Networked Multimedia Event Exploration

Preetha Appan Hari Sundaram
Arts Media and Engineering Program
Arizona State University
e-mail: { preetha.appan, hari.sundaram }@asu.edu

ABSTRACT

This paper describes a novel, interactive multimodal framework that enables a network of friends to effectively visualize and browse a shared image collection. The framework is very useful for geographically disconnected friends to share experiences. Our solution involves three components – (a) an event model, (b) three new spatio-temporal event exploration schemes, and (c) a novel technique for summarizing the user interaction. We develop a simple multimedia event model, that additionally incorporates the idea of user viewpoints. We also develop new dissimilarity measures between events, that additionally incorporate user context. We develop three, task driven, event exploration environments – (a) spatio-temporal evolution, (b) event cones and (c) viewpoint centric interaction. An original contribution of this paper is to summarize the user-interaction using an interactive framework. We conjecture that an interactive summary serves to recall the original content better, than a static image-based summary. Our user studies indicate that the exploratory environment performs very well.

Categories and Subject descriptors

H.5.1 [**Multimedia Information Systems**]: *Artificial, augmented, and virtual realities,* H.5.5 [**Hypertext/Hypermedia**]: *Navigation,* H.5.2 [**User Interfaces**]: *Theory and methods, User-centered design*

General Terms

Algorithms, Design, Human Factors

Keywords

Networked media, events, browser, summarization

1 INTRODUCTION

In this paper we develop an interactive, exploratory environment that allows a network of users to explore multimodal events of interest. The goal is to allow users to explore the activities of a set of friends. This is an emerging problem in several contexts: (a) sharing of media is important in online social-networks such as Friendster [4], where a set of friends share a set of multimedia experiences. (b) In the continuous archival of personal media [9], and an multimodal event browser is key to efficiently navigating such a large event set, and help provide insight.

There has been prior work in event definition (e.g. [6,10]). Events have been defined formally as a change in state of a system [10]. However, there is no formal relationship of events to user context, and additionally there is no notion of a *viewpoint* associated with events.

There has been prior work in exploring image collections. Prior work in [9,12], provides a summary view of media through thumbnail display of images. Temporal arrangement of media can be explicitly queried for, and similarly other types of Boolean queries can be used to search for media effectively. However, this visualization provides no information on spatio-temporal, or semantic relationships amongst media. Displaying images as clusters around locations in a map as done in [14], gives an idea of how the media are distributed across space. In this work temporal relationships between images are lost. There has been much work in creating image based summarizes in particularly for videos [8,17]. However, this work does not address the issue of *summarizing the event browsing experience* of the large collection of images. If at all, the summary of interactions is presented using traditional image storyboards.

In this paper we develop a simple multimodal event model. We use the dictionary [1] definition of an event: *something that happens*. In our framework, events have the following properties associated with them: name, location, time, media elements (set of images, sounds, text), as well as participants. There are two novel aspects to our model – (a) a set of viewpoints is attached to each event and (b) the semantic relationship between event is dependent on user context. We also develop an event-centric user context model, that adapts to user behavior.

We develop an task-driven event exploration environment – (a) visualize what a particular user has done after (or prior) a specific event (*event conditioning*), (b) visualize the event sequence that lead to two users to meet (*event support*) and finally (c) for a given user, determine *interesting events*.

The environment has three exploration environments – (a) spatio-temporal evolution, (b) event-cones, and (c) viewpoint-centric evolution. Our *spatio-temporal visualization* combines space and time information. We dynamically generate maps using online geographic data available in XML format [3], on which events related to a selected user, unfold over time and space. For each event, it will also create a slideshow of images and text. *Event cones* are snapshot of all the events associated with the selected users. The visualization is important since it allows a non-linear exploration of events, much like our natural thought process. The *viewpoint centric* visualization allows the user to browse the event space by following the individual viewpoint through the event space.

We also present a solution to the problem of summarizing an interaction with the exploratory environment. This is done by analyzing the user behavior with regard to the events, and picking the events most relevant to the user context. The resulting summary is presented as an interaction – using the viewpoint centric visualization. Our user studies indicates that the exploratory environment is very well received.

MM'04, October 10–16, 2004, New York, New York, USA. Copyright 2004 ACM 1-58113-893-8/04/0010...$5.00.

The rest of this paper is organized as follows. In the next section we define the problem. In section 3, we define our event model, and define event distances. In section 4 we present our model for user context. In section 5, we present our event exploration framework, and in section 6, we present our solution for summarizing the event exploration. In section 7, we present experiments, and then we discuss limitations and our conclusions.

2 PROBLEM STATEMENT

In this section, we enumerate the tasks needed to solve two problems – (a) creating an exploratory interface for a group of friends exploring a distributed networked media collection and (b) summarizing the user's interaction with the system, while exploring those events.

