Toward an Understanding of Context-Awareness and Collaborative Narratives in Mobile Video Creation

by

Nels Christian Anderson

B.Sc., University of Colorado, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Computer Science)

The University of British Columbia

August 2007

© Nels Christian Anderson 2007
Abstract

As computing technology and multimedia become increasingly intertwined, there is growing evidence of a shift in the way users engage with multimedia. Computing technology is facilitating a transformation—media consumers are becoming media producers. This democratization of creativity, which began with blogging, has now moved into the realm of visual multimedia. One of the major agents of this change is the programmable smart phone. Its portability allows for spontaneous, serendipitous and simple media capture in ways not previously possible. These mobile phones host a number of media capture devices along with their communications technologies, including still and video cameras. However, most of these are merely miniaturizations of traditional media capture devices, and thus prevent more sophisticated interactions. Leveraging the paradigms of context-aware computing, we can build systems that allow users to interact with media capture in entirely new and useful ways.

We have designed a system called Scheherazade that facilitates context-aware capture of video via mobile phones. This system also allows for these video clips to be combined with clips provided by other users, creating collaborative video narratives. We conducted a user study using this system, to determine the usefulness of different types of context. In performing this analysis, we have discovered contextual information that is gathered through user interaction and contextual information gathered automatically are not equally useful when engaging with video captured on smart phones. We discuss design considerations for utilizing manual or automatic context and directions for future work in this area. The research presented here is an initial exploration into how context-awareness can enhance media capture and sharing on mobile devices.
Table of Contents

Abstract .................................................................................................................. ii
Table of Contents .................................................................................................. iii
List of Tables ........................................................................................................ iv
List of Figures ......................................................................................................... v
Acknowledgements ................................................................................................. vi
1 Introduction ......................................................................................................... 1
2 Related Work ....................................................................................................... 4
  2.1 Definition of Context-Awareness .................................................................. 4
  2.2 Five Types of Multimedia Tasks ................................................................. 7
  2.3 Categorizing Context-Aware Video Applications ....................................... 9
  2.4 Context-Awareness and Multimedia Tasks ............................................... 12
  2.5 Relationship to Lightweight Videowork ..................................................... 14
  2.6 Manhattan Story Mashup and the Design of Scheherazade .................. 15
3 Scheherazade Design and Architecture ......................................................... 17
  3.1 Mobile Phone Client .................................................................................... 18
  3.2 Context and Video Management Server .................................................. 21
  3.3 Search and Composition Web User Interface ......................................... 24
  3.4 Summary ..................................................................................................... 28
4 User Study ......................................................................................................... 30
  4.1 Hypotheses ................................................................................................. 30
  4.2 Experimental Design ................................................................................ 31
  4.3 Results ........................................................................................................ 36
  4.4 Discussion ................................................................................................... 43
  4.5 Summary ..................................................................................................... 44
5 Discussion and Future Work ............................................................................. 46
  5.1 Considerations for Utilizing Manual Context ........................................... 46
  5.2 Considerations for Utilizing Automatic Context ....................................... 51
  5.3 Future Work ................................................................................................ 55
6 Conclusions ....................................................................................................... 58
Bibliography .......................................................................................................... 59
Appendices ............................................................................................................. 63
  Appendix A – Study Materials ......................................................................... 63
  Appendix B – UBC Research Ethics Board Certification ............................. 72
List of Tables

Table 2.1 - Systems Categorized by Context .............................................. 10
Table 2.2 - Systems Categorized by Multimedia Task .................................. 13
Table 4.1 - ANOVA of Creation Time .......................................................... 39
List of Figures

Figure 3.1 - Scheherazade System Architecture ................................................................. 18
Figure 3.2 - Scheherazade Client Architecture .................................................................. 19
Figure 3.3 - Scheherazade Tagging Ontology ..................................................................... 21
Figure 3.4 - Scheherazade Server Architecture .................................................................. 22
Figure 3.5 - Scheherazade Database Schema ..................................................................... 23
Figure 3.6 - Scheherazade UI Architecture ....................................................................... 25
Figure 3.7 - Screenshots of Composition User Interface .................................................... 27
Figure 4.1 - Creation Time for Participants Using Automatic Context First ...................... 38
Figure 4.2 - Creation Time for Participants Using Manual Context First .......................... 38
Figure 4.3 - Creation Time for Participants Completing Study without One Week Delay ... 40
Figure 4.4 - Creation Time for Participants Completing Study with One Week Delay .......... 40
Figure 4.5 - Participant Measure of Usefulness of Context for Personal Media .................. 41
Figure 4.6 - Participant Measure of Usefulness of Context for Others' Media .................... 42
Acknowledgements

I must first thank my exemplary supervisors Dr. Buck Krasic and Dr. Rodger Lea. Even though this work ended up a bit outside of their traditional research domains, they provided continual guidance and tremendously valuable advice. I am also grateful for my colleagues in MAGIC, including Mike Blackstock, Matthias Finke, Nicole Arksey and all the rest. Plato knew that our knowledge is little compared to that of which we are ignorant; thank you all for helping me close that gap a little more.

I also offer boundless appreciation to the other Computer Science grads. They indulged me far too much, as I press-ganged them into countless seminars and workshops that they weren’t really interested in. Being part of this vivacious community kept me sane and smiling throughout this entire process.

My parents, Ernie and Cindy, have never once failed to provide me with succor and encouragement. Because of them, I have never doubted my ability to accomplish what I set my mind to. Thank you both very much. Finally, I want to thank Tila, for all of her caring and support. Without you, this endeavor would have been substantially more difficult, and much less exciting.

Nels Anderson

University of British Columbia, August 2007
Chapter 1

Introduction

As computing technology and multimedia become increasingly intertwined, there is growing evidence of a shift in the way users engage with multimedia. Computing technology is facilitating a transformation—media consumers are becoming media producers. This democratization of creativity, which began with blogging, has now moved into the realm of visual multimedia. One of the major agents of this change is the programmable smart phone. Its portability allows for spontaneous, serendipitous and simple media capture in ways not previously possible. These mobile phones host a number of media capture devices along with their communications technologies, including still and video cameras. However, most of these are merely miniaturizations of traditional media capture devices, and thus prevent more sophisticated interactions. Leveraging the paradigms of context-aware computing, we can build systems that allow users to interact with media capture in entirely new and useful ways.

Very little work has been done to investigate how context-awareness can enhance mobile media creation, especially video capture. Davis et al. have investigated context-awareness in terms of mobile photography [22, 23]. This work has culminated in a publicly available beta application called ZoneTag [1]. ZoneTag provides users with context-aware tag suggestions for photos they have taken via their mobile phone. These tagged photos can then be uploaded directly to Flickr.

Manhattan Story Mashup [27] is another mobile photography application, but it is distinct in that is focuses primarily on making media available for reuse. Conducted as a game event, users are given specific keywords as subjects to capture with their mobile
phone. These photos are uploaded to the game’s server, where it is associated with the original search term. These keywords-picture pairs serve to tell a media story, but they are also available for use in other media stories. This allows for the photos to be repurposed and used in creating “mashups” of new and old media.

MobiCon [16], from the VTT Research Institute, does integrate context-awareness with mobile video capture. After video is captured using this system, the system provides the ability to label the video using a variety of predefined concepts (such as “vacation” and “buildings”) as well as manually entering other keywords. Other information is automatically inferred, including the user’s name, current region and country, date and time when the clip was recorded, length of the clip and GPS information if a GPS device is available.

These systems, and similar systems in this area, have focused their analysis on design features and architectural details. However, there has been a lack of analysis in terms of user interaction with the contextual information. In some cases, it is not even clear how users interact with certain types of context information at all. This makes it difficult to understand how useful users perceive various types of contextual information to be, and how this context may improve or hinder multimedia tasks users wish to accomplish.

In our investigation, which this thesis details, we place a heavy emphasis on qualitative and quantitative analysis of usefulness of different types of contextual information. We believe this can better inform future designers of context-awareness media applications. To accomplish this, we have designed a system that draws upon aspects of all three of these systems, as well as other system that make use of context-
awareness and video. We have called this system Scheherazade and it facilitates context-aware capture of video via mobile phones. This system also allows for these video clips to be combined with video clips provided by other users, creating a collaborative video narrative. We conducted a user study using this system, to determine the usefulness of different types of context. In performing this analysis, we have discovered contextual information that is gathered through user interaction and contextual information gathered automatically are not equally useful when engaging with video captured on mobile devices.

