Viewing and Annotating Media with MemoryNet

Rakhi Rajani, Alex Vorbau
Mobile and Media Systems Laboratory
HP Laboratories Palo Alto
HPL-2004-36
March 2, 2004*

E-mail: {rarajani,alexv}@hpl.hp.com

social computing, social navigation, home, user studies, user experience, design

In this paper we describe an investigation into how we might share and annotate media objects (namely photographs) among people in our personal networks. We describe a prototype, the MemoryNet Viewer (MNV) and present results of a user study. We conclude with future developments.

* Internal Accession Date Only
To be published in and presented at CHI 2004, 24-29 April 2004, Vienna, Austria
Approved for External Publication

Viewing and Annotating Media with MemoryNet

Rakhi Rajani, Alex Vorbau
Hewlett Packard Laboratories
1501 Page Mill Road, Palo Alto, CA, 94304
{rarajani,alexv}@hpl.hp.com

ABSTRACT
In this paper we describe an investigation into how we might share and annotate media objects (namely photographs) among people in our personal networks. We describe a prototype, the MemoryNet Viewer (MNV) and present results of a user study. We conclude with future developments.

Categories & Subject Descriptors:
H5.2 User Interfaces - *Prototyping; Evaluation/Methodology*

General Terms:
Design; Human Factors

Keywords:
Social computing; video content/communications; home; user studies; user experience design; awareness; sharing; annotation

INTRODUCTION
The rate of digital capture is growing and we collect vast amounts of images for our personal media collections. Despite this, the sharing process is often limited to one directional means where we author finished media that is then sent on, or we rely on 'display' sites to host our images. The opportunities for interaction with our media are not fully realized and responses to our collections are lost or at least limited.

Sharing in more natural situations is a dynamic and interactive activity that allows for stories to develop and links to emerge between media objects and the people who own it [5]. The technological solutions we currently employ often inhibit these processes.

The MemoryNet project is exploring how we might create an open system of interconnected personal media that can be used to facilitate relationships in personal networks.

CHI 2004, *APRIL 24–29, 2004, VIENNA, AUSTRIA.*
ACM 1-58113-703-6/04/0004.

With MemoryNet we are developing an infrastructure where:

1. The sharing process is as simple as viewing and interacting with your own media
2. Communication is spontaneous and driven by interaction with your media.

Current applications and services provide some of this functionality. Asynchronous services such as Ofoto [1] and Shutterfly [2] allow users to post their media for others to see and also facilitate the addition of small comments. Synchronous applications such as Picassa's Hello [3] and Picshare [8] give users the opportunity to share and converse about their photographs in real time, and Instapix [4] permits users to synchronize their media with that of their friends.

With the asynchronous services, the users' ability to get reactions to their media is more inhibited and with the synchronous services users have to devote their full attention to the task of sharing and talking about photographs. While both of these provide benefit, and we acknowledge their value in allowing people to share their media, users must devote a significant amount of their time to interacting with the services to get any value, and annotating media is a more cumbersome process. When this is the case, people often tend to put off annotating and viewing their media collections despite their desire to do so [6].

With these observations in mind, we began to explore how we might create ways for media to infiltrate the home space and just be present, giving users the opportunity to view their collections (and those of people in their personal networks) more regularly and passively. By bringing media collections into everyday life activities, we were aiming to facilitate annotation, storytelling and the creation of links between media objects.

THE MEMORYNET VIEWER
The MemoryNet Viewer (MNV) (see Figure 1) is a peer to peer (p2p) system for sharing and annotating media objects among people in our personal networks. Recognizing that people naturally associate their own

media with that of friends and family, we developed a prototype to investigate how these connections might be reflected in a software infrastructure. The MNV was built as the first step in exploring this system of interconnected media.

In building the initial prototype our goals were to investigate how a lightweight method for showing media might encourage sharing, how stories might evolve over time and be told as part of the process of playing related media, and how we might begin to associate metadata with media over time.

We implemented an application that allowed for media to play continuously and for annotations to be added to the media with either voice or text. These annotations allowed for metadata to be added to the images over time.

Playing Media

The MNV (see Figure 1) is currently a lightweight desktop application that retrieves media from the collections of the people on a user's 'buddy list' and displays the photos and videos in the form of a continuous slideshow. It displays the objects in random order, choosing equal numbers of objects from each person's collection. The media objects reside on the owners' computer and are connected to that of another person via the p2p network.

The interface resembles that of an instant messaging application with a buddy list to the left, the media display window to the right of that, and speech bubbles on the right most side displaying annotations.