2.1 Event Exploration

We define "exploration of a media collection" in this paper (in addition to simply browsing) to imply the following tasks (we shall use imaginary users – Alice and Bob):

- *Event conditioning*: What has Alice done since a particular event? (e.g. party at Bob's house)
- *Event support*: What was the sequence of events that led Alice to meet Bob at particular location? (e.g. Alice meets Bob at the mall).
- *Interesting events*: What are the set of events of interest to Alice? Note that interest is user context dependent.

We believe that this list of tasks , while not exhaustive, are most common for a group of users .

2.2 Event interaction summarization

With increase in the number of events in the system, the users will like to view summaries of large event sets. For example, Alice should be able to ask the system to give her the most important events in her life, related to her hobby – (like flying for instance), between the year 1995 to 2000. The number of events in the summary will primarily depend on user time constraints. For determining the event summary, the semantic distance between the user's profile and the *entire* event set is used to compute the summary events.

We are proposing a highly interactive exploratory environment, where the user will browse many events. Clearly, we need a framework *to summarize the interaction*. The interaction summary will be generated using the events browsed by the user in the current interaction. In the following section, we begin by formally defining our event model.

3 EVENT MODEL

In this section we formally define our event model. In the following section we discuss the following: (a) event definition, (b) media associated with an event, (c) viewpoints associated with events (d) friends associated with user

3.1 Event Definition

We use the dictionary [1] definition of an event: *something that happens*. In our framework, event have the following properties associated with them: name, location, time, media elements (set of images, sounds, text), as well as participants. Note that both location and time are specified using a set. Our model of an event follows from the fact that every member of a digital personal media collection has relevance to a real-life user activity – *who, when, where, what* to some user. General definitions of events (e.g. [6]) while important, are not needed in our application. In the discussions that follow, we shall use the following assumptions and notations. We assume that there are k users, N network locations, and a total of M media elements. The set of users is designated as U, and the event set as E.

3.1.1 Media associated with an event

In our framework, while an event can be associated with many media elements, we make the simplifying assumption that media elements can be only associated with one event. Given a set of images uploaded by the user, an image is associated with an event if the media element was captured at the same time and place. The location and time can come from a GPS enabled digital camera. Images also have same meta-data fields as an event, except that they are associated with a single location and time. Formally we can define an association function between a media element m and an event e:

$$A(e,m) = \begin{cases} True & if\ t_m \in T_e \wedge l_m \in L_e \\ False & otherwise \end{cases}, \qquad <1>$$

where t_m and l_m is the time and location associated with the media element, and where T_e and L_e are the ranges of times and locations respectively, associated with the event. Note that in our framework, events create a complete partition the media set.

3.1.2 Friends associated with a user

Given a network of users, we assume that each user provides a list of people whose photos in whom she is interested. When such information is not available, we define the set of friends (F_u) of a user u, as fellow users' v who have participated with u in at least one event e. Formally:

$$F_u = \{v \mid v \in U \wedge C(u,v)\},$$
$$C(u,v) = \begin{cases} True & if \quad \exists e \in E \quad s.t.\ P(u,e) \wedge P(v,e) \\ False & otherwise \end{cases}, \qquad <2>$$

Where C is the co-participation function that is true when both users u and v are present at the same event. The event participation function P is true only when the user was present at the specific event.

3.1.3 Viewpoints associated with an event

Given a set of events, each of the elements in this set can have k viewpoints associated with them. By viewpoint, we mean the set of media (images, text, and sound) associated with an event, that each user submits to the system. This is akin to a personal narrative about an event. Note that there is a distinction between the number of participants in an event, and the number of viewpoints. For instance not all participants might have a viewpoint about the event, and someone who hasn't participated in the event might still have a viewpoint about it. Formally,

$$P(v,e) = \begin{cases} True\ if\ \ v \in V(e), \\ False\ otherwise \end{cases} \qquad <3>$$

where, $V(e)$ is the set of viewpoints associated with event e. With this basic event and viewpoint framework in place, we now describe how we calculate event distances.

3.2 Event Distances

In this section we describe calculation of event distances in our framework. We first define a distance between two events independent of user content. Then, we shall two additional

distances – distance between event and user profile and distance between two events with respect to a common user profile, that depend on a user context model. The user context model is discussed in detail in section 4.1.

Events have the following properties: *who, what, when, where*. Thus events could be related by space (*where*), time (*when*), and participants (*who*) as well as semantically related (*what*). Now we describe calculation of event distances by taking all of the above factors into account. Note that all the distances are normalized to be between zero and one. In all the following, e_i and e_j are defined to be any two events.

3.2.1 Context independent event distances

Distances between events that are context independent are based on the four above mentioned properties. The *temporal distance* (*when)* between two events e_i and e_j increases with the time difference between the two events. This is intuitive because of our assumption that as the time difference between events becomes larger, they no longer influence each other.

$$d_t(e_i, e_j) = 1 - e^{-\alpha \Delta t} \qquad \Delta t = \left| t_i - t_j \right| \qquad <4>$$

Here d_t is the temporal distance, α is the decay factor, Δt is the time difference between the two events, t_i and t_j are the times of events e_i and e_j respectively.