This thesis describes the design, implementation and analysis of user interaction with Scheherazade. Previous work relevant to this investigation is presented in Chapter 2. Therein, we analyze prior context-aware video systems and discuss our categorization of them, which produces two distinct categories of applications. Chapter 3 presents the design and architecture of Scheherazade. Scheherazade consists of three separate components- a client on the mobile phone, a server which manages captured videos and their context and a web-based system that allows for searching for videos by their contextual information and combining them into larger narratives. In Chapter 4, we discuss the user study we conducted to examine the usefulness of different types of context in Scheherazade. In this section, we conclude that manual context is qualitatively and quantitatively more useful than automatically ascertained context. Chapter 5 discusses the implications of this study. Here we present design considerations for those seeking to make use of manual and automatic context-awareness in media capture systems. Chapter 6 provides a summary and concludes this thesis.
Chapter 2

Related Work

In this chapter, we discuss prior research relevant to our investigation. We begin by briefly discussing both facets of this work separately. First, we provide a description of context awareness and then a discussion regarding digital components of video creation and use. We then examine systems that incorporate both video and context-awareness, before focusing more closely on context-aware media capture systems. By examining how these different systems make use of context, we were able to construct a taxonomy of context-aware media applications that coincides with standard practices of media capture. Finally, we conclude by describing a system closely related to Scheherazade, and what it contributed to our design and implementation.

2.1 Definition of Context-Awareness

In the past four decades, computing technology has evolved from massive, monolithic objects that not only required specially-trained operators to use, but specifically-constructed buildings to house, to portable pocket-sized objects used by people from all walks of life to facilitate socializing, entertainment and other non-work tasks. These systems are not only interesting in what they do, but how they do it, where it is done, and so forth. The idea that the context in which these systems are used can drive adaptation, even core functionality, of these systems has become very relevant. Context-aware computing researchers examine these ideas and try to determine in which ways context can improve computing technology.
This research was pioneered by Mark Weiser [29, 30] in 1991 under the vision of ubiquitous computing, or ubicomp. Ubicomp endeavors to create an environment where a variety of computers are distributed and continually available, rather than as distinct objects. This served as a foundation for context-aware computing, which refers to systems that are aware of the circumstances in which they operate and can adjust their functionality accordingly. The first tangible research project in this area was the Active Badge [28] system from Xerox's PARC in 1992, but the term “context-aware” was not seen until 1994 [25].

One of the major challenges in context-aware computing is identifying what context truly represents. Many descriptions of context only define it by enumerating examples, or by providing synonyms. But saying that context is the “environment” of an application does little to provide a functional definition. This issue hearkens back to the earliest work in this area and arguably is still manifest in contemporary research. Without a solid grasp of what context means, it can be very difficult to understand and discuss a context-aware system.

When the term context-aware was first introduced, it was defined by example. Simply, context meant location, nearby people and objects and changes to them [25]. Other researchers expanded this list, adding time of day, season, temperature, focus of attention and even the user's emotional state [5, 11, 23]. A more categorical definition was also provided [5], which identified three categories for context: computing environment, user environment and physical environment. However, these environments are themselves defined by listing examples. The common weakness of all these definitions is that they are either imprecise or rigid.
Anind Dey recognized this and presented a flexible and encompassing definition of context [12]. He posited that definitions like those above were too specific. He notices that these definitions attempt to identify which aspects of a situation are contextually relevant, but this is generally unattainable, as the important aspects of context changes from situation to situation. He offers the following definition as an alternative:

"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."

The strength of this definition is that it allows those building a context-aware system to enumerate what information is relevant context in their system. Location, one of the most commonly identified piece of context information, is no longer necessarily context for any context-aware application, as it would be using the previous definitions. However, there is still a simple categorization of context. Location, identity, time and activity are classified as primary context, which all other context information is classified as secondary. The relationship here is any piece of secondary information can be ascertained from primary information. For example, a user's identity can serve as an index into some information space to find their phone number or email address. Dey's work was well received by the research community and has become a canonical reference for those seeking an understanding of context-awareness.

Dey's definition is especially important when examining context-awareness and video. Unlike many applications, which deliver new functionality using context
information, video is already a well-established medium. Rather than delivering something new, context-aware video must find ways in which video can be enhanced using context-awareness. Next we discuss our categorization of the five different aspects of digital video creation, to provide a better understanding of where context can be leveraged to improve interaction with video and software.

2.2 Five Types of Multimedia Tasks

As computing technology has evolved, providing support for digital multimedia has becoming increasingly important to those users of computing systems. Making proper use of this digital multimedia, rather than just digitizing existing media paradigms, is rife with both challenge and opportunity. But to better understand these challenges, we need to understand how software fits into the process of media creation and consumption. Digital multimedia tasks belong to one of five categories: capture, authoring, distribution, media management and presentation.

Capture refers to the devices and software that actually records multimedia artifacts. This includes devices like digital cameras and camcorders, digital audio recorders, webcams, and smart cell phones that have the above devices integrated into them. With the exception of smart phones and some webcams, these capture devices are generally single-purpose and provide little or no functionality beyond basic media capture.

Authoring software are the tool suites that facilitate transforming raw video into a final multimedia object via editing and adding special effects, which can be quite significant in some media. This would also include the software systems that manage
metadata or adding annotations. Examples of this include software suites like Adobe Premiere [35], Apple's Final Cut [36] and iMovie [37]. Audio-only software like Pro Tools [38] would also be a part of this group.

Aside from physical media formats, such as DVDs, software systems are the primary form of distribution for most multimedia. Especially in the last few years, Internet multimedia has become familiar to many more people. Online video distribution systems are quite popular, with YouTube alone serving over 100 million videos every day [34]. With Internet-focused distribution as a tangible possibility, much more is feasible than is with media artifacts. Digital distribution, especially streaming systems, allows for delivery of dynamic, user-centric content.

Media management systems are responsible for sorting and storing large amounts of media in an organized and coherent fashion. Generally designed toward retrieval and search, this software is represented by Personal Video Recorder (PVR) systems, such as TiVo [39], as well as more general-purpose media systems, like Windows Media Center Edition [40], MythTV [41] and Apple's FrontRow [42]. Audio-centric systems like iTunes [43] and Windows Media Player [44] could also be considered to be part of this category.

The ultimate goal of media creation is to craft something and present it to an audience. To facilitate this, there are the single-purpose devices, such as televisions, radios and movie theatre projection systems. These devices are extremely familiar to almost everyone who consumes media, but also largely devoid of innovation. Since their invention, these devices have not changed in any fundamental way. While quality, size and appearance have changed, movie theatre patrons still sit in a darkened room watching
a large screen as was done 80 years ago. Software media players, such as those on personal computers, mobile devices and smart phones, on the other hand, are a more fertile ground for experimentation and innovation.

Now that we possess an understanding of the five different digital multimedia tasks, we discuss a categorization of systems that make use of context-awareness and media.

2.3 Categorizing Context-Aware Video Applications

One major goal of this analysis is to determine in which ways, if any, context-aware video systems differ from other context-aware applications. Also, identifying categories of context-aware video systems would be useful in analyzing and building future context-aware video applications. We examined a host of context-aware video applications to determine if any patterns or taxonomy in the various video systems could be discerned. If we could discover in what ways the various systems were similar in the techniques and methodologies they used to address their challenges, it would be possible to leverage this information when creating a context-aware video system of our own.

Dey presented a survey of context-aware systems [24] that has now become a canonical reference for those seeking to understand the various qualities and features of context-aware computing. Endeavoring to do the same with a specific subset of systems, we used Dey's taxonomy and applied it to a diverse group of context-aware video applications [1, 3, 6-9, 10, 13, 16-22, 26, 27, 31, 33]. From this, we have produced the classification seen below in Table 2.1
Table 2.1 - Systems Categorized by Context

<table>
<thead>
<tr>
<th>System Name</th>
<th>Description</th>
<th>Context Type</th>
<th>Context Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>I</td>
</tr>
<tr>
<td>Ubiquitous Media</td>
<td>Office tele-presence</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Abaris</td>
<td>Therapy recording</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Home Media Space</td>
<td>Smart telecommuting</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MobiCon</td>
<td>Phone-based video capture and tagging</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mobile Cinema</td>
<td>Location-aware film</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Media Adaptation</td>
<td>Extensible media adaptation framework</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Framework</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vox Populi</td>
<td>Bias rhetoric generation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RealityFlythrough</td>
<td>Multi-source, 1st person video coordination</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Smart Media Player</td>
<td>Context-adaptive player</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MMM/MMM2/ZT</td>
<td>Phone-based photo tags</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Replayer</td>
<td>Context-aware video and system log sorting</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AMV Tagging Survey</td>
<td>Tagging video at the scene level</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TEEVE</td>
<td>3D tele-presence</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Manhattan Story</td>
<td>Mobile photography and storytelling mashups</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Examining this classification, we see no particular types of context are notably dominant or absent. While the primary context of identity is present in the majority of the systems, it only dominates by a small margin. In the systems we surveyed, activity was used as often as location and time context, unlike Dey's survey. Many of the context-aware video systems we examined were novel because they integrated non-video information with captured content, such as the digital pen input from Abaris [13] and user
bias in Vox Populi [3]. We believe that activity is more commonly utilized in video-centric systems and supporting it should be a design consideration.

It is also important to examine what types of actions the systems take based on the context they sense. These actions are classified as the **presentation** of information and services to a user, automatic **execution** of a service, and the **tagging** of context to information for later retrieval. We discovered that among videos systems, the actions the system supports also dictates how it affects the media the system uses. Those systems that support presentation or execution rarely create or alter existing media artifacts. Systems that support tagging, on the other hand, emphasize creating new media artifacts and metadata for those objects, but have little to do with how that media is used.

