Proceedings of the IASTED International Conference
HUMAN-COMPUTER INTERACTION
November 14-16, 2005, Phoenix, AZ, USA

DESIGNING AN EXPERIENTIAL ANNOTATION SYSTEM FOR PERSONAL MULTIMEDIA INFORMATION MANAGEMENT

Juan Camilo Pinzon
Department of Computer Science
San Francisco State University, San Francisco, CA 94132
juanpin@sfsu.edu

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

ABSTRACT

The challenges of Personal Information Management (PIM) are proliferating with the ever-increasing availability of systems for digital multimedia information capture. Assimilating such information requires supporting appropriate user-media interaction paradigms as well as developing data models that can represent the true multi-media nature of the information. This paper examines the problem of facilitating user annotations of media in light of the aforementioned context. Annotations, while universally acknowledged to greatly enhance the information content and interaction experience, are often tedious and complex for users to enter. In real-world usage, these problems deter users from annotating media data, which in turn greatly reduces the efficacy of media management systems. In this paper we show how principles from experiential computing and unified multimedia modeling can be used to address many of the challenges that contribute to the complexity of annotating media. The resulting experiential annotation system supports direct user-data interactions using audio, graphical, and text modalities. An integrated annotation-presentation-interaction allows the use of these modalities, potentially in conjunction, to reduce the tedium associated with the annotation process. Additionally, an event-based unified data model and spatio-temporal representation of information are used to retain data context and thus further ameliorate the "annotation bottleneck". User studies are used to determine the efficacy of the proposed approach.

KEY WORDS

Media Annotation, Experiential Computing, Personal Information Management, Unified Multimedia Modeling, UI Metaphors

1. INTRODUCTION

Digitally capturing and storing data and experiences from one's life is becoming increasingly common owing to the availability of powerful and inexpensive media capture devices ranging from digital cameras and camcorders to cell phones with built-in cameras and PDAs. The ease of information capture using such devices contrasts sharply with the increasingly significant challenges related to the management of the captured multi-media information.

The process of media management often employs the step of media annotation to aid in populating the data model, query-retrieval, and media presentation/browsing. In the context of PIM for instance, the media annotation step typically requires users to add information about specific media data such as a photograph. Once entered, the annotation can be analyzed, propagated, or used as metadata. In spite of these obvious advantages, studies on usage patterns involving real world PIM systems show that users often avoid annotating their data and typically consider the process complex and tedious [1]. It is our belief that a variety of factors associated with the assumptions, formulation, and support of the annotation process contribute significantly to its *apparent* complexity. These factors include:

- *Choice of annotation modality*: Often, the modality used for annotations is text. Textual data is easy to parse, analyze, and store. However, making large number of text entries is a complex, tedious, and error-prone process. Furthermore, by abstracting away from our natural senses, textual annotations impoverish the user experience.
- *Dependence on media type*: As a rule, most PIM systems treat information in media specific manners. Annotation too is made to be media specific. This prevents the annotation process from describing semantics spread across media types and reduces its expressive value.
- *Minimal support for data and user context*: In addition to media specificity, many PIM systems employ data model that do not provide support for temporal, spatial,

or more complex relationships (such as cause-effect) such relations often underlie data captured from the real world. Consequently, such contextual information about the data is fundamentally unavailable during annotation. To compound the issue, most annotation systems, exceptions such as [3] aside, do not maintain user state or context. This increases the cognitive load on the user during the annotation process and adds to its complexity.

- *Annotation as an isolated preprocessing step*: Many systems view annotation as part of the data entry step that is conducted in segregation from any information querying, browsing, or visualization capabilities that may be supported. Such a view precludes the use of potentially powerful user-data interaction functionality during the annotation process.

- *Skill level of the user group:* The success of a PIM system is predicated on its acceptance by the general populace. Consequently, such systems need to support user-data interaction capabilities that are simultaneously powerful as well as simple and natural to use.

Recent research in Multimedia has espoused design of systems and interfaces that (1) are 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. Termed "experiential systems" [4], the main motivation behind this approach is to allow users to use their senses and directly interact with the data. Experiential systems have been applied to different problem domains including information exploration [2], and PIM (excluding the issue of media annotation) [5].