The *geographical distance* d_g (*what*) between events is defined as the distance in miles between their respective locations by the maximum mile distance (among every possible pair of events).

$$d_g(e_i, e_j) = \frac{d_m(e_i, e_j)}{d_{\max}} \qquad <5>$$

Here, d_m is the distance in miles between the locations of events and d_{max} is the maximum distance (amongst every pair of events in the entire data set). The distance can be easily calculated by mapping the event address to the latitude / longitude, and then computing the distance between two latitude / longitude pairs.

The *participant distance* d_p (*who*) is calculated as ratio of the number of common participants between two events, by the minimum of number of participants in both events.

$$d_p(e_i, e_j) = 1 - \frac{|P(e_i) \cap P(e_j)|}{\min\{|P(e_i)|, |P(e_j)|\}} \qquad <6>$$

Where $P(e_i)$ is the set of participants of event e_i, $|P(e_i)|$ indicates the cardinality of the set $P(e_i)$.

We define *semantic event distance* (*what*) as follows: we use the concept distance (ref. equation <14>) to define the semantic distance between events. Since every event is annotated with a few sentences of text, we first remove stop words and perform stemming to reduce each event's annotation to a set of words. These form the basic concepts of the event. Then the semantic distance between the two events:

$$d_1(e_i, e_j) = \frac{1}{n} \sum_{a=1}^{n} \min_{b \in 1..m} \{d_c(w_a, w_b)\},\ w_a \in e_i, w_b \in e_j, \qquad d_s(e_i, e_j) = \frac{d_1(e_i, e_j) + d_1(e_i, e_j)}{2}, \qquad <7>$$

where d_c is the distance between concepts, and where w_i, and w_j are concepts belonging to events e_i and e_j respectively. In the absence of *a priori* knowledge of importance of each event parameter (*who, what, when, where*), we assume that all of them are equally important. Hence the comprehensive event context independent distance d_{comp} is calculated as follows:

$$d_{comp}(e_i, e_j) = \omega_1 d_t + \omega_2 d_g + \omega_3 d_p + \omega_4 d_s, \qquad \omega_1 + \omega_2 + \omega_3 + \omega_4 = 1, \qquad <8>$$

where ω_i are the weights and are assumed to be equal.

3.2.2 Distance between user profile and any event

The distance between each event and the user profile is defined as the average of the distance between each concept in the event and all the concepts in the user profile.

$$d_{ep}(e_i, U_p) = \frac{1}{n} \sum_{a=1}^{n} \min_{b \in 1..m} \{d_c(\alpha_a, \beta_b)\},\ \alpha_a \in e_i, \beta_b \in U_p, \qquad <9>$$

Here U_p is the user profile, α_a and β_b are concepts in the event and the user profile respectively, and d_c is calculated using equation <14> in section 4.1.

3.2.3 User-context dependent event distance

The distance between two events changes when conditioned with respect to a common user context. We are using the intuitive idea that events close to a common context, will also become closer to *each other* conditioned on the common user context. For example, consider two events A and B dealing with riding bikes and roses respectively. In the case that we have a user who is interested in biking as well as gardening, then the semantic distance between events A and B *conditioned on the user,* should decrease. Note that the semantic distance between the two concepts (bikes and roses) is very large, when computed using WordNet – equation <14> in section 4.1.

The user context dependent semantic distance is calculated by multiplying their comprehensive distance (from equation <8>), with the average of their distances from the common user profile U_P. Thus,

$$d_{ce}(e_i, e_j) = d_{comp}(e_i, e_j) \bullet \left(\frac{d_{ep}(e_i, U_p) + d_{ep}(e_j, U_p)}{2} \right), \qquad <10>$$

Note that this also helps maintain the triangle inequality constraint i.e. $d_{ce}(e_1,e_2) \le d_{ep}(e_1,U_p)+d_{ep}(e_2,U_p)$. This constraint would be violated when the two events are not semantically related (e.g. bikes and roses as mentioned above).

In this section, we discussed our event model and event distance calculation in detail. In particular we derived context independent distances between two events and further derived distances between two events when conditioned on user context.

4 CONTEXT

In this section we present a definition of context in our system. This builds upon prior work on the user context model [15]. A key improvement in this paper is that context model developed here is *event centric*, as opposed to being media centric [15].

4.1 User Context Models

The dictionary [1] defines context as *the interrelated conditions in which something exists or occurs*. These conditions could be meta data from the events that the user has explored, personal information about the user (e.g. "female", "skiing" etc.), the user's past actions, and the tasks she wants to solve etc.

We have adapted prior work on multimedia context models in [15]. The formal model is defined using a concept-net – a graph $G = <V, E, W>$ where the nodes $v_i \in V$ represent the concepts, the edges $e_{ij} \in E$ represent the type of relationship (semantic,

spatio-temporal, feature-level) between the nodes i and j and $w_{ij} \in W$, specifies the strength of the relationship between the two nodes. A concept node is associated with a specific instance that could be an image, video, an audio segment, text etc. Thus, the context is defined to be the union of concept-nets:

$$C = \bigcup_{i=1}^{k} G_i \qquad <11>$$

where C is the context, k is the total number of concept-nets and G_i is the ith concept-net.