We group those context-aware video systems that support presentation or execution actions together, calling them **presentation-focused**. Presentation-focused applications emphasize using context to drive playback functionality. They use context, especially location, to present relevant media or to enhance tele-presence systems. These systems largely resemble other context-aware system surveyed by Dey. The general architecture and functionality is similar, except in these systems, the primary action is playing video or modifying a live video stream instead of executing or modifying some application behavior. One such system is Mobile Cinema [8]. Crow et al. believe using location will add another dimension to a media experience. Instead of having the same experience for all viewers of a certain video, Mobile Cinema endeavors to create dynamic, customized video sequences that change with a user's location. Another such system is the Home Media Space [21], which uses information like a family member entering a tele-commuter's home office to execute actions, in this case shutting off the
web-cam for the family member's privacy until they have departed. Other systems of this type are Ubiquitous Media, Media Adaptation Framework, Smart Media Player and TEEVE.

The other class of context-aware video applications, which we call creation-focused, consists of those systems that support a tagging action. These applications focus on synthesizing information acquired through context-awareness with video to create better metadata for captured video. In creation-focused applications, the end result of the system (that is, the produced media artifact) is often not context-aware at all. MobiCon [16], for example, uses context to annotate videos that can later be viewed by anyone with a media player. The context-awareness in MobiCon is used to create better metadata for captured video, to improve archival search and sharing. Other systems we categorized as creation-focused are Abaris, Vox Populi, RealityFlythrough, MMM/MM2, ZoneTag, Replayer and the AMV Tagging survey.

By conducting this analysis of video and context-awareness, we have discovered two distinct categories of systems making use of video context-awareness. In the next section, a similar trend is discovered when examining context-awareness video applications in regard to the five multimedia tasks we discussed above.

### 2.4 Context-Awareness and Multimedia Tasks

In the previous section we categorized the systems we surveyed by their various context-aware features. We felt it was also important to examine and categorize these systems according to what areas of videowork they improve upon. As we discussed above in section 2.2, there are five areas where software engages with digital media
creation: capture, authoring, distribution, management and playback. Below, in Table 2.2, we have categorized each of the systems surveyed above by how they interact with the various multimedia tasks.

<table>
<thead>
<tr>
<th>System Name</th>
<th>Capture</th>
<th>Authoring</th>
<th>Distribution</th>
<th>Management</th>
<th>Playback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitous Media</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abaris</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home Media Space</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MobiCon</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Mobile Cinema</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media Adaptation Framework</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Vox Populi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RealityFlythrough</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart Media Player</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>MMM/MMM2/ZT</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Replayer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>AMV Tagging Survey</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEEVE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Manhattan Mashup</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When examining this table in coordination with the presentation-creation distinction drawn above, we can see a correlation between type of context-aware applications and the videowork aspects they modify. Distribution of media based on context is functionality shared by both types of applications. However, the authoring and management aspects are modified by creation-focused applications, while presentation-focused applications modify playback. Capture is shared by the two types of applications, with a bias toward presentation-focused applications. This implies that
certain types of context are better suited for augmenting certain aspects of digital multimedia software. This information can be quite valuable in designing applications. In knowing what kind of application is being designed, the types of contexts that will be most valuable for its usage can be emphasized, instead of trying to support a myriad of different contexts that might not be useful.

2.5 Relationship to Lightweight Videowork

Kirk et al. presented a survey and analysis of user video capture practices [15]. They describe two distinct categories of videowork- lightweight and heavyweight. Heavyweight videowork is the more traditional "home movies" videowork, characterized by use of single-purpose capture devices, deliberate and planned capturing of events, use of editing software and transferring the media to a physical artifact (e.g. DVD). Contrarily, lightweight videowork is spontaneous, unedited, captured by mobile phones, and is shared either on the captured device or via the Internet.

Comparing this to the taxonomies we presented above, we can see a very similar trend. Creation-focused video systems have many of the same features of lightweight videowork. Similarly, presentation-focused video systems align closely with heavyweight videowork. This similarity provides a useful foundation for building systems that want to explore the features of lightweight videowork and creation-focused context-aware systems.

We decided to narrow our focus and examine the use of context in creation-focused systems. This distinction between lightweight and heavyweight systems provided us with several useful insights. First, it was important to target mobile phones
as the primary capture and authoring devices, as lightweight videowork occurs almost exclusively on these devices. Second, our system should not assume users will engage in traditional video editing, as occurs in heavyweight videowork. Finally, use of the Internet is the prima facie mechanism for sharing this kind of media. We incorporated all three of these observations into our design for Scheherazade.

2.6 Manhattan Story Mashup and the Design of Scheherazade

One system we surveyed that deserves emphasis is an event Nokia organized called Manhattan Story Mashup [27]. Using N80 smartphones, players were presented with keywords from stories created by authors on the Manhattan Mashup website. Their task was to capture an image that clearly represented the keyword they were dispatched. Once all the keywords in a story were photographed, the story was shown on the Reuters Sign or other large public displays in Times Square. The keyword/image pairs were also made available for other authors to use, creating new narratives from existing content (hence the “mashup”). Over 250 players participated in this event and it was considered to be a strong success, both from Nokia and the players' perspective.

Our design goal for Scheherazade was to take the basic format of Manhattan Mashup, extend system support to video and add types of context used in the creation-focused systems we surveyed above. One striking feature of many of the systems we surveyed was a lack of qualitative and quantitative analysis of user interaction with the system. With Scheherazade, we wanted to study usefulness of different types of context available for use. Thus, we designed a user study to evaluate this, presented in Chapter 4.
In this chapter, we have provided a taxonomy of context-aware video systems and discovered two different categories of context-aware video systems. This categorization was also apparent when examining these context-aware systems in terms of what multimedia tasks they augment. The distinction between lightweight and heavyweight videowork cemented the features our systems should support. In the next chapter, we provide a detailed description of the system architecture of Scheherazade.
Chapter 3  

Scheherazade Design and Architecture  

Having examined research involving context-awareness, multimedia and collaborative narratives, we desired to extend this area of investigation to video capture on mobile devices. With little prior work relating specifically to video capture, we looked to similar work using mobile phone photography as a foundation for our investigation. By creating a study similar to those that have been conducted previously with mobile photography, we can examine the similarities and differences between these two types of context-augmented mobile media capture.

To conduct such a study, we needed to design capture software for the phone to enable acquisition of contextual information, as well as a framework for managing the capture media and context. We decided to call this system Scheherazade, after the storyteller heroine of The Book of 1001 Nights. Scheherazade consists of three distinct components: the mobile phone client, which associates context information with video clips, the back-end server which manages videos and their context once uploaded from the phone, and the web UI which enables searching for video clips using their contextual information and compiling them into larger narrative sequences. Each of these components is described in detail below. The overall architecture of Scheherazade can be seen in Figure 3.1.
3.1 Mobile Phone Client

Examining prior works, especially MobiCon [16] and MMM [9], we designed Scheherazade’s phone-based software with three goals in mind. First, capturing media should resemble the existing processes on the phone as closely as possible. Second, the system should allow for quick time of capture annotation of videos. Third, based on user reactions to MobiCon [16], we wanted to provide offline capture and archiving of captured media. Users of MobiCon reported that uploading captured videos was frequently sluggish and the wait prevented them from capturing other videos. GPRS is generally quite slow, especially when dealing with large artifacts like video, and 802.11 wi-fi connectivity is not always available where the user may choose to go. By allowing for videos and their contextual information to be archived for later uploading, a user’s ability to capture media is not artificially impaired by their network connectivity.
The mobile client for Scheherazade was developed for Nokia's Series60 platform using the Series60 port of the Python programming language. We selected this platform due to the availability of several N80 smart phones in our laboratory. These phones afford both a three megapixel video camera and wi-fi networking. The client software handles acquisition of most of the relevant context for the video. A representation of this is shown in Figure 3.2. To meet our first design goal, instead of writing new video capture software as part of the Scheherazade client, we used S60 Python to launch native applications. Using the video camera software provided by Nokia allowed us to provide
access to the sophisticated video capture system, familiar to those who have used similar software on other devices. Being a native application developed in Symbian C++, it was also higher performance than a similar application developed in S60 Python would be. We did consider drawbacks to this approach, most notably that the transition between the native video software and the Scheherazade client would be jarring and slow. However, initial pilot testing confirmed that the transition wasn’t excessive and speed was sufficient for the required tasks. This satisfied our first design goal of making video capture on the phone resemble the standard practices as much as possible.