The key characteristics of experiential systems exhibit a striking fit to the requirements of media annotation identified by us. For instance, the directness in experiential systems corresponds to the need for using natural and direct annotation modalities during annotation. Similarly the property of experiential systems to maintain user state and context can aid the annotation process by providing contextual cues. Other property-requirement correspondences can also be analogously established. This convergence between the requirements for reducing the annotation complexity and characteristics of experiential systems forms the basis of our design approach. We perceive the primary contributions of our research to be the following:

- From a fundamental perspective, we consider annotation to be a result emerging from experientially supported user-media interaction rather than a one-directional information flow from the user to the system. Furthermore, the annotation process can potentially encompass distinct (but semantically correlated) media. To achieve this goal, we have built on our prior research in event-based unified multimedia modeling in the following; we briefly describe our prior research and refer the reader to [6] for details. An event in our approach is an observed physical reality having spatio-temporal character that function as the fundamental information generator. It may be observed through different sensors and information about it stored using different types of media. Such potentially heterogeneous media are said to support a given event. Using the notion of an event as the focal point, a data model can then be used to unify different media types. By considering the annotation problem in presence of such a unified data model, we are able to annotate both specific media data as well as information organization abstractions that span different media types.

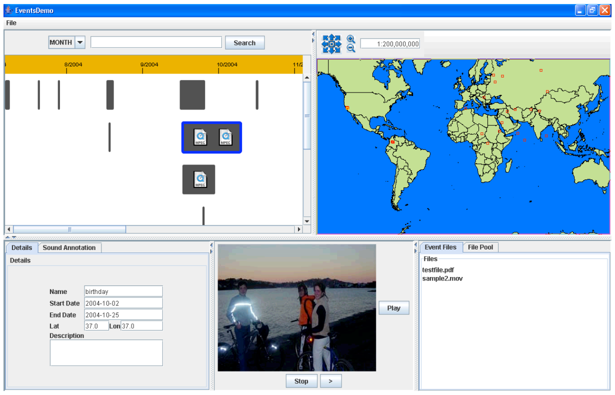
- From an operational perspective, we have developed an integrated annotation-presentation-interaction interface, which provides support for data and user context. Within this interface, various views of the data (such as those across time, space, and other attributes) are tightly linked to each other. Interactions, changes, or annotations in terms of any single view are instantaneously propagated and reflected in all the views. The experiential annotation system also supports direct and multi-modal interaction/annotation facilities. For instance, based on location attributes, events and underlying media from a specific geographic region can be selected through a map. Subsequently, this data can be directly voice annotated.

2. DEVELOPING AN EXPERIENTIAL MEDIA ANNOTATION SYSTEM

2.1. *eVITAe*: An experiential PIM system

Our research on developing an experiential media annotation system extends our prior work on experiential personal information management involving the *eVITAe* (electronic vitae) system [6]. In the following, we outline the *eVITAe* system with the goal of describing its existing functionality, which is used to support our current research on experiential media annotation. For details on *eVITAe,* we refer the reader to [19]. The *eVITAe* interface (Figure 1) consists of three modules that are tightly correlated; the *timeline module*, the *location module*, and *the event details module*. These modules work together to display multimedia data and support multimodal interactions with them. In *eVITAe*, changes made in any single module are automatically propagated to all the other modules appropriately. Such designs are called *reflective user interfaces*.

The *timeline module* presents a temporal characterization of events. Here, events are represented through rectangular event-planes that span the time interval during which the event occurred. The event-plane contains icons of media files that support the event. Event-related information, such as event name or location can be displayed by clicking an event plane. Our current research (see following section) augments this functionality by presenting the (audio) annotation, on selection of the corresponding event plane. The *location module* was developed using the OpenMap library [11]. OpenMap provides a complete solution to develop and display locations information using longitude and latitude. Two basic functionalities are implemented: a mouse listener, to select events on the map and a second mouse listener that reacts to selections such as the selection of a region on the map. Once an event selection is detected, it is highlighted across all modules. If a user selects a specific area on the map, as detected by the second listener, a similar set of actions is repeated for each event occurring within that area. Finally, the *event details module* is used to display any detailed event information including the date, text annotation, exact location and event name.



Details

Name: birthday
Start Date: 2004-10-02
End Date: 2004-10-25
Lat: 37.0 Lon: 37.0
Description:

Files
testfile.pdf
sample2.mov

Figure 1: The *eVITAe* interface: Timeline module (top left), Location module (top right), Voice annotation module (bottom left), Event details module (bottom left) and Media Player (botton middle).