The user context comprises of concept nets composed from (a) the initial user profile that the user inputs (stating her interests, background etc.), (b) the event viewing history (that establishes what events the user has seen) (c) her behavior (which viewpoints the user visited, how much time he/she spent on each of the media elements), (d) her tasks (what he/she is trying to solve through summarization).

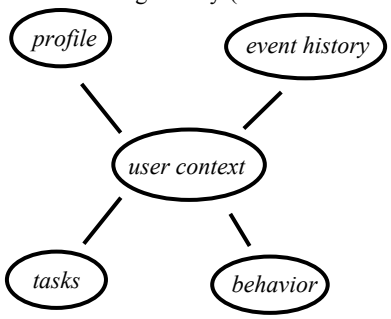

Figure 1: User context**:** The user context model is composed of – (a) profile (b) event history (c) behavior and (d) tasks

Figure 1 illustrates this relationship.

The initial user context is modeled using the textual inputs she is asked to provide as a part of her user profile - details like gender, ethnicity, age, profession, cultural interests etc. The relationships between the concepts are given by generalization / specialization of WordNet. The user's context is then the union of such concept nets. The context is updated with the event viewing history (including textual concepts as well as media), the user behavior (amount spent per viewpoint, for each event) and the tasks, as the user starts interacting with the system.

We have shown how to construct a context model that encapsulates *who* the user is, *how* he interacted with the system and *what* are his tasks with respect to the summary.

4.2 Semantic Distance between event concepts

In order to assign importance scores to events towards determining which of the events need to appear in the summary, we need to calculate the distance of each event from the user context. We use a modified version of the implication measure defined in [15]. This is summarized in the next two paragraphs.

Every concept α has two entities associated with it – the parent concept, and the children concepts (as given by the WordNet hierarchy). Each entity implies the concept with a different weight – ω_1 (the parent) and ω_2 (all the children). Hence the *implication* that the concept α is true given that another concept β is true is computed as follows:

$$I(\alpha \mid \beta = T) = \omega_1 I(parent \mid \beta = T) + \frac{\omega_2}{k}\sum_{i=1}^{k} I(c_i \mid \beta = T),$$
$$I(\beta \mid \beta = T) = 1, \qquad <12>$$
$$\omega_1 + \omega_2 = 1,$$

Where I is the implication strength, and k is the number of children of the conceptα, c_i is the ith child of the concept, and where ω_1 and ω_2 are the weights attached to the implications of the parents and the children respectively. The weight ω_1 is computed to be inversely proportional to the number of children of the *parent* concept, α. i.e.

$$\omega_1 \propto 1/m,$$

where m is the number of children of the parent.
The distance between the two concepts is determined as follows:

$$d_U = 1 - I(\alpha \mid \beta = T),$$
$$d(\alpha \mid \beta = T) = d_U / \sqrt{f_\alpha \bullet f_\beta}, \qquad <13>$$
$$d_2(\alpha, \beta) = \big(d(\alpha \mid \beta) + d(\beta \mid \alpha)\big)/2,$$

where d_U is the un-weighted distance between the two concepts. f_α and f_β, represent normalized knowledge priors for concepts α and β. The priors are used to re-weight the distance. These could be set by the user, as part of her context model. We easily make the distance symmetric.

The above distance measure assumes the sense of a word to be the one that is most likely to be used. This information is obtained from the tag count of each word sense. However, this would lead to errors in cases where the event was annotated with a word whose sense was *not* the most popularly used one. For example, an event annotated with the word "party", would lead to its sense of "political party" be picked rather than the intended sense of a celebration. Hence we use the tag count of each sense of a word to calculate the probability of each sense of the word, and thus modify the distance measure as follows:

$$d_c(\alpha, \beta) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{\alpha,i}\; p_{\beta,i}\; d_2(\alpha_i, \beta_j), \qquad <14>$$

where, concept α has n senses and concept β has m senses. The probability of occurrence of each sense of the word (p_α) is calculated in the following manner:

$$p_{\beta,i} = \frac{t_{\beta,i}}{\sum_{i=1}^{n} t_{\beta,i}}, \qquad <15>$$

where $t_{\beta,i}$ is the tag count of the ith sense as given in WordNet [13]. We have described how we calculate distances between two semantic concepts. Concepts in the user profile evolve with time due to the user interaction. Now, we will describe how concepts get added to the user profile, depending on how the user interacts with events in the system.

4.3 Concept evolution

In our system, concepts get added to the user profile as the user interacts with events. We use a *leaky bucket* model from [15] for modeling evolution of context. Concepts automatically acquired by the system into the user profile are slowly lost over time at a fixed rate, unless they are reinforced by the user visiting related events. This concept reinforcement is analogous to forming associations of a current event with an earlier event that generated feelings of a similar experience.

The weight of concepts in the user profile grow and decays exponentially with time. Concepts that don't get reinforced periodically are lost. We used the growth and decay equations given in [15], to calculate concept weights based on time spent by the user in each event.

