Social Practices around Personal Videos using the Web

Rodrigo Laiola Guimarães¹, Pablo Cesar¹, Dick C. A. Bulterman¹, Ian Kegel² and Peter Ljungstrand³

¹ CWI: Centrum Wiskunde & Informatica
Science Park 123, 1098 XG
Amsterdam, The Netherlands
+31 20 592 9333
{rlaiola, p.s.cesar, dick.bulterman}@cwi.nl, ian.c.kegel@bt.com, peterlju@tii.se

² BT Research & Technology
Adastral Park, IPS 3RE
Ipswich, UK
+44 (0)1473 642916
³ Interactive Institute
Box 1197, SE-164 26
Kista, Sweden
+46 702 889759

ABSTRACT
Social multimedia is changing the way people interact with each other, transforming old practices on political activism, social participation and interpersonal relationships. Sharing dynamically created video segments is a prime example of this social transformation. This paper reports on a long field trial studying social practices around personal videos within a tight social circle. Our results show that users appreciate the importance of video sharing for feeling more connected with group members and for building shared experiences. Our results also indicate that a new generation of video management systems is needed to maximize the impact of group sharing, with a focus on more efficient automatic processes to manage the social experience. A fundamental user requirement is finer-grained privacy controls over video content distribution. Paying special attention to existing social theories and human-centered methodologies this paper brings implications to be considered in the design of future home video management and sharing systems.

Categories and Subject Descriptors
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Audio, Video.

General Terms
Design, Experimentation, Human Factors, Management.

Keywords

1. INTRODUCTION
Media capture is a ubiquitous activity. Shared viewing of that captured content typically is not. In spite of the dramatic impact of user contributed content sites (such as YouTube and various national video sites), the amount of ultra-personal content being shared with family and friends (to say nothing of wide anonymous audiences) is minimal. A conservative estimate media use indicates that average owners of smartphones and portable cameras capture hours of videos yearly, but that only minutes (or seconds) of content are being shared. The question we have is: why is this?

Modern video cameras allow a user to exercise a reasonable level of creative control during video capture. Shots are created using cameras with high resolution, often resulting in HD video output. Lenses are sensitive enough to allow capture to take place in dimly lit venues, yet small enough to fit in a shirt pocket. In general, audio quality is high, and recording skew is minimal. Finally, most cameras allow a dazzling array of metadata to be captured, including everything from camera details, locations and faces to be identified during capture.

In spite of these advances, content is not becoming more accessible: where face/person recognition in images is becoming commonplace, the successful high-level analysis of video content is still in its infancy. Most captured home video is poorly lit and densely shot. This makes video analysis difficult, which in turn makes indexing, finding and reusing captured content by end-users frustrating.

The problems of (re)using personal content have several other dimensions as well. At a high level of abstraction, a fundamental problem is that third-party users expect a higher level of production content (in terms of shot selection, story pacing and logical narrative) than most amateur authors have available to them. At a lower level of abstraction, searching through captured content is simply too laborious to provide a productive user experience. Put another way, as the digital family archive gets larger, the more difficult is to find something interesting.

In this paper, we consider the problem of reusing personal video content. Given the characteristics of much end-user content, we have chosen to focus on studying the social practices of video capture. Rather than analyzing video content, this article presents a qualitative study of the people behind the camera. In this paper we are interested in better understanding the social practices around personal videos within groups of people with strong ties [8]. We aim at helping users to create and share personal video stories. We have approached these challenges from three different perspectives:

1. **behavioral profiling**: can we capture and use end-users’ recording behavior to allow for a personalized exploration of a shared media repository?
2. **video creation**: can video editing techniques be employed to help users in the creation of personalized video stories from a shared media repository?
3. on-demand personalization: can video editing tools take into account the relationships between producers, users, and media for generating personalized videos?

Our ultimate goal is to better understand the requirements for a next generation video management and editing tools, so that users can make more use of their personal footage to share experiences with family members and friends using the Web.

This paper uses a novel model for media collection and sharing. We have studied small-scale collaborative content collection. Consider a high school band concert, where a dozen or so parents capture video during the concert. Each of these parents is primarily interested in their own child, but all of them will create content that will contain (potentially) interesting shots of others in the band. How can these supplemental shots be integrated in to personal narratives created by individual parents?