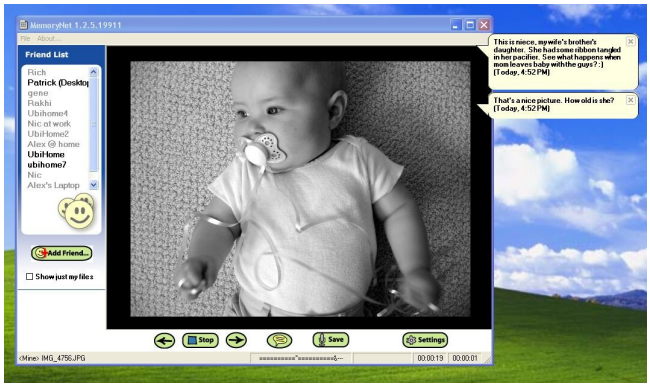

Figure 1: The MemoryNet Viewer

The general use model for the viewer has been to leave it running on a separate desktop monitor with the media rotating continuously. We also kept it running on a plasma display in our 'home lab' and projected the display onto a prominent wall in our general lab space to support the notion of more 'passive' viewing and opportunistic annotation.

Annotating Media

Users add text annotations to media by clicking the appropriate button (similar to Liechti et al's model [7]) while the object is being displayed, and typing in their comments. The next time that media object is displayed the text annotation will appear in a speech bubble. Audio commentaries are added using one of two methods. In the first, the user can press a button to begin recording and another to stop.

In early testing, however, we observed that users were spontaneously commenting, laughing and telling stories, especially when in groups. We soon realized that our initial method of capturing audio did not capture these spontaneous reactions and so the second method of audio annotation involved buffering audio while the media was being displayed. This passive recording option records from the moment the media object is first displayed to when it transitions to the next object. If the user clicks the 'save audio' button at any time while the photo is being displayed, the annotation is saved, otherwise the recording is discarded.

Recognizing that people occasionally regret what they say, or that questions and comments get 'old', annotations are physically stored on the computer of the person who made the annotation. The annotation is requested from the user (by the system) each time its media object is displayed on another machine, and if it has been deleted, will not be played.

While spontaneous audio recordings were viewed favorably upon playback, there are still many experience issues that we need to address such as how to search audio, how to delete blank spaces in the audio clip and what affect a deleted annotation has upon any story thread that is forming.

EVALUATING THE MEMORYNET VIEWER

We carried out structured interviews with users recruited from our labs to evaluate the concept of the viewer as a method for sharing and annotating digital media with ease. Although all users were asked the same set of questions, they were also given the opportunity to comment more generally to express their thoughts and wishes.

The prototype application was in use by seven people for at least 2 weeks (although five out of the seven had been using it for four weeks) prior to the interviews. The major constraint of the study was that the system was being used in a work environment. Although this inhibited some of the social uses we had envisioned, it still provided valuable data and actually changed the social atmosphere of the work environment.

Results

General Perceptions

Initial discussions around this concept led us to question what users really wanted to do with their digital photographs [6]. While we do not claim to have answered this question, four of the seven users said this method of displaying, viewing and commenting upon their (and others) photographs *"just made sense"* and gave them a *"way to pull my photos out"*. Users commented upon wanting to *"dump them somewhere"* and have them just be there.

Four of the seven users commented upon how they experienced satisfaction from seeing their own photos being displayed as much as they did other peoples. It brought their media collections out and *"allowed me to see photos I'd forgotten I'd taken"*.

Five of the seven users had the MNV running on a dedicated screen alongside their regular monitor. On an average day, users spent approximately 10-15 minutes *actively* interacting with the viewer. This involved concentrated viewing of the media, reading/adding annotations and uploading more media. However less dedicated time was also spent glancing at pictures *"every now and then"* Although this may not seem like a considerable amount of time it is in fact in line with what we had hoped for. We wanted the viewer to be fairly passive - something that did not require large amounts of interaction time on the part of the user to reap any benefit. Users commented upon how the photographs caught their attention at moments, and they could choose whether or not to pay attention. The presence of the images was somehow satisfying or pleasing in that it made users feel like they had some connection to the other people on the network.

Annotations

The text annotation feature was used more commonly in the work environment. With work colleagues, users commented upon how they were more conservative in the annotations they offered so as not to offend out of ignorance of details. All users mentioned that auto capturing of *"the 'aha' experience*" from distant family members would be something that they would look forward to playing back.

The most common reasons for adding annotations included:

1. To make a funny comment.
2. As a response to someone else's annotations.
3. Adding some knowledge (either to your own or someone else's images.
4. To ask questions about content/context.

When people were alone they usually added text annotations rather than audio annotations to the media, mainly because they felt 'stupid' talking to their computer. In social situations however, when a number of people were present in a room and paying attention to the viewer, more buffered audio annotations were saved.

Participants could see, how in a home setting, audio annotations might be more appropriate and entertaining, providing extra value when sharing annotations with family and friends. Users *"want to hear their reactions to your photos and the events in your life"* especially when it's from people they have an emotional closeness to.