We have discussed how we model events and user context in our system, calculation of semantic distance between concepts and

concept evolution in the previous sections. In the following section we will describe our event exploration framework.

5 EVENT EXPLORATION

In this section we discuss our event exploration interface, by breaking it down into two sub-problems – (a) task-based media selection (b) media presentation.

5.1 Task based media Selection

In the task-based media selection sub-problem, we select a subset from the entire media collection that enables the user to efficiently perform each task described in section 2.

Now for each of the tasks described in section 2, we now present our solution to the media selection problem. Given the user set U, media set M, event set E, current user u, the set of friends F_u of u, we describe below a set of exploratory tasks that a user would like to perform. Then we present solutions to find a subset of the media set to perform the task. As before, let use two imaginary users Alice and Bob.

5.1.1 Event conditioning

What have Alice's friends done since (or prior) to a particular event? (e.g. party at Bob's house). The set of media elements that need to displayed will involve those in events that Alice's friends participated after the specific event. Note that *not* all media elements that belong to events in which Alice's friends (F_a) participate are relevant. This is because they can contain people that are of not interest to her. Formally, the conditional media set S_b, given a specific event b is defined as follows:

$$P_b = \{e \mid e \in E \wedge \exists f \in F_a \ s.t.\ P(f,e) \wedge t_e \geq t_b\},$$
$$S_b = \{x \mid x \in M \wedge \exists e \in P_b \ s.t.\ A(e,x) \wedge W(x,F_a)\}, \quad <16>$$

where, P_b is the set of events that occur with Alice's friends *after* event b, and where S_b is the set of media elements and where the W function is true if any of Alice's friends (F_a) are present in the *who* field of the media element; t_e and t_b are the times associated with events e and b respectively. A slight change ($t_e \leq t_b$) to equation <16> will help her determine media elements *prior* to event b.

5.1.2 Event support

What were the sequence of events that led Alice to meet Bob at particular event? (e.g. Alice meets Bob at the Mall). The set of media elements that need to be displayed are those occurred prior to the current event, and which contain either Alice or Bob in them. Formally:

$$P_c = \{e \mid e \in E \wedge P(Alice,e) \vee P(Bob,e) \wedge t_e \leq t_c\},$$
$$U = \{Alice, Bob\}, \quad <17>$$
$$S_c = \{x \mid x \in M \wedge \exists e \in P_c \ s.t.\ A(e,x) \wedge W(x,U)\},$$

where, c is the event where the two meet, S_c is the media elements to be displayed.

5.1.3 Interesting events:

What are the set of events of interest to Alice? An event is deemed interesting to a user, if the semantic distance between the meta-data description of an event and the user profile is less than a certain threshold. In our present implementation, we use a simplification – an interesting event is one that took place at a location visited by the user, or if the event contains at least one friend. Formally:

$$I_v = \{e \mid e \in E \wedge \exists f \in F_v \ s.t.\ P(f,e) \vee l_e \in L_v\}, <18>$$

where I_v is the interesting event set, F_v is the friend set of user v, and L_v is the location set (i.e. these are all the locations visited by the user v), and where l_e is the location of the event e.

5.2 Media Presentation

We now discuss the media presentation framework. After selecting the task-centered media, the next step is present it in an optimal manner. The primary goals while developing our media presentation techniques were as follows:

- They should allow users to explore relationships amongst themselves and their friends.
- They should enable users to understand how events unfold time and space.

We discuss three novel presentation schemes in this section – (a) Spatio-temporal evolution, (b) event cones. (c) Viewpoint based evolution

5.2.1 Spatio-Temporal Evolution

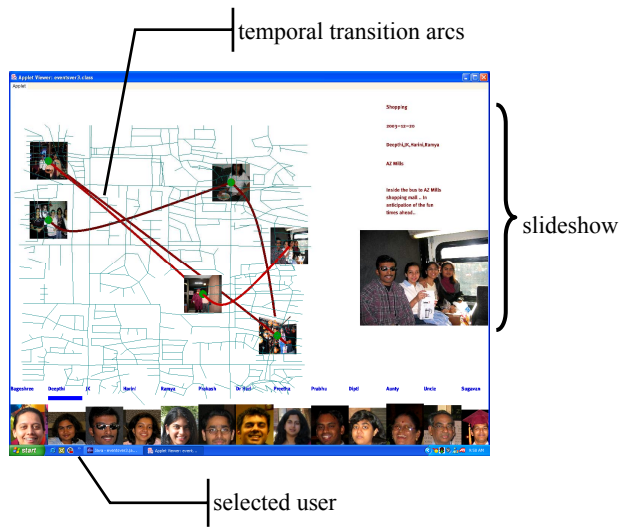

Figure 2: Spatio-Temporal evolution of digital media in our system. Once a user selects a person from the network, the system shows the events that the user participates in, over space and time.

