WHY DID YOU RECORD THIS VIDEO? AN EXPLORATORY STUDY ON USER INTENTIONS FOR VIDEO PRODUCTION

Mathias Lux

Klagenfurt University
Institute for Information Technology
Universitätsstr. 65-67, Klagenfurt, Austria
mlux@itec.uni-klu.ac.at

Jochen Huber

Technische Universität Darmstadt
Telecooperation Group
Hochschulstr. 10, Darmstadt, Germany
jochen.huber@acm.org

ABSTRACT

Why do people record videos and share them? While the question seems to be simple, user intentions have not yet been investigated for video production and sharing. A general taxonomy would lead to adapted information systems and multimedia interfaces tailored to the users' intentions. We contribute (1) an exploratory user study with 20 participants, examining the various facets of user intentions for video production and sharing in detail and (2) a novel set of user intention clusters for video production, grounded empirically in our study results. We further reflect existing work in specialized domains (i.e. video blogging and mobile phone cameras) and show that prevailing models used in other multimedia fields (e.g. photography) cannot be used as-is to reason about video recording and sharing intentions.

1. INTRODUCTION

The advent of highly capable devices and software for video recording and editing enables users to quickly capture scenes, edit and share them. This leads to an overwhelming amount of videos available online for example through YouTube. Still users complain about complicated procedures as software and devices do not meet their expectations. In that sense, the expectations of a user are defined by the current situation (i.e. the context in which the user is for instance recording a video) and the goal to be achieved (i.e. to capture an emotionally important moment). The actual goal pursued in a specific situation thence characterizes a user's intention: why she captured, edited or shared a video in that very moment.

A professional video blogger for instance has different intentions than a recent grandmother. While the blogger wants to deliver a message, or eventually express herself [1], a grandmother might be interested in capturing the first steps of a grandchild to revisit and recall the moment later on. Actual intentions or goals are most likely very individual and diverse. Still, if we assess those intentions and classify them, we can use the intention of the video producer within multimedia systems to provide more meaningful services. For video retrieval, a system might be able to match the intention of the video creator to the intentions of people who actually search for videos. Such an approach fosters more appropriate recommendations and search results by filtering out videos that are not appropriate to the viewer's intention.

Thus our research focuses on possible taxonomies and classification schemes, which allow for grouping similar intentions. User intentions in the field of multimedia production have mostly been investigated in the field of photography. Only little effort has been spent on analyzing user intentions for video recording. Moreover, we are not aware of any taxonomy in that field. In this paper we contribute an exploratory study with 20 participants reporting on 48 situations when they took videos with arbitrary recording devices. We first discuss prior work, describe main research goals, and outline the study setup. Afterwards, we discuss the instances collected throughout the study to investigate if and how existing taxonomies from other fields can be applied to user intentions for creating videos. Based on our results we conclude by outlining the requirements and need for a new user intention classification approach in video production and sharing.

2. PRIOR STUDIES

The most prominent work in the field of user intentions has been presented by Broder [2] and Rose and Levenstein [3] on user intentions in web search. Intentions in the field of multimedia production however have been discussed only rarely in research. Most effort has been spent on the analysis of user intentions when taking photographs. In the following, we discuss studies in that field. Moreover, there has been only little research on user intentions for video recording, which we discuss in the subsequent subsection.

2.1. Studies on photography

Directly related to our work is the work of Kindberg et al. In [4] they present the results of a study on the use of mobile phone cameras. They define a 6-way taxonomy based on two

main dimensions: "function vs. affection" and "sharing vs. closed group sharing vs. individual use". Their study shows that people mostly (41% in their study) took photos out of personal reflection and reminiscence (affection/individual). The work of Kindberg was reflected in a study in [5] on a more general population, not restricted to mobile phone cameras. The study showed that borders between affective and functional intentions are fuzzy and also the 3 classes of sharing intentions cannot be differentiated easily.

In [6] image based communication within families is investigated. Within field trials, families were equipped with devices to create and edit images and to share them between family members. They found that intentions differed over generation. Parents typically made functional use of the system and monitored their children, while grandparents shared their feelings with their children and grandchildren. The kids, however, used the devices to create and share stories and funny situations.

2.2. Studies on videos

While Kindberg et al. and Makela et al. focused on photos, Bornoe and Barkhuus investigated motivations for video microblogging in [1]. Video microblogging has been promoted as a tool for social collaboration, with the main goals of bloggers being (i) self expression, (ii) entertainment, and (iii) self presentation.

Puikonnen et al. [7] investigated the use of mobile phone cameras for taking videos in a study with 11 participants. Although they did not primarily focus on user intentions for taking videos, they found that in most of the scenarios people wanted to preserve a moment of interest for them, which is only occasionally shared.

