audio, and video chronicling their day-to-day activities. The availability of sophisticated multimedia capture capabilities provided through affordable devices such as digital cameras and cellular phones capable of capturing and storing images, video, audio, and text, underline the continuation of this trend.

The proliferation of such multi-media data is however, a critical challenge to the state of the art in multimedia data management. To understand the essence of the issue, we note that significant research has occurred in the management of media such as images, video, and audio. However, such research has typically focused on the development of powerful features to describe the corresponding media and the use of similarity functions to answer queries based on these features [7]. This approach simplifies the multimedia management problem, because a database is assumed to contain only a specific type of media [3].

Furthermore, the approach of storing, querying, and managing different types of media separately creates media specific silos and fragments the semantics of the information across them. Two key challenges therefore arise: (1) *The unified modeling problem*: How to design a generic multimedia management framework, that would allow interactions across multiple media in semantically meaningful manners? (2) *The interaction design problem*: What kind of display and interactions, including query-retrieval capabilities, should such a framework support?

Recent research by Singh et al. [8][9] propose the notion of an "event" as the cornerstone of one possible approach to unified multimedia modeling. In this approach, an event signifies an occurrence that is captured through, possibly multiple types of, sensors, each of which results in the description of this occurrence through the corresponding media. For instance, a vocal performance may be captured using two types of sensors: microphones and video-cameras, yielding its description using two types of media: audio and video. The event (i.e. the vocal performance), then acts as a semantically meaningful unifying point for the two types of media. Expanding on this idea, an event-based data model is proposed, where events are parameterized by attributes such as time, location, and application specific attributes such as participants. Each event also subsumes the different media that capture it. Finally, various types of relationships, for example spatial and temporal are defined among events. In an implementation, the event-based layer is envisaged to serve as a bridge across media specific management layers. For further details we refer the reader to [8][9]. It may be noted, that the intuitive notion of an event is a natural one for organization of information. For instance, in the extensive studies conducted by Rodden and colleagues [11], participants identified "events" as the highest rated feature for organizing information. The key contribution of our prior works [8][9], was to formalize this notion and apply it for unified multimedia modeling. As identified above, a pre-eminent challenge in this context is that of designing paradigms for user-data interactions. In this paper, we address this specific problem by focusing on the context of personal multimedia management. The contributions of this paper include:
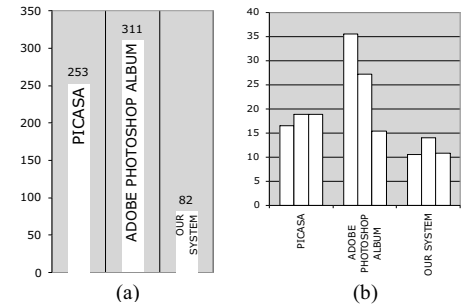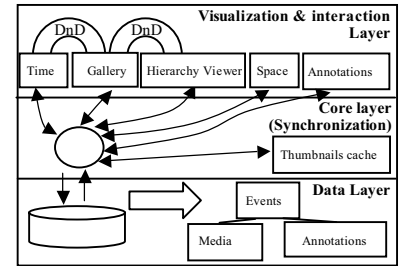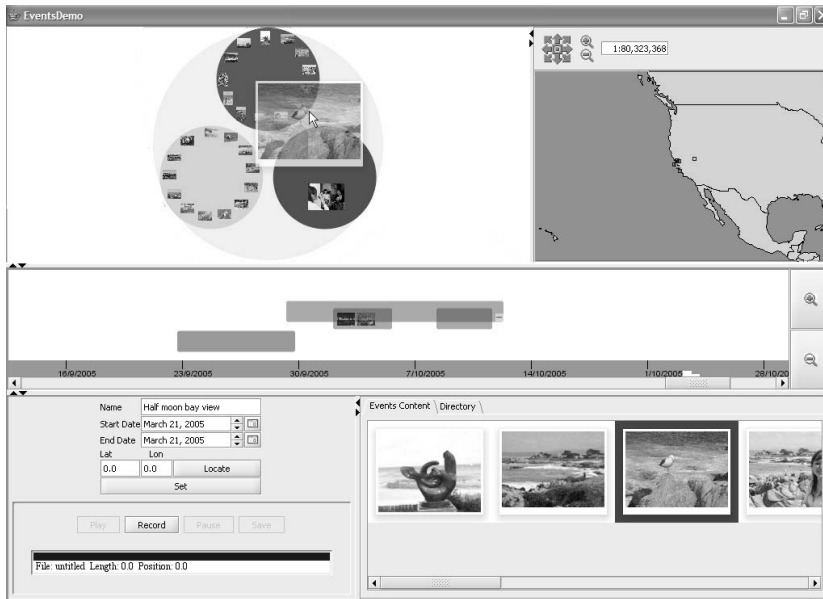
Figure 1. Screenshot of media being transfer from one event to other using DnD. Zoomed location of the event is shown in the map as well as its hierarchy extended in the timeline, also the correspondent event is displayed in the hierarchy viewer. (left). System architecture diagram (top right). Average number of clicks necessary to organize 50 media items (a). Average number of clicks necessary to find the contents of 3 different queries (b).