2.2. Supporting experiential media annotations

We develop our description of the experiential annotation functionality by enumerating the design features that address the key complexities of the annotation process:

- *Annotation modality*: Text is a cumbersome modality for entering large number of annotations. Furthermore, its non-direct nature increases the complexity of entering information by using it, while at the same time the quality of user experience in terms of naturally interacting with the media decreases. Our design of an experiential annotation, system maintains support for text annotations. However, audio rather than text is supported as the primary annotation modality (voice annotations).

- *Media type independent annotation*: By using a unified event-based multimedia data model, annotations can be associated both with specific media as well as to heterogeneous media. Event annotations are automatically propagated to underlying media. This parallels group annotation capabilities supported by some PIM systems, albeit with greater rigor in terms of semantic consistency. Further, separate annotations can also be associated with each specific media supporting and event. This enriches the descriptive capability and in turn, user experiences.

- *Support for context*: Various types of contextual information (such as spatial, temporal, or event-level) are made available for providing context support during annotations. This is strengthened through the use of reflective interfaces, which simultaneously provide different perspectives about the data.

- In the experiential annotation paradigm, the annotation process is completely integrated with other information management and interaction capabilities such as browsing, querying, and visualization of spatio-temporal event relations. Further, the reflective nature of the user interface aids in real-time propagation of information. Such a closely integrated framework allows users to treat annotation as a bi-directional process; where the system can be tapped to obtain information relevant to the annotation.

In addition to the three modules described in the previous section, two new modules are used to support the annotation process: the *voice annotation module* and the *media player*.

3. SYSTEM DESCRIPTION

In *eVITAe*, the timeline, location, and voice annotation modules work together to build queries to the events database, searching by date, location and annotations. The event detail module and media player are used to visualize the data. The five modules react to changes and actions made in each of the other modules.

To achieve synchronization in all the modules, a pool of objects is built at the core of the system were listeners of each module would be attentive at any change on the properties of the objects in the pool. Since keeping the interface updated is a very expensive operation because of the continuous querying to the database, the usage of views at the database and some simple caching techniques are necessary. For implementation purposes MySQL database is used, using a pool of connections through JDBC to optimize its performance.

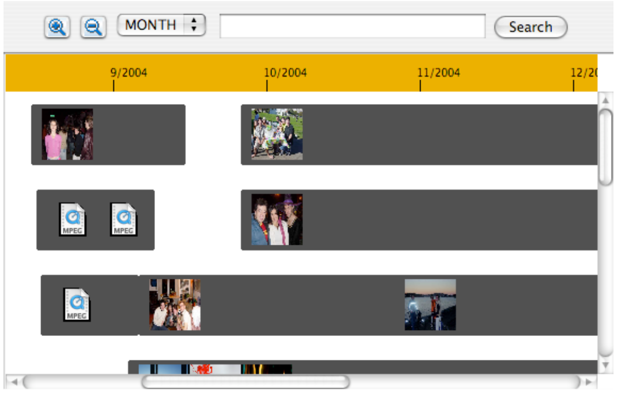


Figure 2: The Timeline module

The timeline module (Figure 2) displays the events in terms of time. Depending on the time granularity used to display the events, the rectangles, which represent the events will display thumbnails of the files contained within. Clicking on a rectangle will select the event showing the details, the specific location and the voice annotation. If an icon contained within the rectangle is clicked, all the actions mentioned before will be performed and the contents of the file will be opened on the media player. Popups, which display the basic information of each event, are also provided.

In the location module two listeners are implemented to set the location of the events two procedures are followed: manual insertion of the latitude and longitude and, as a second option, reading the EXIF[15] header of the files supported by the standard to obtain its GPS data. The map module reacts to actions done in the details and timeline modules, showing the events available, according to the search parameters set in these modules. When an event is selected, its location is highlighted on the map. A region in the map can also be selected to identify all events associated with that location.