Associating context with the videos captured was the primary purpose of the Scheherazade client software. The contextual information we wanted to acquire was based on our survey of context-aware video systems, as detailed in chapter 2. The client software automatically captures the following contextual information: date and time of media capture, location via a Bluetooth GPS unit and username of the media author. Additionally, after capturing a video, the Scheherazade software supplies the user with a simple menu-based dialog for associating simple textual tags with the media they just captured. We used a simple ontology of tags, seen in Figure 3.3, with the top level being a simple binary choice between the clip being shot inside or outside (analogous to the “internal/external” scene distinction in film treatments). This choice determines the content of the next series of menus. The user is given a choice between four over-arching categories, each of which include four subentries, or typing in their own tags on the phone’s keypad. After selecting or typing a tag, the user can enter more tags if they desire. By providing this tagging interface at the time of capture, we achieved the second of our design goals.
Once all desired tags have been entered, the user can either upload the video directly from the phone, or store the video and all its contextual information to an archive for later uploading. The contextual information is stored in a textual format, making the memory storage requirements minor when compared to the size of the captured videos. The user was also able to select an “automatic archive” option, making all media captured archived by default. It is then possible to upload all the videos and their context in batch, likely once the user has reached an area of reliable 802.11 wi-fi connectivity. This satisfied our third design goal, segregating the capture process from the upload process. In chapter 4, we will discuss user experiences with the client software and lessons we learned.

3.2 Context and Video Management Server

The context and video management aspect of Scheherazade are handled by a JBoss server and MySQL database. JBoss and MySQL were selected due to proven success in building similar context and media management systems rapidly and easily. Our requirements when designing this aspect of the system was focused primarily on two things: providing a simple uploading procedure and allowing stored videos to be accessed
by queries on their contextual information. The emphasis was on simplicity and designing for mobile interaction from the outset. Figure 3.4 provides a diagram of this component's architecture.

Uploading took place using HTTP POST and simple TCP sockets. The video file itself was transferred via TCP to the Scheherazade webserver. Each video was transcoded using ffmpeg from the phone's default .mp4 file format to the Flash video format (.flv) required by the Flex UI detailed below in section 3.3. The first frame of the video was also copied as a .jpg to use a thumbnail when the video was manipulated via the Flex UI. All three of these files were placed on a world accessible location on the webserver. Perhaps most importantly, all the uploading and transcoding could be
accomplished by a single HTTP POST request, in our case from the Scheherazade’s S60 Python client. This made the uploading process simple, transparent to the user and possible using only the phone’s 802.11 wi-fi connection. Architecting the system in this way made uploading simple and possible from a single key-press on the mobile phone, meeting our first design goal. Contextual information about a video was uploaded to the
Scheherazade server via HTTP POST, where various servlets inserted this information in the MySQL database. The database schema can be seen in Figure 3.5.

Some additional context acquisition occurred once the video was uploaded, namely the association of weather with the videos. As the contextual information from the phone included data, time and location (from GPS), Scheherazade’s webserver then contacted Yahoo’s weather service to determine the conditions for that place and time. This information was also added to the database. An uploaded video served as the primary point of reference, meaning that queries on any of the various categories of contextual information would ultimately resolve to a video entry. This means videos can be retrieved by their contextual information, our second design goal.

3.3 Search and Composition Web User Interface

The third component of the Scheherazade architecture needed to provide users with the ability to search for video by contextual information, watch those videos and add them to a dynamic timeline to create a larger video narrative. It was important that this component be web-based, but also provide a dynamic interaction environment. We decided use Adobe’s Flex framework for development, preferring it over an AJAX environment. Flex provides a Flash-based environment for developing interactive, asynchronous web content. Being based on Flash, interaction with video is robustly supported (although this did require the .flv transcoding mentioned above). Figure 3.6 shows the architecture of this component and Figure 3.7 shows screenshots of the UI. Scheherazade’s composition interface consists of three distinct elements. The first is the search pane in the upper-left, which allows for users to search for videos using contextual
information associated with those videos. That information consisted of: the filename of
the video, name of the video's author, tags associated with the video, weather conditions,
date (using simple "before," "between" and "after" delimiters) and location of where the
video was captured. This provides the search functionality required of this interface.

The location search term requires further explanation. First, users could simply
search for videos captured either "inside" or "outside." As mentioned above, this is
analogous to the "External" and "Internal" scene designations in traditional
cinematography. Alternatively, users could use the GPS information associated with
videos for a more specific search. However, as specific numerical GPS coordinates were
unknown and functionally unintelligible to users, we instead associated sets of GPS
coordinates with certain semantic locations around UBC campus, e.g. "Forest Sciences
Centre" or "Student Union Building." The GPS location of each video was categorized
as one of these semantic locations, or simply "unknown" if the video's coordinates were
outside of
all semantic locations. It was also possible to bypass semantic locating and enter specific
ranges of GPS coordinates.

Upon performing a query, search results were displayed in the pane in the upper-right. The video itself was displayed, along with standard video player controls. The context of the video was also presented, in text format next to the video itself. The display of contextual information was not largely interactive; the only exception is that clicking on the GPS coordinates would launch a Google Map displaying that location. Viewing the result videos and their context is straight-forward, satisfying our second design requirement.
The lowest pane on this interface was the timeline used for creating longer narratives from the shorter clips. Users could drag and drop clips from the results pane into the timeline pane, or drag them inside this pane to alter their order. This would insert the thumbnail of the video, the first frame mentioned above, into the pane. Clips could also be removed from the timeline by highlighting and pressing delete on the keyboard. Users could view the timeline in sequence by pressing the “Play Narrative” button below the timeline.

The narrative is represented by a SMIL [45] file generated by the Scheherazade JBoss server. The Synchronized Multimedia Integration Language offers HTML/XML-like syntax for working with multimedia, including video. When the option of playing the narrative is selected, the Flex UI determines which videos are in the timeline and their order, and then passing this information to the Scheherazade JBoss server. A SMIL description is then created for this narrative, which specifies the videos’ URLs on the webserver and their sequence. This SMIL file then is then placed on the webserver, and a reference to it is returned to the web UI. Finally, the web UI launches the URL of the SMIL file, which will open whatever media player is registered for the SMIL MIME type (usually RealPlayer). Again, this is transparent to the user and expeditious, so when the “Play Narrative” button is pressed, the appropriate media player launches almost instantly. Our final design requirement is satisfied by allowing easy creation and viewing of these video narratives.

3.4 Summary
Scheherazade was designed to support our investigation of how users create and utilize media when augmented with contextual information. We needed to develop software components both on the mobile phone, to facilitate context acquisition, and on a central server, to manage the media. An interaction component was also required, so users could search and combine their video clips using the associated contextual information. We designed and developed each of these components, using S60 Python and the phone’s native video camera for the client, a JBoss/MySQL server for media management, and a Flex web interface for searching, viewing and compiling video clips. With these components developed and tested, we were able to conduct our study of mobile video capture and context-awareness, discussed in Chapter 4.
Chapter 4

User Study

To evaluate Scheherazade and validate our hypotheses about the usefulness of different types of context, we conducted a study of Scheherazade usage. From this study, we examine both the quantitative data regarding the usage of contextual information, as well as qualitative responses from participants about their experience with the system. The study consists of two tasks imitating how software like this might be used on a day-to-day basis by users. The first task consists of capturing a number of videos and then annotating them. The second requires users to compile some of these video clips into contiguous narratives, using both clips of their own and clips that other users have captured. Below we detail how the study was conducted and the results from the data we gathered.

4.1 Hypotheses

Our first hypothesis was that participants would find manual context more useful than automatic for locating both their own and others’ video clips. Kirk et al. [14] discuss the lack of success automatic annotation techniques have had in regards to digital photograph collections. The MMM and MMM2 systems developed by Davis et al. [22, 23], found that making greater allowances for time of capture annotation increased both the amount of the media captured and the amount shared. Ames and Naaman [1] came to similar conclusions studying ZoneTag, the public incarnation of the MMM systems.
We also hypothesized that those users who waited a week between capturing the video clips and creating the compositions would find the contextual information more useful. Metaphorically described as the problem of "unlabeled video tapes in a shoebox," we predicted that contextual information would be useful in aiding recollection of captured videos.

4.2 Experimental Design

Our examination of Scheherazade's usage consists of two separate tasks that, when examined in coordination, can allow us to validate or reject the hypotheses outlined above. These tasks, and other details of the experiment, are presented in this section.

Task 1 - Video Capture

The first half of this study required participants to use the Scheherazade client on the smart phone to capture a number of video clips for two different narratives. Participants were instructed on the use of the software and the phone's video camera. Subjects then walked around the UBC campus with an experimenter (in case of software failure or other problems) capturing videos. A rough predetermined path served as a guide, but participants could choose to deviate from this course if they desired. Aside from directing participants and answering questions related to the operation of the software, the experimenter did not interact with the participants. The Scheherazade client was set to automatically archive the videos and their context, so the participant would not have to be concerned with locating and maintaining wi-fi network connectivity.
Participants were asked to capture at least 10 video clips for each of the two narratives, for a total of at least 20 video clips. We requested that clips be about 10 seconds or longer. We asked that the video clips for each narrative be distinct, but participants did not have to film all 10 clips for each narrative contiguously. If participants were unable to think of “themes” for the narratives, it was suggested that they use something along the lines of “student life” and “buildings and construction,” as there are many opportunities to capture video appropriate to these topics around the university campus. Participants were allowed to pursue more “avant-garde” narratives, focusing on themes such as colour or texture rather than a more linear or descriptive narrative.