This paper is structured as follows. The next section provides the background. It discusses the social theories that motivated our work, and sets the end-user production use case around which we have realized this investigation. Section 3 reports on a user study about social practices around personal videos using the Web. Next, Section 4 discusses the implications of the findings in the user study, thus answering our research questions. Finally, Section 5 presents the related work, while Section 6 dedicates to concluding remarks.

2. MOTIVATION

The work presented in this paper is part of a pan-European project called Together Anywhere, Together Anytime (TA2). The goal of this project is to understand how technology can improve relationships between groups of people separated in space and time. One of the scenarios of this project, MyVideos, explores an asynchronous sharing framework in which highly personalized music videos are constructed from a collection of independent parent-made recordings.

The results presented in this paper are part of a long field trial, from December 2009 until September 2010, in which a group of families in the Netherlands collaborated in preliminary interviews, video recording and social sharing of content within a relatively closed community [4]. Our long-term association with these families has provided us invaluable insights on how so-called Web 2.0 technologies fail to meet the demands for interpersonal communication between friends and families. It has provided us as well a deep understanding on current practices around home videos, and how such practices can be matched to existing social theories.

2.1 Social Science Background

Social presence examines how well a mediated system can mimic the feeling of presence of others. But the concept of social presence is solely focused on the present situation and not on how bonds are developed over time. Hence the concept of social presence is not sufficient for understanding how relationships evolve over time as a shared, reciprocal bond beyond the present situation. But we find this in the concept of connectedness [15]. While communication is essential for connectedness, it does not presuppose a continuous open communication channel, but merely sufficient communication and contact to allow an existing relationship to be maintained. Even media with a low richness and naturalness such as text messages can create the feeling of connectedness. It is enough for one part to be aware of the other and the other person to confirm his awareness [13].

Even though the emotional energy constituting a shared bond must be recharged from time to time by situations of social presence, connectedness can help maintain a relationship. Figure 1 illustrates a schematic view of the perceived strength of a social bond over time, showing reoccurring shared events (interaction rituals) in the Durkheim sense [6], with a fading strength of the social bond in between. The peaks in the figure correspond to shared events and gatherings, mediated or not, referred to as interaction rituals (e.g., the concert). During these events, participants carry out some joint activity, re-affirm their relationships, while both building on and extending their common pool of shared memories. The smaller peaks correspond to social connectedness actions (e.g., watching a video that was recorded at the concert or sharing a video). Reflecting on the results from the field trial, we can infer that future home video sharing systems should be designed for mediating the smaller peaks, providing tools and mechanisms for maintaining and fostering relationships over time.

2.2 An End-User Media Production Use Case

In the spring of 2010, we put into practice the theory introduced above at the spring concert of one high school in Amsterdam. A group of amateur musicians performed in front of their parents, siblings, and classmates. People in the audience brought their video cameras and recorded content primarily for personal use. Each family was interested in capturing their own child, but they also wanted to record enough contextual information from the concert. Such personal footage was intended to become part of the family archives, to be watched in family gatherings, and to be shared with family members that live away. Eventually, the recordings will become shared memories of what happened in that occasion. This scenario is a typical example of small-scale social event, fitting the social theory represented in Figure 1.

After the concert had taken place, some of the people from the audience provided us their video material. In total around 200 media objects were collected for a concert lasting about 1h and 35 minutes. Twelve (12) cameras were used; two of them used as the school camera setup.

3. USER STUDY

In the summer of 2010 (July-September) we run a user study with the some of the people that participated in the concert. More interested in subjective results that in statistical data, our approach was largely exploratory and interactive. This section reports on the findings regarding video recording and editing.

---

1. http://www.ta2-project.eu/
Both sessions were started with an ice-breaking activity on the whiteboard (see Figure 4), followed by discussions around the research questions.