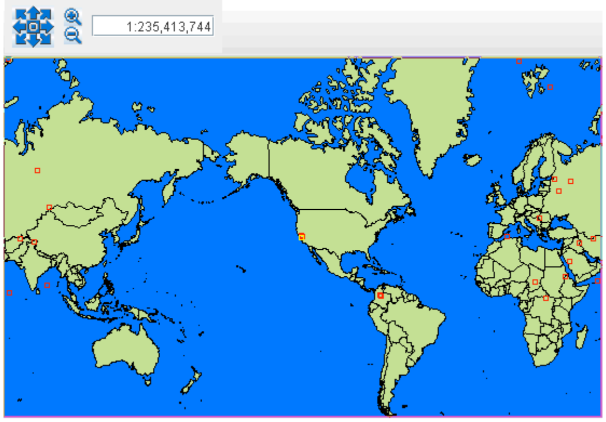


Figure 3: Map Module (OpenMap)

Two new modules are used to support the annotation process. Of these, the *Media Player* (Figure 4) developed using the Quicktime for Java API [14], enables opening a variety of different media such as video files, audio, images and documents (in PDF format). The second module, called the *sound annotation module* has an interface resembling that of a sound recorder (Figure 4) and allows audio annotations to be added or modified at any time. It can be invoked on events or media identified through interactions using the timeline or location modules.

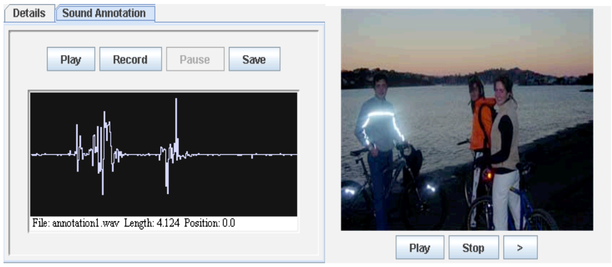


Figure 4: Sound annotation and media player modules.

Due to the explorative nature of the system and use of audio annotations, a way to search through the annotations information is needed. To satisfy this requirement, a voice recognition engine (Sphinx4 [12]) is used to convert the audio annotation to text prior to committing the audio annotations to the database. In order to minimize errors in speech recognition, only those results are used for which the accuracy scores from recognition are sufficiently high. This text information can then be used for search through the annotations.

4. EXPERIMENTS

To evaluate the usability and the quality of experience provided to the users, the enhanced version of *eVITAe* (with the experiential annotation support) was compared with a set of other PIM systems: WWMX [18], Preclick[16] and Picasa [13]. The same media was stored in each of these PIM systems, which consist in a random group of photos and videos relative to four worldwide known events (Athens Olympics, soccer World Cup 2002, anniversary of 9/11 and the 2003 new year's eve celebration in Paris) set in a way that the users had to make some effort to figure out what those events were; all of the media had the corresponding creation date and some of the media was modified to have the EXIF header with the corresponding GPS data for the location necessary in WWMX and *eVITAe*. A questionnaire with thirty-nine questions was designed based on the user experience and insights to the usage of each system. Eight of these questions were repeated for each application for comparing purposes. Ten participants were selected, whose expertise on computer were diverse, to test the system.

The test began by asking some basic questions such as "Are you aware of the concept of media annotation?", where six out ten users respond affirmatively and five of those six had annotated at least once their media like mp3 or photos. To ensure optimal acclimatization, each participant started by using simple systems (such as Preclick) meant for managing specific media (photographs) and then moving up to more complex systems such as WWMX and *eVITAe*.

Every single participant responded that they would not be willing to spend time annotating their media if they had to type it. The overwhelming negative response supports the importance of choosing an appropriate (and natural) modality for annotations.

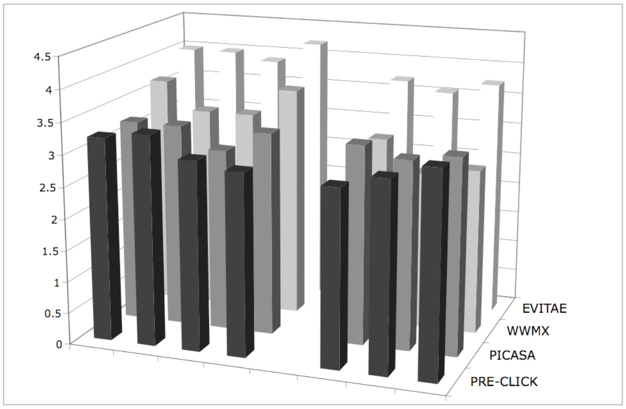


Figure 5: Average user score in usability test.