Beyond capturing video appropriate to their chosen narratives, subjects were not instructed on what types of sequences to capture. Filming campus denizens not involved with the experiment was allowed. However, if the video capture was potentially invasive (i.e. more than would be expected in a common public space), we asked that participants get the accompanying experimenter to explain the nature of the experiment and the confidentiality of the captured media to get the uninvolved person's permission.

Subjects were instructed on how the process of textually annotating videos functioned and that they would later be able to use these tags to search their own media. But they were not told whether future participants would also be able to search and access videos captured by previous participants. Thus, participants did not know if they were simply tagging videos for their own reference, or if their tags would possibly be used by others.
Upon completing this task and returning to the laboratory, the video clips were uploaded to the Scheherazade server via wi-fi and transcoded as described above.

**Task 2 - Narrative Creation**

The second component of this investigation had participants making use of the videos they captured during the first task. The task consisted of creating two longer narrative sequences using both their own video clips and video clips captured by others. We asked that subject use at least five of their own clips for each narrative (of the at least 10 captured for each), and at least three clips from someone else. To avoid bias, we provided the same set of “stock videos” for everyone, but participants were unaware that this was the case. These videos were captured by an experimenter a few days before the study began, using the same software, techniques and time constraints as were used in the study.

As described in Chapter 3, participants used the web UI to search for video clips and compile them into a longer sequence. The types of contextual information that could be searched changed between creating the first and second narrative. To create one narrative, it was only possible to search using the manual context (author and tags) and the other only allowed searching over the automatic contexts (GPS/semantic locations, filename, date, weather). The order of the two different versions was counterbalanced across participants to minimize training effects. Thus, half of the participants used the automatic context version first and the other half used the manual context version first.

Subjects were only able to access the web UI; they did not have direct access to the videos they captured or their contextual information. Tags used to annotate videos
had to be recalled from memory by the participants, as no list of tags used was available. The following semantic locations were available as search terms for the GPS-based location context: Forestry, Tech Ent III (Starbucks), ICICS, Engineering Complex, Biology/Bioinformatics, Student Union Building, Brock Hall, Buchanan, Main Mall and Koerner Library. Participants were cautioned that due to possible unreliability of the GPS tagging, videos might not be associated with the proper semantic location.

Procedure

The experiment was conducted as follows: 1) After consenting to participate in the study, participants were briefed on the nature of the experiment. They were then provided with the experimental equipment and given a tutorial on the functionality of the phone's video camera and the Scheherazade software. 2) Participants were given 50 minutes to travel around UBC campus on foot and capture videos using the mobile phones. 3) Half (8) of the participants were asked to return in approximately one week to complete the experiment, the other half return to the complete the second half of the study just after capturing the video clips. Upon their return, the first portion of the questionnaire was completed, assessing background information and experience with the Scheherazade client. 4) The web UI was used to arrange the two video narratives from the captured clips. 5) After participants had created and viewed their two video sequences, they completed the unfinished section of the questionnaire that dealt with the web UI and searching using the various contexts. An informal interview was conducted with the participants to follow-up on answers to the questionnaire. The questionnaire given to participants is included in Appendix A.
Design

This experiment used a 2 x 2 mixed design. The wait between video capture and narrative was between subjects, while the contextual information available for search (automatic vs. manual) was within subjects. The order of presentation for the contextual information was counterbalanced to minimize possible training effects.

Participants

Sixteen subjects (5 female, 11 male), aged 19 to 40 participated in this study. All but two participants owned a mobile phone. They were provided with a $10 honorarium for their participation. Additionally, the video narratives created by the participants were rated by three volunteers that did not otherwise take part in the study; the creator of the video narrative with the highest overall rating received an additional $10. All participants were recruited by departmental email lists or word of mouth.

Apparatus

The first phase of the experiment was conducted using Nokia N80 smart phones and Holux GPSlim236 Bluetooth GPS receivers. The experimental software was coded in Series60 Python. The details of this software are discussed in Chapter 3.

The second phase of the experiment used a desktop PC running Windows XP with a 3 GHz processor, 1 GB of RAM, an nVidia GeForce 6600 GT video card and a 19 in. monitor configured at 1280 x 1024 pixels resolution. The UI was viewed using the
Mozilla Firefox 2.0 web browser. The web UI was developed using Adobe Flex. Details regarding these components are also available in Chapter 3.

**Duration**

The video capture portion of the study took approximately 50 minutes to complete. While the exact duration of the capture task was not measured, no participant took less than 45 minutes or more than 55 minutes to complete this portion of the experiment.

The narrative creation tasks, across all participants, averaged 23.81 minutes in length. Time required to create the individual narratives depended on which type of context was being used to search for video clips and narrative was created first. The statistical and design significance of these results is discussed in detail below. The entire duration of the experiment, including instructions, video capture and questionnaires, was approximately 90 minutes.

**4.3 Results**

We evaluated our hypotheses using the experiment described above. A summary of the results of this experiment are presented below.

**Measures**

The primary qualitative measure was the time needed to create the two different narratives. Video screen capture of the narrative creation tasks were used to determine time spent. To ensure that a poor search/composition interface wasn’t a source of
significant delay for some participants, we asked them to evaluate the ease of use of the web UI on the questionnaire. All participants except one either agreed or strongly agreed that creating narratives was easy. When asked about the search interface, 11 of 16 participants either agreed or strongly agreed that it was easy to use. The other 5 respondents said they felt neutral.

Participant reactions to the various types of contextual information were collected via the questionnaire. Participants were asked to rank each type of context on a five-point Likert scale for how useful it was to finding their own video clips, and then rank it again for how useful it was in finding video clips created by others.

**Effect on Creation Time due to Context Type**

Time taken to create each narrative is summarized in Figures 4.1 and 4.2. Figure 4.1 shows creation time for participants that used automatic context for video retrieval first, while 4.2 shows the times for those searching using manual context first.

The mean time for creation of narrative searching using automatic context for those participants that used automatic context first was 15.46766 minutes, and creation time for the manual narrative for the same group was 9.92083 minutes. For participants that searching using manual context first, the mean for creation of narrative using automatic context was 11.45208 minutes and using manual gave a mean of 10.78125 minutes.

Time of creation was relatively diverse, but we noticed that the difference of the means between the groups that used automatic context vs. those that used manual context
Figure 4.1 - Creation Time for Participants Using Automatic Context First

Figure 4.2 - Creation Time for Participants Using Manual Context First
Table 4.1 - ANOVA of Creation Time

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>19.83252</td>
<td>1</td>
<td>19.83252</td>
<td>1.522763</td>
<td>0.22746</td>
<td>4.195982</td>
</tr>
<tr>
<td>Columns</td>
<td>77.31852</td>
<td>1</td>
<td>77.31852</td>
<td>5.936601</td>
<td>0.021444</td>
<td>4.195982</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.5446</td>
<td>1</td>
<td>0.5446</td>
<td>0.041815</td>
<td>0.839452</td>
<td>4.195982</td>
</tr>
<tr>
<td>Within</td>
<td>364.6731</td>
<td>28</td>
<td>13.02404</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>462.3687</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

at first seemed substantial. Training effects clearly impacted the time taken, so we needed to determine if the difference in means was due to randomness and training, or if the type of context used makes a difference. Using a two-factor ANOVA analysis, summarized in Table 4.1, we can see the p-value of the columns (which corresponds to which context type was used first) of .021444 is beneath our significance alpha of .05, meaning there is statistical significance between time taken using automatic context and manual context.

**Effect on Creation Time due to Delay**

We also examined whether or not waiting a week between capturing the video clips and compiling them into narratives had any effect on the time needed to create the narratives. Figure 4.3 and 4.4 summarize creation time of the two groups. The mean time for creating the narrative using automatic context for the group that completed the study contiguously was 12.80307 minutes, and the mean for the narrative created using manual context was 9.43333 minutes. For the group that waited a week between capture and composition, the mean time for automatic created narrative was 14.11667 minutes and the manual created narrative was 11.26875 minutes.

Referring again to Table 4.1, looking at the variance within the sample (i.e. those who delayed vs. those didn’t), the p-value of 0.22746 is greater than the
Figure 4.3 - Creation Time for Participants Completing Study without One Week Delay

Figure 4.4 - Creation Time for Participants Completing Study with One Week Delay
significance alpha of .05, meaning there is no statistical significance in the time of creation between those who waited and those who did not.

**Contextual Preference for Personal Media**

On our questionnaire we asked users to evaluate how useful they felt each of the different types of context was for locating media of their own, and separately, media from others. We used a Likert scale of 1-5, where 1 meant the subject strongly disagreed with the usefulness of the context type, and 5 meant they strongly agreed. The summary of contextual preferences in regards to locating personal media is shown in Figure 4.5.