Participants

Seven people among relatives and friends of the performers participated in the user studies. The rationale used for recruitment was diversity. We wanted to gather people with different roles for better understanding the social needs of our potential users. Participants’ occupations included student, social scientist, software engineer, art designer and visual artist, resulting in a variety of needs that may influence the video capturing, editing and sharing practices. All the participants were Dutch, and all of them at the time of the study were living in the Netherlands, but an uncle of a performer that lived in the USA (the only one that was not present in the concert recordings). Figure 2 shows the profiles of the participants.

Since our main focus is to better understand small groups of people with strong interpersonal ties, this study had to be realized with a fixed selection of users. It would have been impossible to do crowdsource testing, since we wanted to explore the fact that people had a social connection with the recorded footage. In general we believe that this sample of users provides us a relevant picture of the ways they currently record videos of people they care about, and how they use such footage to share experiences within their (probably restricted) social group.

Methodology

We used different methods for data collection including in person interviews (each participant individually) and questionnaires. Because we suspected their practices would be different, and as a point of contrast, we followed a semi-structured approach. We ran the experiment in the lab, and it consisted of 2 sessions (approximately one hour each). The initial session was used to collect background information about video recording habits. It was used as well as an opportunity to better understand how the participants conceptualize the concert (see sketches in Figure 3).

The second (in-depth) session was dedicated to explore the video editing practices and media sharing routines of the participants. For this second session we used the footage that they had recorded in the spring of 2010 during the high school concert. Both sessions were started with an ice-breaking activity on the whiteboard (see Figure 4), followed by discussions around the research questions.

All participants reported they normally record videos in social/family events, such as the high school concert. The average shot duration and the percentage of time they spend recording vary, depending on the event and the goals (for instance, a father told us he records a lot during his daughter’s concerts because it might be useful in her musical education).

Validating previous studies [11], most of the participants barely look into their footage after the event has taken place. One problem is the relatively high number of media assets - Around 200 media assets from 12 cameras for a concert lasting 1h35min. Another problem is that the footage, as captured, cannot be easily explored. When prompted they agreed that it would be good to have better tools to add annotations. But the reality is that participants were quite resistant about correcting wrong metadata, arguing that this process demands too much effort. This however, could be different if a more incidental and casual user interface for annotations were provided.

"It is not my problem (correct the wrong metadata)... people don’t have time to play with their own footage." (Uncle of a performer)

Figure 2. The makeup, age, and gender of the participants.

Figure 3. Users’ sketches illustrating the concert setup.

Figure 4. Engaging participants in an ice-breaking activity.
Regarding video editing, most of the users declared that they had an editing suite at home. For basic edits our Windows users were familiar with Windows Movie Maker. Participants described to us how they would create a movie about the high school concert using their preferred editing tool. They would choose some of the clips and drag them to the timeline. Then, they would use visual effects, transitions and sounds that are provided with the video-editing suite. In general, participants indicated that they would tell the story of the concert using their personal videos.

When questioned whether their video material would be enough to create a compelling video, they mainly answer negatively. They agreed that content captured by other people that participated at the same event could be interesting for others. For example, it could be used for breaking the monotony of a single camera viewpoint, for providing complementary views of the performers, and for getting interesting events that were not captured by one’s own camera. However, most of the users asserted that current tools do not allow for easy watching and repurposing other parents’ footage. Although being familiar with end-user video editing tools, for most of the participants video editing is time consuming and a complicated process. Moreover, video editing can provoke high CPU consumption, slowing down the computer. For that reason, one mother reported she carries out minor editing tasks, such as clipping, on her own video camera; since her personal computer cannot handle vast amounts of video material efficiently.

Media sharing and social relations

All participants said they appreciate watching videos on YouTube, and most of them have a Facebook account. However, only a few of the participants actually are engaged to collaborative tagging of videos and/or photos [12]. When prompted whether and how they share their videos, our participants repeatedly said that in general they do not post

---

2 The services and technologies mentioned in this paper, if unknown, could very easily be identified via a simple Internet search, therefore they will not be Web-referenced.
personal videos on the Web. While the youngest participants argued their personal videos were not interesting enough, for our older respondents privacy was the main concern not to share personal videos on the Web.