- A multi-stage approach to event identification, that utilizes, as the first step, spatial and/or temporal characteristics (such as GPS coordinates and timestamps), of media files to cluster them into groups that correspond to the intuitive notion of events. Later stages employ simple yet highly efficient user interactions to possibly modify or semantically enhance the event descriptions.

- Development of highly intuitive graphical operations that can be used to perform event-level manipulations, such as merging events, altering events, and creating new events.

- Development of visualizations and interactions that allow reasoning about the semantically hierarchical nature of events. Creation of capabilities for performing drill-down and roll-up operations on event hierarchies and visualizing their spatial and temporal characteristics.

Further, all the aforementioned features are designed using the idea of experiential environments [4] and reflective interfaces. Essentially, this allows our system to (1) avoid complex user-data interaction metaphors, (2) support same query and presentation space, (3) present characteristics of the information that are independent of (but do not exclude) media types, (4) maintain user and data context, and (5) aid in perceptual analysis, exploration, and information assimilation.

It should be noted that the problem of personal information management is an active research topic. See for instance [1][2][5][6][10][11] and references therein. The fundamental distinction of our approach from such efforts lies in our emphasis on unified multimedia modeling and designing interactions in such settings.

2. SYSTEM DESCRIPTION

The system is composed of three layers (Figure 1): a *Visualization and Interaction Layer* to support display and interaction with various semantically relevant event characteristics, a C*ore Layer* that contains the event-level operations and synchronization of the system, and a *Data Layer* in charge of event and media storage and retrieval. This research focuses on the first two layers, which we describe in detail below.

The *Visualization and Interaction layer* includes UI-components to visualize spatial, temporal, and hierarchical characteristics of events as well as the media and annotations associated with them. Within this layer, interactions between components is reflective and supported through Drag and Drop (DnD) handlers that serve as a way to communicate between different interface elements. These components together constitute an event-based organizer for managing semantically related heterogeneous media. The *core layer* is in charge of synchronizing and providing all the programming interfaces needed by the UI components to support interactions, issue queries and perform automatic operations like spatial-temporal clustering of media to form events.

Seven modules comprise the Visualization and Interaction layer: (1) an interactive map, (2) a universal media player, (3) event detail display, (4) an audio annotation tool, (5) an event hierarchy viewer, (6) an interactive timeline, and (7) a media gallery. In the following, we briefly outline the modules (1) – (4) and focus on the design of the event hierarchy display, the event timeline, and the media gallery.

Operation	Input/Output	GUI-Action
Create Empty Event	$\{\} \rightarrow$ event	Draw a rectangle to indicate approximate start and end dates in the timeline or right click on the hierarchy viewer to create a new event, and then enter details.
Create Event(manual)	$\{MMO\} \rightarrow$ event	Select multiple media items from the gallery, and then right click to create a new event with the media selected.
Create Event(s)(using a cluster algorithm)	$\{MMO\} \rightarrow \{event\}$	Select multiple media items from the gallery, and then right click to select the clustering options. Multiple events will be created automatically based on the events dates.
Insert Media	$\{MMO\}$, event	Select one or more media items inserted into existing event then DnD into the desired event on the hierarchy viewer or timeline.
Remove Media	$\{MMO\}$	Select one or multiple media items, and then right click to remove to an archive.
Remove Event	$\{event\}$	If event is empty then delete it, if an event is a super-event then detach it from sub-events, if an event contains media items they are removed to an archive.
Attach Event (to super-event)	$\{event\}$, event	Select an event to be attached then DnD to the desired event. If the event is not within the time range of the super-event, it would not be attached.
Detach (from super-event)	$\{event\}$, super-event	Select an event to detach then DnD it outside the super-event.