The highest ranked types of context were the author of the media, the textual tags and the semantic GPS location of the videos (manually searching for GPS coordinates was not used by any subject). Three context categories received a positive
response (response mean of 3.5 of more): author had a mean score of 4.3125, tags 3.875 and location 3.75. Weather received the lowest score at a mean of 2.75.

We again examined the difference between those participants that waited between capture and composition and those that did not. Statistically, there was no significance between the usefulness evaluations of those who waited.

**Contextual Preference for Other’s Media**

Figure 4.6 summarizes subjects’ responses in regards to media captured by others. Again, three categories received a positive response: author with a mean of 3.9375, location with a mean of 3.6875 and tags with a mean of 3.6875. Filename was the lowest category, at a mean of 2.6875.
Statistically, there was again no significance between those who waited and those who did not.

4.4 Discussion

Context Type Affects Usefulness

We found that using automatic context vs. manual context does significantly impact how long it takes to compile narratives. Automatic context was slower for creating video narratives, even when used second after participants were familiar with the system (and even some of the video content).

Users also prefer manual context to most automatic context. Across all participants, the mean Likert score for the manual contexts was 3.9531, while the mean of the automatic contexts was 3.1875, nearly a full point lower.

Users found manual contexts more generally useful, and it allowed them to perform the requested task faster. Because of this, we accept our first hypothesis that manual contexts are more useful than automatic contexts.

Delay Does Not Affect Usefulness

Contrary to the above, delay between capturing media content and compiling it into larger narratives does not affect how useful different types of context are or how long it takes to perform the narrative creation task. There was no statistical significance in the creation time between those who waited and those who did not. The total time for the second part of the experiment for the waiting group was 22.2364 minutes, while the
waiting group was 25.3333 minutes. This is likely explained by the two longest overall times (30.4833 minutes and 32.4167 minutes) being in the wait group.

User responses on the Likert scale questions also do not vary significantly between those participants that waited and those that did not.

**Context Types are Useful Equally for Personal and Others’ Media**

There was no way to determine from the analysis of creation time whether context types were more useful for locating personal videos vs. videos from other users, but analyzing participant response on the questionnaire indicates that there is little difference. There was no statistical difference between any of the context types when comparing the personal and other scores. The mean score across all context types for use with personal video clips was 3.4896, while use with others’ clips was 3.3959. The largest difference was between author with a mean of 4.3125 for personal videos and 3.9375 for others’. While not statistically significant, this difference is largely due to participants not knowing all the names of the other video authors. Participants frequently used the author context for locating others’ videos by virtue of finding one video by that author as part of another contextual search and then having that seed their author-based search.

**4.5 Summary**

In this chapter, we discussed the details of the user study we conducted to examine user interaction with Scheherazade. Our study tested two hypotheses: that manual context would be more useful than automatic context and that context would be more useful as time elapses between time of capture and time of use. Between our
examination of the time required to perform the two narrative creation tasks and results on the participants' follow-up questionnaire, we concluded that we have sound evidence confirming the first hypothesis, but insufficient evidence for the second. In the next chapter, we shall discuss the implications of these conclusions and suggest further investigations motivated by these results.
Chapter 5

Discussion and Future Work

The preceding chapter discussed our study of user practices and preferences when using Scheherazade. In this chapter, we shall be discussing the implications of those results in regards to previous work involving mobile media creation. We shall also identify future investigations to further examine the results of our experiences with Scheherazade.

5.1 Considerations for Utilizing Manual Context

As discussed in chapter 4, in terms of both qualitative and quantitative measurement, manual context information proved more useful to participants for the task of creating video narratives. Additionally, participants expressed preference for manual context, specifically the tag annotations, frequently in interviews and in commentary on questionnaires. Participant 14 (denoted as P14) used manual context first and provided the following comment on the questionnaire: “For the 1st narrative where keywords/tags were searchable, the videos were literally at your fingertips.”

This user preference for time of capture tagging has been seen before in studies of MobiCon [27] and ZoneTag [1]. However, both of these studies were not comparative in nature, in that it is not clear how users were able to interact with alternatives to manual annotation. MobiCon, for example, mentions use of GPS information, creator’s name, region, country and other automatically available information, but little information is available about how users were able to interact with this information and what their
responses to it were. We found this does not provide clear recommendations for improvement and further investigation. Thus, here we present four considerations for those looking to incorporate manual context into a media capture system:

- Do not require manual annotation
- Provide dynamic menus for tagging
- Provide a flexible, adaptable ontology of “suggestion” tags
- Manual context can lead to new de facto categories of context

The first issue related to manual annotation that we identified is that users frequently find it burdensome to be required to annotate each video (our system did require this). It has been well-established that time of capture annotation encourages substantially more tagging than post-hoc annotations systems [1, 9, 10]. But forcing users to annotate after capturing each media artifact can actually discourage any annotation beyond what is required. Additionally, as lightweight media capture is mobile by definition, users might not always have the time or ability to annotate media at the moment when it is captured.

One way to deal with this issue is to allow for a set of tags to apply to a sequence of media. As media capture frequently takes place around certain events [1], users can create a set of tags that apply to that event and reuse those tags easily until they transfer to another event. Several participants in our study commented on this, expressing a wish to be able to apply the same tag to a series of captured media. While our participants were capturing videos for a specific purpose, we still believe that allowing for use of prior tags, or simply skipping tags completely, would be useful in normal capture practices.
How users interact with inputting annotations is another area where we believe there are important considerations. We identified two distinct, but related, considerations when dealing with how users input manual context. First, using a menu to add annotations is vastly preferred over inputting new text via the phone's keypad. As discussed in chapter 3, we used an ontology-based menu with the option of typing in new tags via the keypad. However, if a menu-based system is being used, it is beneficial to have this menu be adaptive. Several Scheherazade participants commented that it would have been useful if any custom tags entered on the keyboard were then stored and made available as a menu choice, instead of having to type a more common custom tag many times.

Some participants also conveyed that they chose a less accurate tag from the menu so as not to have to type a more accurate tag using the keypad. P13 took this to the extreme, choosing a completely unrelated tag to refer to the main subject of one of the two narratives. They labeled approximately half of the videos with the “business” tag, even though these videos had nothing to do with a business. When asked about this, P13 remarked that they found typing on the phone keypad “very annoying” and knew what the “business” tag meant in the context of their videos. P13 said that when tagging in this fashion, they didn’t think there would be any difficulty in locating their clips during the second experimental task. While no other participants co-opted an unrelated tag to this extent, some participants did mention the desire to. When asked why they did not, they commented that they didn’t want the tags on their videos to be misleading to others. A more dynamic and user-customizable menu system would have avoided these issues.
The other consideration we identified that relates to menu-based annotation is the issue of menu size and scale. This can be a difficult issue to address and finding the right balance can be difficult. Some of our participants commented upon the size and content of the tag menus, with some saying they felt it was too broad. Conversely, others commented that it was too restrictive. While menu-based systems are clearly preferred to keypad-only annotations, populating an ontology that will be broad enough to serve many users, but still specific enough to be pointed, and small enough to find the appropriate terms quickly is quite difficult. We believe the best way to address this issue is to have a flexible ontology that adjusts itself to the user's context of capture. In the next section, we will discuss how automatic contextual information can facilitate these flexible ontologies.

Finally, we offer one last consideration in regards to manual contextual information. Aside from user preference toward manual context, a tagging system also offers the benefit of allowing the context types the system supports to be extended in an informal, ad-hoc way. For example, a system might not explicitly support automatically identifying lighting conditions in a certain video clip, a suggestion made by several of our participants. Even if it wasn't feasible to support this type of contextual information automatically, it is possible for users to support this manually by the use of custom tags. This type of emergent behavior has been seen in the past, with geotagged photos on Flickr. By providing tags "geo:lat=" and "geo:long=" with appropriate GPS coordinates, photos can be browsed by where they were taken. Eventually Flickr enhanced their service to better support geotagged photos, but initially geotagging only existed as a de facto practice among interested users.
User-driven extensibility of this nature can be quite valuable, as it allows smaller communities of users that all share a particular interest to create their own categories of contextual information unique from general tags. Two of our participants commented on this directly. P8 remarked, "Having manual tags give flexibility of adding more context types." P7 said in response to being asked what other types of context they would like to see supported, "Not sure, tagging kind of covers everything else." With the extensibility of well-supported manual context, systems designs need not go to great lengths formally supporting types of automatic contexts users might not find useful. As was evidenced with geotagging photos on Flickr, manual context can allows users to create their own categories of context, demonstrating usefulness and interest. If an ad-hoc category does prove to be popular and useful, the system designers can then make the necessary adjustment to support it more methodically.