“It is personal... if I make a personal shot, a close-up of my daughter, for example, and I do this for personal reasons, I never do this for the others.” (Mother of a performer)

In the case of the footage from the high school concert, when asked whether and how they would like to share videos, teenagers expressed they would rather download video files to their local computers, send a link of a particular video by email or share on YouTube and/or Facebook. The parents however, informed us that a ‘Burn to DVD’ functionality of some selected videos would be also convenient given that grandparents usually do not have Internet access at home.

“I put the footage one a disk, a DVD disc. just pieces, that’s like that... that’s my disk! And the other thing is grandmother doesn’t have Internet, so it should be in a disc, I put it on my laptop and I show to grandmother.” (Mother of a performer)

For some of our participants interactivity functions (e.g. show me more from this person) while watching videos were seen as an interesting feature. However, some parents were opposed to that.

“I think children and young people would start manipulating the system then, seeing what they can do etc... but then it becomes a game!” (Mother of a performer)

4. DIRECTIONS FOR FUTURE SYSTEMS SUPPORT

In this journey, we have learned many lessons. One important implication is the need for a new generation of automatic tools for analyzing user-generated content. There is a need for better tools that can handle poorly lit and densely shot footage. Moreover, these tools should be capable of recognizing special moments (e.g., a solo) and of filtering out segments in which the performer is not well portrayed. Regarding privacy concerns, participants felt that they could rely on a system, if a trusted institution – like the school of their kids – managed it. In general they considered it to be a safe place to contribute their assets.

In most cases, participants were actively looking for video clips of their close friends and relatives. They complained – and desperate – when the quality of the video was not good enough or when the metadata of the videos was wrong (see Figure 7). And most importantly, they wanted to immediately share video clips with members of their close circle. “Can I send it now?” was a common reaction of the participants after seeing a video clip they especially liked.

Based on the participants’ comments, suggestions and reactions, we got a strong sense that current tools are not enough to attend their needs. Current video sharing platforms on the Web do not allow for a collection of families that may have limited interactions to be brought together by contributing media assets for common use.

The most relevant implications of this work can be summarized as follows:

On considering user profiling based on user’s recording behavior

Content analysis: most of our participants reported they record videos in social events, but they barely look at those afterwards.

Tagging is considered an important process since it allows for easier search of subjects of interest, e.g. events and performers. However, users do not want to spend time annotating their own material, neither video content recorded by others. Thus, we need better annotation processes such as metadata extraction tools that work under user-generated content conditions (e.g. with poor light, cameras with different video quality and encodings etc.). Other solution might be to allow annotation of content while the person is recording.

Exploration of the assets: as with recommender systems [2], a profile of the user can facilitate exploration of the video assets. One way of profiling them is to use their previous activity (log) or to request them to fill some personal data in the system. While these approaches can provide relevant results for a statistically significant group of people or group of actions, they do not take into account small groups of people, and their recording behavior. Figure 8 shows the results of analyzing the metadata associated to the media captured during the high school concert. In the figure, the recording behavior of a mother towards her kid is compared with the average behavior of the rest of the parents. Based on the results, we can see that recording habits and relationships provide an important cue that should be considered in the profiling process. We can notice that the affection level towards a performer greatly influences the overall time a recorder spent capturing that person. In this paper we argue that apart from traditional profiling approaches, user-recording behavior might provide complementary knowledge (research question 1), when the overall number of interactions with the system is not statistically significant.

On considering video editing processes to create good presentations:

Our participants see the value of sharing personal media within their social group, but they do not edit their own videos because it is cumbersome and time-consuming. More importantly, they find that their own material is not sufficient for creating a compelling video story. Generally, users do not ask other parents for their material, even though they would like to see the footage other people recorded. Even if they ask for other recordings, current tools do not allow for gathering all the material in a convenient and easy way to be explored. If we consider that it is difficult to find what it is relevant in one’s own footage, it becomes daunting when the media space is contributed by a dozen of parents; creating videos out of the event is almost an impossible mission. Therefore, automatic-authoring processes that take into account users choices and behavior will play an
important role to help users in creating appealing videos (research question 2).