This visualization will address the task of *event conditioning*, as well the *interesting events* task. Prior work in [9,12] , show collections of digital media organized either by space or time. Our visualization expands on these by combining space and time information. We dynamically generate maps using online geographic data available in XML format [3], on which events unfold over time and space. Media associated with events appear at different locations, ordered in ascending or descending temporal order (as chosen by the user). The spatio-temporal visualization scheme is shown in Figure 2.

The interaction begins by the user selecting a friend. The system responds by showing the spatio-temporal events, over time and by location. Since each event is associated with a collection of media elements, we present them (images, sounds and text) as a slideshow over time. The time assigned to rendering the images associated with each event is estimated using comprehension

times estimates using visual complexity as in our prior work [16] event The temporal transitions between locations is indicated using arcs, and saturation of the arc color is used to indicate time (the most recent arc is the most saturated).

We now present the solution to the problem of visualizing *interesting events*. We begin by displaying all the events associated with selected user. When the user moves her mouse over a specific event location, then all the friends of the selected user are who *participated in the same event* are highlighted in green. The other friends of the selected user, who participated in *other events at the same location,* are highlighted in purple.

The user can click on each highlighted user – clicking on the friends highlight in green, will result in spatio-temporal evolution for that person, starting with the current event.

5.2.2 Event Cone

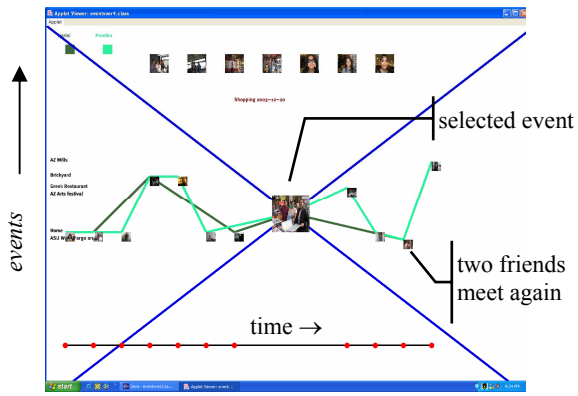

Figure 3: An event cone. This is a snapshot of all the events associated with the selected users. The events are ordered along the vertical axis and the horizontal axis shows time.

In this section we present our solution to the task of *event support*. While the spatio-temporal evolution does enable us to understand what happens to a person over time, it doesn't allow, it doesn't allow us to develop a proper understanding of the event relationships.

We browse a photo collection in a non-linear manner (i.e. we do not follow temporal order). Pictures trigger memories – and we pose queries – "What happened to my friend Alice after that day?," or "What led to Alice and me meeting at that event?." While they can be answered by constructing and executing complex Boolean queries and then searching amongst the matching images, our visualization attempts to provide a summarized snapshot of time and space. We call such a summary – *event cone*.

When the user selects her event of interest in the spatio-temporal browser, an event cone of that event is shown, which shows how different friends of the selected user who participated in the event, came together at that point of time, and also what happened to them afterwards. Each user has a distinct path and is individually colored. When the paths intersect, they imply that the friends meet.

5.2.3 Viewpoint Based Evolution

In this section we present our visualization scheme that allows for viewpoint based exploration of events. The idea that users desire *agency* [7] – users proactively interact with the system to control the flow of story, is incorporated in this visualization.

We construct an exploratory environment that provides users with different ways to navigate through the same event space. As discussed in section 3.1, every event has a set of viewpoints associated with it Users are presented with a dynamic visualization that lets them explore events based on the viewpoints associated with them. Initially, media associated with the selected user's events (according to the time range chosen by the user) appear on locations in the map. Figure 4 illustrates this. (Note that we have not included snapshots from the application since the images were too small to communicate the idea well).

As seen in the illustration below, the *k* viewpoints associated with an event, are displayed around it. Also the set of supporting events for the current event and one succeeding event (i.e. for the currently selected viewpoint) are shown. The set of supporting events for each event are calculated as the union of all events that support each of the *k* viewpoints of the event .

Now users can select view points around the current event, to see a slideshow of media associated with the event (and the selected viewpoint). Also supporting events and succeeding events for each event, changes according to the chosen viewpoint. The user can change the current event at any time, by simply clicking on any event (supporting as well as future event). Then, the system creates the supports and future events for this new current event. Note that the viewpoints associated with this new current event will be different.

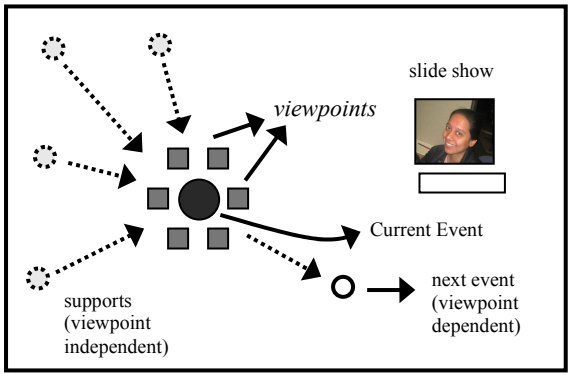

Figure 4: Viewpoint based exploration: Every event has a set of supporting events and one succeeding event, according to the currently selected viewpoint. The user can dynamically change the viewpoint, thus changing the slideshow associated with the viewpoint, as well as the future event.