In summary, we identified four important considerations in regard to utilizing manual contextual information. First, do not require every captured media artifact to be annotated, but provide mechanisms for assigning tags to sequences of media. Second, using menu-based tags is preferred to keypad entry, but the later should still be supported for terms not available in the menus. Additionally, saving terms users enter and adding them to the menu is beneficial. Our third consideration is that providing dynamic menus that adjust to their user's context is preferable to static menus. Finally, supporting manual context also affords the ability for users to create new de facto categories of context. These ad-hoc categories can spare the need to support them more formally, at least until it proves more valuable to do so.
One of our primary goals in this investigation was to separate and contrast usefulness of the automatic and manual contextual information. In doing so, we do not mean to create a false dichotomy between automatic and manual context. However, by examining them separately, we can gain a better understanding of how users interact with each type of context. Below we discuss important considerations to keep in mind when using automatic context.

5.2 Considerations for Utilizing Automatic Context

While manual context allowed subjects in our study to complete the narrative creation task faster and was their preferred choice of context, there are still ways to incorporate automatic context into a system that augment use of manual context. Most participants found the use of location contextual information useful, adding further credence to the idea that automatic context has benefits to offer. In this section, we discuss four considerations for designing systems that incorporate automatic context:

- Provide semantic meanings to location context
- Make system usable offline
- Keep the greater scheme of user practices in mind
- Allow for viewing and editing of contextual information

The first consideration specifically regards location contextual information. As location features prominently in many context-aware systems, we feel that it is especially important to provide usable location context. In most cases, location context is determined automatically in a numerical fashion, either using GPS, cell tower IDs, wi-fi
access points or another method. Data acquired in this fashion maybe be useful programmatically, but it is almost always meaningless to users. It is therefore beneficial to translate numerical coordinates into semantic locations that have meaning to users. Semantic locations have been demonstrated to be more useful and user-friendly than quantitative coordinates [2]. Participants in our user study universally preferred semantic locations to numerical GPS coordinates.

We believe the most useful way to integrate location into mobile media capture is to dynamically generate suggestion for manual annotations, as this has seen success in context-aware photography. This allows for association with semantic locations, but instead of the rather inflexible method of associating a fixed range of coordinates with a single location, users can use the suggestions to select the semantic location(s) that are the most meaningful to them. Saving the coordinates for the purposes of displaying video clips by capture location on an overlay map might also be useful, as this has proven popular with digital photographs. Augmenting manual context interaction with semantic locations derived from automatic context can bring the advantages of both to bear.

Related to our first consideration, our second observation in regards to utilizing automatic context is that it is essential for a context-aware media capture system to be usable while entirely offline. Automatic context is frequently acquired via network connectivity of some kind, usually Bluetooth or Internet data transfer. This can yield many benefits, but the consequence of this is that a constraint has been placed upon media capture that does not exist in traditional capture devices. If the network connection is sluggish or unavailable, it’s often unclear how well the capture systems will function, if they will function at all. Users do not expect network connectivity to be a
requirement for media capture, nor is it reasonable for designers to ask this of them. Thus, it is imperative that capture software provide operation in an offline mode, even if some of the functionality is limited. As media capture of this nature is highly mobile by definition, users may not have the time or ability to interact with context acquisition systems at the moment of media capture. Designers must allow for this scenario, lest requirement of network connectivity become a hurdle to users capturing media.

Thirdly, it is important to keep the greater scheme of user practices in mind when engaging with automatic context. A good example of this is the automatic file naming conventions on the Nokia N80s we used with Scheherazade. Files captured on these devices had a filename that consisted of the date, plus an incrementing integer for uniqueness, e.g. “06072007003.mp4.” While context-based filenames like this are certainly more useful than purely random integers, as is the case with some digital capture devices, using date is most likely redundant. Participants in our study were able to search by date using calendar menus and before-between-after delimiters, and they preferred this to wildcard-based results using video filename. The mean Likert scale score for usefulness of filename was 2.8125 out of 5, making it one of the lowest-scoring categories of context. Subjects ranked filename below explicit date searching virtually across the board.

In prior investigations into media capture and storage practices [15], it was reported that users frequently organize their photos into folders named by date and event. This would make having filenames set by date extraneous. We believe that it would be more useful to use some of the manual annotations to set the filename of the clip. P10 commented upon this on their questionnaire saying, “The filename could be more useful
if it is related to the context of the clip.” While giving media clips a date-based filename makes sense in isolation, when examined in contact of broader media management practices, a date-based filename offers little value. This is just one example of the importance of understanding user interaction with the system at large and designing context systems to support those larger practices.

The final, and perhaps most general, consideration when working with automatic context is allowing users to view and edit the context of their captured media. On their questionnaire responses, 11 participants agreed or strongly agreed with the statement “I would like to be able to view the contextual information associated with a video before uploading it.” 10 participants also responded with agreement or strong agreement to whether they would like to be able to edit contextual information before uploading. Several participants commented upon mild inaccuracies in the GPS-semantic location association, saying they would like to be able to correct this. Contextual information gathered automatically has clear potential for usefulness, but as this information is procured without user input, it is important to provide the ability to correct inaccuracies and other problems. As we suggested above, instead of assuming automatic context will be accurate, it may be more useful if automatic contextual information supplements manual context. By keeping a human involved in the context acquisition process, at least for context with potential accuracy issues, some of the issues involved with using automatic context can be avoided.

The results of user interaction with Scheherazade show that users find manual context more useful. However, this does not mean we believe automatic context is not useful or should be excluded. As we demonstrated with these four considerations,
automatic context can be utilized quite successfully to augment contextual information gathered manually. The possibility of examining synthesis between these two context systems, as well as other possible future investigations is discussed next.

5.3 Future Work

Our examination of user interaction with Scheherazade was a largely exploratory work, as little of the work in this area has attempted to qualitatively and quantitatively evaluate usefulness of contextual information in mobile video capture. By conducting this initial investigation, we can provide solid motivations for future experimentation.

Foremost, it would be valuable to examine user interaction with a video capture system that uses dynamic annotation menus based on user context. This has been demonstrated to be useful for mobile photography, and we believe it would be useful for mobile videography as well. Due to success with ZoneTag and similar systems, we do not believe there is enough reason to conduct an experiment focusing solely on user-specific menus for manual video annotation. Instead, these features could be combined with areas of investigation, such as those discussed below.

Another worthwhile area of future work would be examining the relationship between manual annotations and previewing videos by way of thumbnails and keyframes. There has been a significant amount of research on use of keyframes for video browsing, e.g. Boreczky et al. [4]. In using Scheherazade, at times, the first frame thumbnail of videos wouldn’t load immediately due to network latency. Some users either disregarded or ignored this problem by referring solely to the video’s tags for an idea of the clip’s content. P16 mentioned this, saying they found tags especially useful, “... because they
point to video content (thus, you don’t need to preview every single clip).” We believe motivation exists for examining the relationship between tags and keyframes in browsing videos.

Due to time constraints and a limited amount of hardware, our investigation took place in a laboratory setting. It is possible that there might be some interaction details and practices that we were unable to observe due to this. Thus, conducting a long-term study that made use of more organic media capture practices would be worthwhile. Additionally, by conducting all the video capture beforehand, it would be possible to allow users to search over the video clips of all the other participants. This could provide a much larger set of videos, which we believe would further emphasize the usefulness of being able to searching over contextual information. Two of our participants even commented upon this. P6 remarked, “Usefulness of context info goes up exponentially as the # of available videos increases.” P9 said, “I can see that ‘context’ and ‘tags’ would be more important in a larger scale system (more videos + participants).”

As discussed in chapter 4, our hypothesis that video context becomes more useful as time between capture and use increases could not be proven. However, this is not definitive, but rather it may simply be the case that a week is not a sufficiently long time to see these memory-augmenting advantages. The investigators of ZoneTag came to a similar conclusion, saying:

“...we found that the memory function of tags was still not a popular motivation. While it is likely that tags will provide this function as a currently-anticipated benefit in the future, adding context to facilitate
remembering details about photographs was not a primary motivation for
tagging in the present.’’

A lengthier study with more standard capture practices from users may allow the
memory function of tags to be more apparent. In general, it would be a valuable
experiment, as it would provide better insight into how users would make use of
such a system on a daily basis, instead of the more task-focused laboratory
setting.

In this chapter, we have discussed several important considerations when
design systems that incorporate both automatic and manual context. We believe
that manual context should be the locus of user interaction, but automatic context
can be leveraged to enhance functionality. In the next chapter, we provide a
summary of this entire work.
Chapter 6

Conclusions

In this work we set out to investigate the usefulness of different types of context in capturing video clips on a mobile phone and creating mashups of those clips. Little previous work had been done in this area, and that which had focused either on mobile photography or simply recording and saving videos. Very few related systems focused on providing qualitative or quantitative analysis of user interaction. We specifically designed our user study of Scheherazade to draw out user experiences, and provide a more thorough understanding of how users engage with context-aware media capture.

Our analysis determined that manual context is faster and preferred by users over automatic context. However, we believe that automatic context can still be very useful in supplementing the use of manual context, to provide interaction mechanisms that are useful and encourage users to engage with the system. By involving users where necessary, we can avoid many of the problems associated with the more programmatic and numerical automatic context. We can also leverage some of the benefits of automatic context without its drawbacks by providing users with dynamic and flexible methods of interacting with manual context, so as to make those interactions as simple and effective as possible.