On considering on-demand editing based on the social relationships:

Our participants do not want to spend time and effort, so they appreciate the use of automatic processes for creating videos. However, while automatic processes can provide video stories with not so much effort, these videos do not always reflect the user intent. While fully automatic video generation happens to be almost satisfactory, our participants miss the personal touch. In this case, recommender systems can play an essential role in helping users to identify relevant alternatives for portions of an authored video. We call this on-demand personalization. Where video-editing systems targeted to end-users should consider who is making the video, what is the intent, and with whom s/he is going to share the story. The final goal is to provide users a number of helpers that allows them to create the story they want to tell. Using an automatically generated multi-camera video as the starting point (with no effort), future systems should consider the supporting mechanisms for slightly modifying it, making it more personal and intimate (research question 3).

5. RELATED WORK

A number of research efforts have focused on the aggregated behavior of people and social media. Examples include location-aware photo sharing systems [5], studies on micro-blogging around live events [14], and collection management of YouTube videos [10]. Unfortunately, studies that focus on the individuals and their context are scarce. This paper provides such study, focusing on personal videos as digital Web artifacts to be created, explored and shared with a delineated community.

Not to forget that sociological theories [7] and user-centered approaches [3][9] also have tackled different aspects of the multimedia workflow. For instance, human-centered studies [11] investigate what people do with media, balancing the preponderance of techno-centric work with appropriate user-centric insight. Similarly, in our work we pay special attention to social theories and human-centered methodologies.

While our work deals with recording, management and watching video material, the social practices surrounding it share many similarities with how people use photographs. Van House et al. [16] reported findings from four inter-related empirical studies of present and emerging uses of photos in everyday life. While making home-visit interviews, people consistently apologized for their disorganized photos and claimed they would better annotate and arrange their images – someday. This indicated a shared value of selected, edited, labeled and ordered images, however, rarely achieved. The interviewees also generally saw little need for metadata beyond date of capture, since they were only intended for close friends and family and they knew everyone in the images anyway. While several people commonly used Flickr as a repository for their images, only a few of the participants commonly added metadata tags to their images. Those who tagged generally did so for other Flickr users – basically for strangers, as a sort of civic duty.

Another general finding relates to the sheer volume of data that is easily generated by digital image capturing technologies. The participants in the study did very little to label or organize digital photos stored locally. They were quickly overwhelmed by the size of their collections and the opacity of computer-based storage, with indecipherable file names. This indicates a general need for techniques that can assist and lower the task of indexing and organizing captured photo/video material into coherent units.

Logical sequencing of images/videos of recorded events is important in the sense making process of viewers. However, typically a narrative is also needed to guide the interpretation of the images and to provide further meaning and significance. According to Van House et al. [16], the sequencing of photographs (as in a photo album or a stack of prints) is important for the construction of meaning to the viewer, however also usually accompanied with a narrative. While sequencing is inherent in all video material, this points to the importance of logically arranged video clips, guided by some form of narrative to make the resulting material intelligible to the viewer, in particular if oral, co-present narration is not possible. The same argument is supported by Walker and Mouldon [17] who examined over 40 photo albums by non-professional photographers. According to their findings, most albums are
created according to some sort of narrative. They conclude that photo albums as such are generally insufficient for a viewer to make sense of; rather, they are intended to be presented to one or a few individuals at a time with an accompanying oral narrative. Finally, our work is also related to home video management and navigation [1]. However, we go one step further by considering the possibility of exploring and utilizing material contributed by other participants of a shared event.

6. CONCLUSION AND FINAL REMARKS
Recognizing the importance of looking at common media assets as a cornerstone for the sharing of group experiences [18], the scenario explored in our research revolves around personal videos. In particular, it focuses on tools that help small groups of people sharing videos of a high-school concert. Unlike many collaborative editing systems, the primary purpose of the content sharing is to serve as a resource that can be used by members of the extended community to serve a collection of individual needs. For data collection we used multiple methods for an extended period of time, including: focus groups, in-person interviews, interactive tasks, and in-situ video recordings of a school concert.

7. ACKNOWLEDGMENTS
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2007-214793. Special thanks to all partners involved in the MyVideos demonstrator.

8. REFERENCES