Thus this is a very dynamic environment that allows users to explore events/viewpoints in a non linear manner. The time spent by the user per event is logged in order to add concepts associated with the events to the user's profile. This information is used to summarize the events and the user's interaction, as described in the following section.

6 SUMMARIZATION

We describe in this section how we summarize the interaction the user had with the system, while browsing events. The same principles apply while summarizing events as well.

6.1 Interaction Summarization

We use the context model described in section 4.1 in order to summarize event interaction. We present the summary itself as an interaction. This enables users to revisit events in a manner similar to their original interaction, rather than seeing a passive slideshow of summary events, or a story board of important media from the events. The summary events are computed in the following manner, assuming that (a) the user provides k, (the number of events she wants in the summary) and the time range of the events he wants to summarize (b) the user has interacted with the event space that he wants to summarize. The interaction summary can be created in the following steps:

Building context model: Concepts are added to the user profile, weighted by the time she spent in each event. The memory model described in section 4.3 is used to grow and decay concept weights.

Selecting top level summary events: The top k events are picked by decreasing order of distance from user profile. The distance between each event and the user context is calculated using equation <9>.

Selecting semantically related support events: In order to determine the supports of each event, we need to determine the conditional distance between two events. This is the pair wise context dependent event distance formula (ref. equation <10>), to find m other related events that are close to the top k events picked above. These m events act as supporting events to the events in the summary. This step has a two fold purpose. First, it increases the coherence of the presentation by providing support to understand why a particular event occurred; also it provides the user interactivity.

We eliminate redundant supports in the following way. We examine the supporting events and check if there are supports that support multiple events, if so they are picked for the summary presentation. Also for events e_k that have more than one support, we remove those supports that if they directly or indirectly support other supporting events of e_k. The events chosen thus are presented to the user as an *interactive summary*, using the viewpoint based evolution scheme, described in the previous section. Note that the user has the choice to explore media within each event.

6.2 Event Summarization

Thus we have seen how user's interaction with the events can be used to summarize their interaction. If instead we would like to summarize that occurred in a certain time-range – say all parties that Bob attended between 1997-2000, we do as follows.

For event summarization, we follow the same steps as above , except that this time we are summarizing *events that the user hasn't interacted with* , based on information in her basic profile and his interaction with some other events (if any). So if there are events that are semantically related amongst events that the user has seen already, and the event set she wants to summarize, they would still appear in the summary, because we use a context dependent event distance formula (ref. equation <10>). For example, if Bob wants to view a summary of Alice's event set and Bob's user profile says he is interested in travel, then an event in which Alice traveled around Europe would be selected.

7 EXPERIMENTS

We designed a pilot user study to test if our interactive system was able to achieve the goals stated in section one. The system was implemented with processing [5]. Five users were chosen, out of which four were graduate students at ASU and one working professional. Two users provided the set of digital media used in the experiment. The numbers in bold indicate that the results were significant (using the t-test) at 95% or better.

7.1 Evaluation of Presentation Schemes

To evaluate our system's presentation schemes, users were asked to evaluate the system's quality (through criteria such as ease of use, picture organization, ability to solve tasks etc). Each user was asked to pick a answer on a scale strongly disagree – strongly agree (1-7). Table 1 and Table 2 summarize the result of this study. Our results show that the browsing system was very well liked, and each of the visualization schemes solved the tasks (described in section 5.2) that they were best suited to solve. Users were also asked to rank the three visualizations on overall preferences and ease of use. The *viewpoint centric visualization* was preferred, but also considered more difficult to use than the others. This could be due to the non linear nature of the of the viewpoint interaction.

Table 1: user feedback evaluating the interface

Exploration	Score
I like this browser	**6.4** / 7.0
This browser is easy to use	**5.2** / 7.0
I am satisfied with the picture organization	**6.6** / 7.0
A month from now, I would still be interested in using the application	**6.6** / 7.0

Table 2: User study evaluating the effectiveness of the visualizations with respect to specific tasks

Tasks	Spatio temporal	Event Cone	Viewpoint based
Can determine how people met at a certain place/time	**5.4** / 7	**6** / 7	**5.4** / 7
Can determine what happened to someone after a certain event	**6** / 7	**6.8** / 7	**6.4** / 7
Can determine other persons who have events in locations of interest.	5.4 / 7	**6.2** / 7	**7** / 7

7.2 Evaluation of the summary

We now describe the experimental setup to our interactive summarization schemes. We conducted two sets of experiments, that compared the interactive scheme with the traditional image storyboard as well as compared different interaction schemes.

We first determined if an interactive summary was preferred to a traditional image storyboard. We created an image storyboard by by ranking events on the time spent and picking the cluster center of the event's photographs to generate the storyboard. Users were asked to state their preference between the story board and the optimal interactive summary. All users preferred the interactive summary..