Table 1. List of operations. Notation: $\{\}$=empty set, $\{event\}$ = one or multiple events, $\{MMO\}$ = one or multiple media objects, $\rightarrow$ = transforms into

The interactive *map* shows the event location and enables users to retrieve events by selecting either a region in the map or by clicking on specific locations. Additionally, it supports standard operations like zoom in/out and panning.

The *universal media player* provides a single interface/player for viewing multiple media types, including video, audio, images, files in PDF format, and text. The *event detail display* presents information pertinent to an event, and includes information such as the event location, start and end dates, event name, and event description (textual or audio). Finally the *audio annotation tool* loads the corresponding annotation and allows the users to record voice annotations.

A critical attribute of event-based information organization is its hierarchical nature, where semantically simpler events aggregate in time and space to form more complex event. For instance, a "birthday party" may be constituted of simpler events such as "cutting the cake", "meeting friends", or "having dinner". Each of these events may themselves comprise of simpler sub-events. Such event hierarchies provide a transition from the lower (simpler) levels which are directly based on the media (data) to higher level, based on conceptual notions (which in our problem domain, are typically user defined). The ability to efficiently represent, display, and interact with such hierarchical information is integral to understanding the underlying information. In our system, the *Event-Hierarchy Viewer* is used to display and interact with hierarchical event information. Our design extends the idea in [13], to present events as circles containing either media thumbnails, or other circles denoting sub-events (see Figure 1). The presence of media thumbnails provides a visual notion of the event contents. Interactions supported through this display allow users to "drill-down" or "roll-up" into the hierarchy. This interface also reflects any change on the event structures and enables basic event structure operations like creation, removal and alterations on the hierarchy structure.

In the *timeline*, events are represented by rectangular panes, which may contain media thumbnails. To address the issue related to time granularity, the timeline uses a zoom in/zoom out abstraction that interactively allows a user to view information at the level of days, months, or years. To display hierarchies, overlapping events are displayed in different levels and events that belong to the same hierarchy are displayed in a layered manner. Given a specific hierarchy of events, the default view displays the top level events. The presence of sub-events is indicated by a small (+) icon that acts as a toggle. By clicking it, users can expands and contracts the entire event hierarchy in the timeline.

The *gallery* component supports preview of the contents within an event and to serve as a bridge between the file system and the event database. This component is divided in two. The first section displays media for any selected event. The second section allows users to select a directory in the file system, obtain a preview of the media contained in it and import the media files into the system. During the import, temporal clustering of the media files into events is accomplished using a sliding-window approach described below.

In the *Core Layer*, temporal clustering of media files using their time of capture, serves as the initial point to organize media by events. The idea of clustering media by time is based on the "bursty" nature of capturing events through media. The media is first sorted chronologically. Next, the temporal difference (Δt_i) between the timestamps is calculated, along with the average temporal difference (Δt_{avg}) for adjacent media (m_i, m_{i+1}) within a sliding window. In our experiments, a window size of 2 hours provided highly accurate results. In this approach, a new event is created if $\Delta t_i(m_i, m_{i+1}) > \Delta t_{avg}$ and the media timestamps fall within the sliding window. Otherwise the media is added to the current event.

It should be noted that, while highly effective in practice, such clustering may not always provide perfectly accurate results. Furthermore, user input is invariably required to represent implicit semantics of a situation in order to come up with event hierarchies that are rich in terms of user insights and experiences. A series of operationally simple, yet semantically powerful event-level operations are therefore defined to help users manipulate events. These operations are all implemented as highly intuitive drag and

drop operations. Table 1 describes these operations and the corresponding GUI actions.

3. EXPERIMENTS

We present results from three experiments to demonstrate the efficacy of our design. In the first experiment, the effectiveness of the time-based clustering approach to define events was measured. Three different media sets (containing 400, 153, 171 media items respectively), that belonged to three different users were involved. Precision and recall values were used to measure the effectiveness of the clustering algorithm in correctly assigning media to events (as determined by each of the users). The results from this experiment are shown in Table 2.