This work comprises some initial steps toward a better understanding of how context-awareness can improve the ways media is recorded and shared via mobile devices. By emphasizing manual context and knowing the ways automatic context is and is not useful, we can begin to fully understand how context-awareness can improve media capture and sharing on the increasingly ubiquitous mobile phone.
Bibliography


[34] “YouTube serves up 100 mln videos a day”, Reuters, July 19th, 2006.
[38] http://www.digidesign.com/index.cfm?navid=28&langid=100
[45] http://www.w3.org/AudioVideo/
**Purpose:** The purpose of this study is to understand how the process of capturing video on smart cell phones can be enhanced by utilizing data available on these devices that is unavailable on other capture devices. By synthesizing captured video with contextual information about the when/where/what of the video, we believe this can improve the workflow for creating narratives based on short video clips.

**Objectives:** The objective is to understand how creating narratives based on video captured mobile devices can be made easier and more compelling. We want to better understand how contextual information can be synthesized with video to create better metadata and improve the workflow of creating videos based on these traditionally short clips. The contextual information we are examining is a combination of information that can be automatically inferred and information that requires manual entry from the content author. We want to examine usefulness of these two categories of contextual information.

Over a 1 ½ hour period we request your consent to participate in our study by capturing two different series of video clips and then arranging them into two corresponding narrative sequence. The capture task will require approximately 1 hour, and the narrative creation task will require approximately 30 minutes. You maybe be requested to complete these two tasks contiguously, or to complete the capture task and then return in approximately one week to complete the latter portion of the study. You shall be notified of this before you are asked to grant your consent.

Upon completion of each task, you will be asked to fill out a questionnaire, which will take approximately 5 minutes. No prior skills are required. You will be provided with instructions on the type of video to be captured and be shown how both the phone-based video capture software and the narrative creation software are operated. We will be providing the smart phone and all other related materials.

At all times during the study, the investigators will be able to answer any questions that you may have about the procedures, and ensure that they are fully understood.

**Dissemination:** The findings collected from this study may be used in a graduate Masters thesis, academic conference/journal publications, presentations, and workshops.

**Confidentiality**
The identities of all participants will be kept confidential. In field notes and in reporting of the results, participants' identities are hidden through the use of pseudonyms. Any personal information will be stored securely in a password protected computer account, or a secured lockable filing cabinet on campus.
Recruitment Form

Media and Graphics Interdisciplinary Centre (MAGIC)
2424 Main Mall
Vancouver, B.C., V6T 1Z4
(604) 822 - 8990

Study Recruitment for

Toward an Understanding of Context-awareness and Collaborative Narratives in Mobile Video Creation

Principal Investigator: Dr. Charles Krasic
Co-Investigators: Nels Anderson, Nicole Arksey, Mike Blackstock & Dr. Rodger Lea

We are looking for UBC students who would like to participate in an video capture and narrative creation study here on the UBC campus. If you have a free hour and a half to try our mobile-technology based video capture system please keep on reading.

Purpose: To understand how contextual information can be combined with video captured on mobile devices like smart phones to make creating narratives easier and more compelling. We are investigating how you use the video capture technology as well as how you use the contextual information to create sequences of related video clips.

Objective: The objective is to make creating video on mobile devices easier and more compelling. We want to better understand how contextual information (such as GPS coordinates and content tags) can be synthesized with video to create better metadata and improve the workflow of creating videos based on these traditionally short clips. The contextual information we are examining is a combination of information that can be automatically inferred and information that requires manual entry from the content author. We want to examine usefulness of these two categories of contextual information.

The Study: You first be asked to spend two 25 minutes intervals capturing at least 10 video clips using a mobile phone related to two different “narratives” (i.e. topics). After this, you will be asked to create video sequences using your captured video clips plus some additional clips to represent the two narratives. You will either be asked to do so immediately or approximately one week after capturing the video (we can arrange before you agree to participate). After completing both the video capture and narrative creation, you will be given a brief questionnaire and we hope you can give us your thoughts on the process. If there are questions at anytime, we the investigators will be available to assist you. Observation notes and questionnaire results may be used in a graduate thesis or other types of publications. The video clips you capture will not be used in this fashion. All information will always remain strictly confidential and anonymous.

There will be a $10 honorarium for this study.
Questionnaire

Context-Aware Narrative Creation Questionnaire

Part 1

1. In what age group are you?
   6  19 - 29
   7  30 - 39
   8  40 - 49
   9  50 +

2. What gender are you?
   o  Female
   o  Male

3. Do you currently possess a mobile phone?
   o  Yes
   o  No
   If yes, please mark the features your phone possess:
     o  Camera – still photos
     o  Camera – video
     o  Wi-Fi Internet Connectivity (802.11)
     o  Don’t know

4. How often do you use the following?
   1- never, 2- rarely (less than 1/month), 3-sometimes (monthly), 4-frequently (weekly), 5- daily/almost daily

   Mobile phone-based photo capture:
   Mobile phone-based video capture:
   Video editing software:
   (e.g. iMovie, Windows Movie Maker)
   Internet-based video systems:
   (e.g. YouTube, GoogleVideo)
Part 2

Section A:

With respect to the last task, please indicate the extent to which you agree or disagree with the following statements:

SD = Strongly Disagree
D = Disagree
N = Neutral
A = Agree
SA = Strongly Agree

<table>
<thead>
<tr>
<th>Statement</th>
<th>SD</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capturing video was easy.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Adding annotations the video via the list was easy.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Adding annotations to the video via text input was easy.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I would like to be able to view the contextual information of a video before uploading it.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I would like to be able to edit the contextual information associated with a video before uploading it.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Section B:

1. What particular aspect(s) of using the mobile phone to capture video did you *not* like? Why?

2. What particular aspect(s) of the mobile phone to capture video did you like? Why?
Part 3

Section A:

With respect to the last task, please indicate the extent to which you agree or disagree with the following statements:

SD = Strongly Disagree
D = Disagree
N = Neutral
A = Agree
SA = Strongly Agree

<table>
<thead>
<tr>
<th>The search interface was understandable and easy to use.</th>
<th>SD</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating narratives was easy.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by filename of the video clip was useful for finding clips I personally captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by filename of the video clip was useful for finding clips someone else captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by creation date/time of the video clip was useful for finding clips I personally captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by creation date/time of the video clip was useful for finding clips someone else captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by author of the video clip was useful for finding clips I personally captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by author of the video clip was useful for finding clips someone else captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by the location of where the video clip was captured was useful for finding clips I personally captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by the location of where the video clip was captured was useful for finding clips someone else captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Being able to search by the weather when the video clip was captured was useful for finding clips I personally captured.</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>
Being able to search by the weather when the video clip was captured was useful for finding clips someone else captured.

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Being able to search by annotations to the video clip was useful for finding clips I personally captured.

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Being able to search by annotations to the video clip was useful for finding clips someone else captured.

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Section B:

1. What particular aspect(s) of using this interface to create narratives did you like? Why?

2. What particular aspect(s) of using this interface to create narratives did you not like? Why?

3. a) Did you find certain types of context information were not useful for this task?

   b) Do you believe they would be useful in a more general scenario?
4. What other types of context do you believe would be useful in a scenario like this?

5. Other comments?
Appendix B – UBC Research Ethics Board Certification

**CERTIFICATE OF APPROVAL - FULL BOARD**

<table>
<thead>
<tr>
<th>PRINCIPAL INVESTIGATOR:</th>
<th>INSTITUTION / DEPARTMENT:</th>
<th>UBC BREC NUMBER:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charles Krasic</td>
<td>UBC/Science/Computer Science</td>
<td>H07-00944</td>
</tr>
</tbody>
</table>

**INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:**

<table>
<thead>
<tr>
<th>Institution</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Other locations where the research will be conducted:**

<table>
<thead>
<tr>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
</tr>
</tbody>
</table>

**CO-INVESTIGATOR(S):**

- Mike Blackstock
- Rodger J. Lea
- Nicole Arksey

**SPONSORING AGENCIES:**

- N/A

**PROJECT TITLE:**

Towards an Understanding of Context-awareness and Collaborative Narratives in Mobile Video Creation.

**REB MEETING DATE:**

- April 26, 2007

**CERTIFICATE EXPIRY DATE:**

- April 26, 2008

**DOCUMENTS INCLUDED IN THIS APPROVAL:**

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent Forms:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consent Form</td>
<td>1.1</td>
<td>May 3, 2007</td>
</tr>
<tr>
<td>Advertisements:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment Form</td>
<td>1.1</td>
<td>May 3, 2007</td>
</tr>
<tr>
<td>Questionnaire, Questionnaire Cover Letter, Tests:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>1.1</td>
<td>May 3, 2007</td>
</tr>
</tbody>
</table>

The application for ethical review and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

- Dr. Peter Suedfeld, Chair
- Dr. Jim Rupert, Associate Chair
- Dr. Arminee Kazanjian, Associate Chair
- Dr. M. Judith Lynam, Associate Chair
- Dr. Laura Ford, Associate Chair