We evaluated our optimal summarization scheme in the following way. We generated three summaries – summary one (optimal), summary two (temporal) was generated using

temporal distance between events (equation <4>) instead of the corresponding semantic distance (equation <14>). Summary three (random) was generated by picking events at random. Users were asked to compare all three summaries, in terms of whether the events in them were semantically related. Table 3 shows the results.

All users except one stated that summary one was the best in terms of bringing out semantically related events. This occurred because the events picked at random for that user, happened to be semantically related by coincidence.

Table 3: User study results evaluating the interactive summaries

Summarization	Optimal	Temporal	Random
Events in the summary are semantically related	**6.4** / 7	**4.8** / 7	2.8 / 7
The information in the interactive summary is related to the information viewed	**6** / 7	**5.2** / 7	3.4 / 7

8 LIMITATIONS OF CURRENT WORK

We would like to enumerate the limitations of this work:

- *Sense Disambiguation:* Since events are annotated with keywords, there are errors in picking the senses of those words. Although, we use tag count probabilities to address this issue, a more robust model is needed.
- *Common Sense Reasoning:* The context model uses concepts whose weights grow and decay according to his interaction with events. Our semantic distance depends upon WordNet but ignores common sense relationships [2] amongst events.
- *Event structures:* We do not analyze relationships or temporal structures amongst events.

9 CONCLUSIONS

We now present our conclusions. In this paper we have developed an multimodal event browsing tool, that allows a network of friends to share experiences.

We have developed a simple event model that incorporates the idea of a viewpoint. We developed event distances that incorporates user context. We also presented user context models. We developed three novel event exploration schemes – (a) spatio-temporal evolution, (b) event cones and (c) viewpoint centric browsing. Finally we developed an interactive summary to summarize a large event set.

The user studies indicate that the exploratory event browsing environment is very much liked. The results indicate that the *viewpoint centric* interaction was preferred. We also showed that optimal interaction summary was better liked than other interactive summaries; it was also preferred to the image storyboard by all. We plan on doing the following – (a) developing new collaborative annotation techniques to assist photo upload, and (b) use tangible media interfaces [11].

10 REFERENCES

[1] *Merriam Webster Dictionary* http://www.m-w.com.

[2] *OpenCyc* http://www.opencyc.org.

[3] *TIGER maps* http://imagemaps.mle.ie/.

[4] *Friendster* http://www.friendster.com.

[5] *Processing* . http://proce55ing.net.

[6] S. BOLL (2003). *MediAether - an Event Space for Context-Aware Multimedia Experience*, 1'st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003, Berkeley CA.,

[7] K. M. BROOKS (1997). *Do story agents use rocking chairs? The theory and implementation of one model for computational narrative*, Proceedings of the fourth ACM international conference on Multimedia, 317 - 328,

[8] M. G. CHRISTEL, A. G. HAUPTMANN, H. D. WACTLAR, et al. (2002). *Collages as dynamic summaries for news video*, Proceedings of the 10th ACM international conference on Multimedia, Juan Les-Pins, France, 561-569, Dec. 2002.

[9] J. GEMMELL, G. BELL, R. LUEDER, et al. (2002). *MyLifeBits: fulfilling the Memex vision*, Proceedings of the 10th ACM international conference on Multimedia, Juan Les-Pins, France, pp. 235-238, Dec. 2002.

[10] A. GRAHAM, H. GARCIA-MOLINA, A. PAEPCKE, et al. (2002). *Time as essence for photo browsing through personal digital libraries*, Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries, Portland, Oregon, USA, 326 - 335,

[11] H. ISHII and B. ULLMER (1997). *Tangible bits: towards seamless interfaces between people, bits and atoms*. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press**:** 234--241.

[12] H. KANG and B. SHNEIDERMAN (2000). *Visualization Methods for Personal Photo Collections : Browsing and Searching in the PhotoFinder*, (ICME2000), New York:, IEEE, 1539-1542,

[13] G. A. MILLER, R. BECKWITH, C. FELLBAUM, et al. (1993). *Introduction to WordNet: An On-line Lexical Database.* International Journal of Lexicography **3**(4): 235-244.

[14] C. SHEN, B. MOGHADDAM, N. LESH, et al. (2001). *Personal Digital Historian: User Interface Design*, In Proceedings of Extended Abstracts of Human Factors in Computing Systems (CHI 2001), ACM Press,

[15] H. SRIDHARAN, H. SUNDARAM and T. RIKAKIS (2003). *Computational models for experiences in the arts and multimedia*, 1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003, Berkeley CA,

[16] H. SUNDARAM and S.-F. CHANG (2001). *Constrained utility maximization for generating visual skims*, IEEE Workshop on Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001), Kauai HI USA, pp. 124-131, Dec. 2001.

[17] S. UCHIHASHI, J. FOOTE, A. GIRGENSOHN, et al. (1999). *Video Manga: generating semantically meaningful video summaries*, Proceedings of the 7th ACM international conference on Multimedia, Orlando, Florida, USA, 383-392, 1999.