The participants were asked to measure the insight proposed by the interface to navigate and annotate. Every user responded that Preclick and Picasa provided little or no insights about the data compared to WWMX and *eVITAe* with experiential annotation support. The first two systems earned very similar scores and *eVITAe* with sound annotations outscored the one from WWMX. The latter two systems outperformed the regular photo managers and compared to WWMX's score 3.2 and Picasa's 3.1, *eVITAe* performed even better with a score of 3.7 (Standard deviation 0.377). To be more concrete in the question *"Does the interface implicitly help you to annotate the media?" eVITAe* had a score of 3.8 and the other systems scores between 2.8 and 3.2. Two final questions queried users about their experience, comparing the old version of *eVITAe* with the one supporting experiential annotations. In Figure 5 we present the graph of the scores obtained by each system in the study. These scores are also tabulated in Table 1.

All the users agreed that they would assimilate the information better with sound annotation. As for the question "If you have the possibility of recording your annotation instead of typing it, would that increase the chances of annotating your media?" the average score was 4.6, which shows the huge impact recording will have to the process of annotation. Four participants comments highlighted the facility to associate what they were annotating using the map, timeline and event paradigm. All of the users were particularly attracted to WWMX and *eVITAe* because of their ability to assign locations and navigate through a map.

	PRE-CLICK	PICASA	WWMX	EVITAE	±
At first glance are you able to distinguish and give a fast detailed description of the media you are seeing on the screen?	3.22	3.22	3.67	4.00	0.37
How certain you are of the description you gave to most of the media.	3.33	3.22	3.22	4.00	0.37
Does the interface implicitly help you to annotate the media?	3.00	2.89	3.22	3.89	0.44
Did the interface "pre-organized" the media for you?	2.89	3.22	3.67	4.22	0.57
Did the interface help you to put the media into a more specific context?	2.80	3.17	3.00	3.71	0.39
Did the other media help you associate and put the media into a more specific context?	3.00	3.00	2.50	3.57	0.43
Was the browsing experience fast and intuitive?	3.22	3.11	2.63	3.75	0.46
AVERAGE Score Sample size (10 users)	3.07	3.12	3.13	3.88	0.38

Table 1: Usability test results. Scores between 1 and 5 were 5 is the highest score.

These results compliment previous experiments [2] done in *eVITAe* that prove the efficiency of the system and ease of use, by comparing the number of necessary clicks to find photos.

5. CONCLUSION

In this paper we applied and extended the ideas of experiential computing to address a key challenge of PIM: namely that of annotating media. Our approach is based on extending the experiential framework by combining visual and audio query and interaction capabilities. Experimental results indicate the potential of our approach not just in ameliorating the "annotation bottleneck" but more fundamentally in improving, enriching, and making more natural the user interactions and experience with multimedia data. Our future work in this area will focus on improving the processing and fidelity of annotations involving natural modalities such as audio. Propagation of such annotations also merits further research and experimentation.

REFERENCES

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

[2] Gong B., Singh R., Jain R., "Research explorer: gaining insights through exploration in multimedia scientific data", Proceedings of the 6th ACM MM, 2004.

[3] Appan P., Shevade B., Sundaram H., Birchfield D., "Interfaces for Networked Media Exploration and

Collaborative Annotation”, Intr. Conf. on Intelligent User Interfaces, pp.106-113, 2005

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

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

[6] Singh R., Knickmeyer R., Gupta P., Jain R., “Designing experiential environments for management of personal multimedia”. Proceedings of the 12th annual ACM MM. October 2004

[7] Platt J., Czerwinski M., Field B., “*PhotoTOC: Automatic Clustering for Browsing Personal Photographs”*, Proc. Fourth IEEE Pacific Rim Conference on Multimedia, 2003.

[8] Toyama, K., Logan, R., Roseway, A., Anandan, P. “Geographic location tags on digital images”, in Proc. ACM MM, October, 2003.

[9] Bin Wu, Rahul Singh, Punit Gupta, Ramesh Jain: *eVITAe*: An Event-Based Electronic Chronicle. EDBT 2004: 834-836

[10] Wen-Syan Li, K.Selçuk Candan, Kyoji Hirata, Yoshinori Hara. Supporting efficient multimedia database exploration. VLDB Journal, Volume 9 Issue 4. April 2001.

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

[12] http://cmusphinx.sourceforge.net/sphinx4/

[13] http://www.picasa.com

[14] http://www.apple.com/quicktime

[15] http://www.exif.org

[16] http://www.preclick.com

[17] http://www.mysql.com

[18] http://www.wwmx.org

[19] R. Singh and R. Jain, "From Information-Centric to Experiential Environments", Book Chapter in "Interactive Computation: The New Paradigm",D. Goldin, S. Smolka, P. Wegner, eds., To Appear, Springer Verlag, 2005.