Nevertheless there were significant issues with audio annotations. Users have no way of knowing what might be contained in them. With text, you can glance at the words and get a flavor for what is there but with audio this was not possible. Users wanted to navigate the audio to get to the bits that were interesting or skip the long pause at the beginning of some audio annotations. Although the value of audio was clear, the method of implementing both capture and playback most definitely needs work.

In building the prototype we had predicted some of these issues and in the future will be incorporating work from other members of our lab which looks that how one might index audio, and translate short comments to text, allowing it to be navigable.

Conversations

All seven users noted how the application had led to in-person conversations. This was most enjoyable for people when they were co-located, saw an image that interested them and shouted a comment over the wall. But, people also reported the application sparking conversations away from the viewer itself. Five out of seven users noted how they had seen something that they later talked about with the either the owner of the photo or one of the other people on their buddy list. These moments were common and users mentioned liking the fact that it *"sparked new and interesting conversations"*.

Feeling more connected & discovering new links

All users noted that they had learned something about someone that they hadn't known before. The photos and annotations led them to change their opinions of people and their interests, with one user commenting *"I didn't realize X was so daring"*. It helped them learn about peoples families, 'meet' their spouses and kids and see their houses (for example) – it essentially gave people a window to a less formal part of peoples' lives and all users valued that immensely. As participants put up vacation photos, more stories were told. Six out of seven users mentioned joy or happiness as an emotion sparked by the photos – *"it made me smile"*. There was also an

element of feeling an affinity to people as users realized they had similar interests, and also some envy at how much other people travel! There was also some evidence of curiosity, poignancy and the sparking of personal memories.

Wanting More

The interviews highlighted the good and the bad both in terms of functionality and the interface. The interaction model was compelling in its simplicity, yet people wanted more. We deliberately began with little functionality and the 'more' came from the users. Specifically, users wanted:

1. The ability to express connections to their own media eg *"I've been to the same spot"*.
2. The ability to see related media, or media from the same collection.
3. The ability to have live conversations with people with the media objects being present.
4. The ability to control the media they see by person, by theme etc
5. The ability to navigate audio annotations
6. The ability to create groups of people based upon interest, commonalities etc

Again some of these were features we had previously discussed but had decided not to implement at that time.

The Value

The value for people was in sharing their everyday lives with their close friends and family and in providing a constant connection to people, being able to access other people's archives of photos, and just having a window into the world of others.

FUTURE WORK

From a technical standpoint, we are working to grow MemoryNet into a secure infrastructure that supports the management of links between media such that we create a system that strengthens personal networks and improves the preservation of memories by representing these emotional connections in software. We see home archival appliances that store and manage the connections among our media. We are developing user interfaces that display dynamic clusters of media based not just on time but on a phrase, a name, or spoken word and we are developing physical interfaces and interaction models to navigate and annotate the media.

We also recognize that as technology progresses, we retire old machines yet we want to keep the annotations associated with the media that we store because the stories told and comments made are more valuable over time. This is an issue that we still need to address.

We are in the process of developing the infrastructure and appliances for deployments with networks of friends and family. One likely future deployment will include connecting family members with elderly relatives in care facilities. We will use these deployments to investigate how connected media in personal networks might augment personal connections.

Finally, our goal is to research interaction models that promote the ability to express an emotional connection to media by facilitating effortless annotation and storytelling.

CONCLUSIONS

We have described the MemoryNet Viewer and our initial studies to evaluate the concepts and the application. We acknowledge that only a limited amount of data can be collected from seven users but initial results are encouraging. Given this we are pursuing these ideas with further developments to the viewer and other prototypes. We acknowledge that there are significant issues with some of the features and methods used and are working to further develop these and solve some of the user experience problems identified here.

ACKNOWLEDGMENTS

The authors wish to thank Gene Becker and all members of the Contextual Media Systems Group for comments and contributions to the work.

REFERENCES

1. Ofoto. http://www.ofoto.com
2. Shutterfly. http://www.shutterfly.com
3. Hello by Picassa. http://www.hello.com
4. Instapix. http://www.instapix.com
5. Chalfen,R., (1987), Snapshot Versions of Life, Bowling Green State University Popular Press, Bowling Green, Ohio
6. Frohlich,D., Kuchinsky,A., Pering,C., Don,D., Ariss,S., (2002), Requirements for Photoware, In the Proceedings of CSCW'02
7. Liechti,O., Ichikawa,T., (1999), A Digital Photography Framework Supporting Social Interaction and Affective Awareness, In the Proceedings HUC'99
8. Pilu,M., (2000), PicShare, A Capture Centric Remote Collaboration Tool, HPL Tech Report HPL-2000-14