	Trial 1	Trial 2	Trial 3	Avg.
Precision	0,90	0.84	0.81	0.85
Recall	0.94	0.91	0.89	0.91

Table 2. Precision and recall values for experiment 1

The aim of the second and third experiments was to determine the usefulness of the proposed interactions for users in managing and assimilating information. As part of these experiments, we compared our system with two commercial products that share similar objectives, namely Picasa [14] and Adobe Photoshop Album [12]. Fifteen (15) participants were involved in these studies and were first introduced to each of the systems. For the second experiment, the users were provided with 124 media items (including text, videos, and photographs). Each user then organized this media with each of the three systems, in semantically meaningful manners, as determined by each one of them individually. The complexity of the effort, with each system, was measured using the number of clicks. This data is presented in Figure 1 and the proposed system did significantly better than Picasa and Adobe Photoshop Album. This can be attributed to a combination of automated clustering, event manipulation operations, and the different semantic views supported in our system. In the third experiment, the ability of users to execute non-trivial retrievals across the three systems was investigated. A predefined set of 753 media items were pre-loaded in each of the three systems. The users were then asked to find information corresponding to the following three queries: (1) "Find media related to two activities that occurred after a ski trip in 2005 and before the summer season", (2): "Find photographs from two different dinners in 2005/06 holiday season", and (3) "Search for media from two trips during the second semester of 2005". Each user used the applications and executed the queries in random order.

Interestingly, different users used the applications in different ways; some used time as their main reference, and others used the groupings/events to search for the data. Picasa and our system out performed Adobe Photoshop Album because of two reasons: all the items on the interface were chronologically ordered and changes or selections in

any component were reflected in the rest of the interface. However, the simple yet semantically powerful display capabilities of our system, led to the most efficient retrieval, the results of which are presented in Figure 1.

4. CONCLUSIONS

In this paper we have considered the problem of designing user-media interaction in event-based unified modeling of personal multimedia. Such a unified approach avoids fragmentation of semantics across multiple media and ultimately leads to better information assimilation. We propose operations that allow event creation and manipulation. Further, a collection of specialized interfaces allow users to visualize and interact with various semantically relevant event characteristics. Comparative experiments and user studies demonstrate the efficacy of the approach and its potential for a new class of multimedia information management systems.

5. REFERENCES

[1] Cooper, Matthew, Foote, Jonathan, Girgensohn, Andreas, Wilcox, Lynn. "Temporal Event Clustering for Digital Photo Collections.", ACM Transactions on Mult. Comp., Comm. and Apps. 2005: 269-288.

[2] Communications of the ACM, Special Issue on Personal Information Management, Vol. 49 No.1, Jan 2006

[3] Dionisio, J. D. N., Cardenas A., "A Unified Data Model for Representing Multimedia, Timeline, and Simulation Data", IEEE Trans. On Knowledge and Data Engineering, Vol. 10, No. 5, 1998

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

[5] Naaman M., Y.Jiun Song, Paepcke A., García-Molina H. "Automatic Organization for Digital Photographs with Geographic Coordinates". *In proceedings,* (JCDL 04), June 2004.

[6] Platt J., Czerwinski M., Field B., "*PhotoTOC: Automatic Clustering for Browsing Personal Photographs"*, Proc. 4th IEEE Pacific Rim Conference on Multimedia, 2003.

[7] Santini S, Gupta A., "Principles of Schema Design for Multimedia Databases", IEEE Trans. on Multi., Vol. 4, No. 2, 2002

[8] Singh R., Li Z., Kim P., Pack D., Jain R., "Event-Based Modeling and Processing of Digital Media", CVDB 2004: pp.19-26

[9] Singh R., Knickmeyer R., Gupta P., Jain R., "Designing experiential environments for management of personal multimedia". Proc. ACM MM. October 2004.

[10] Ringel M., Cutrell E., Dumais S. T., Horvitz E. "Milestones in time: The value of landmarks in retrieving information from personal stores". In *Proceedings of Interact 2003*.

[11] Rodden, K. and Wood, K. R. "How Do People Manage Their Digital Photographs", Proc. CHI 2003.

[12] Adobe Photoshop Album 3.0 http://www.adobe.com

[13] Grokker, http://www.grokker.com

[14] Picasa, http://www.picasa.com