3. STUDY DESIGN AND METHODOLOGY

We conducted an exploratory study to grasp the main intentions driving people to record and share videos. There were 20 participants, 16 male and 4 female (see Table 1). We conducted semi-structured interviews. Besides assessing demographic information and general usage statistics, we focused on two main dimensions: communication and recording habits. Regarding communication, we asked the participants for instance if they shared videos on a regular basis; if so, with whom, what the videos should have conveyed and how they tried to achieve this goal. Regarding the recording habits, we asked the participants to report on situations when they had taken videos in the past. They were asked to describe the situations, their key motivations for capturing the scene and what they did with the captured videos. Based on these questions, 48 different situations when they took a video were reported. All interviews were recorded and transcribed. We analyzed the transcriptions on word-level using an open coding approach [8].

Table 1. Demographical data, usage classes are daily (1), more than once a week (2), more than once a month (3), and once a month or less (4).

	Gender	Age	Usage
P1	m	25-30	1
P2	m	25-30	1
P3	m	18-24	2
P4	m	18-24	2
P5	m	18-24	2
P6	m	$<$18	2
P7	m	25-30	2
P8	m	$>$40	2
P9	m	$>$40	2
P10	f	25-30	2
P11	m	31-40	2
P12	m	18-24	2
P13	f	18-24	2
P14	f	31-40	3
P15	m	18-24	3
P16	m	31-40	3
P17	m	$>$40	4
P18	f	25-30	4
P19	m	25-30	4
P20	m	18-24	4

Due to the fact that there is no existing taxonomy categorizing user intentions for video production, we investigated existing classification schemes from other fields closely. Most promising candidate was the one proposed by Kindberg et al. [4]. It classifies basically along two different dimensions and proposes 6 classes (see Table 2). Based on our collected data and the existing taxonomy, we focused on whether and how the taxonomy can be applied to video production. More specifically, we focused on two main research questions:

- *Are the classes proposed in [4] disjoint?*
  Questions directly related to this are: If a user creates a video, can we assume that the user's intentions can be assigned to exactly one single out of the six proposed classes? Can people have multiple intentions at the same time?

- *Is the proposed two-dimensional taxonomy sufficient to describe user intentions for video creation?*
  Questions directly related to this research question are: Is "personal reflection" the opposite of "sharing" as indicated by the work of Kindberg et al. (see Table 2, "personal reflection") or do we need an additional class or dimension for classification?

Table 2. 6-way taxonomy as proposed by Kindberg et al. in [4] for intentions driving image capturing.

	Social		**Individual**
Affective	Mutual experience	Absent family or friends	Personal reflection
Functional	Mutual task	Remote task	Personal task

4. RESULTS

The taxonomy shown in Table 2 gives a good starting point for an analysis of the user reports. Our first step was to cluster all 48 instances based on primary intention of the user when recording the video. We analyzed transcriptions and found that most of the reported instances reflected common aspects or background. The clustering revealed 6 groups of intentions different to the ones proposed in [4]. Compared to Table 2 functional intentions are not divided into social and individual by their scope. Also intentions of sharing and preservations are not at opposite ends, but might apply at the same time.

- **Preservation**: Storing a scene to view it later
- **Sharing**: Showing scenes to others
- **Affection**: Capturing a scene due to emotion
- **Functional**: Video is part of a job, hobby, etc.
- **Technical Interest**: E.g. trying out a camera
- **Other**: Unknown or unmentioned intentions, etc.

Each of the 48 instances was assigned to these groups. In contrast to the taxonomy proposed in [4], we allowed assignment of instances to multiple groups where appropriate. During analysis, 9 instances were assigned to one single group, 31 instances had double assignments to two different groups and 8 instances have been assigned to three different groups. It was not necessary to assign a single instance in our dataset to more than 3 groups.

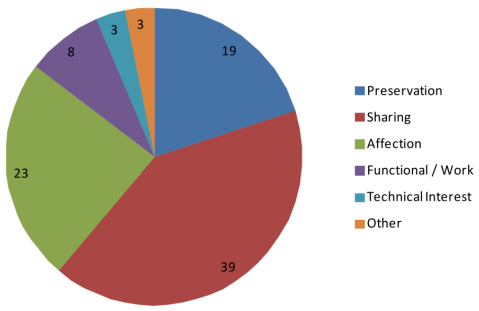

Fig. 1. Assignment of instances to the six groups along with the absolute number of assignments per group.

Figure 1 details the numbers of the assignments. Most notable fact is that nearly all of the videos (39 out of 48) were taken with some sharing intention in mind. 29 of the 39 instances were only shared to family, friends, colleagues or other closed groups. Affection drives the user to capturing the video in 23 cases. 19 instances could be assigned to the "Preservation" category. Three videos, which were "taken out of boredom" or "just for fun" were assigned to the "Other" group. Also a class of minor interest is "Technical Interest", consisting of three videos, which were taken to "try out the camera".

Using the class assignments, we then investigated the pairwise cross tabulations and investigated the pairwise correlation of assignments based upon the phi coefficient φ to find indications whether groups are independent or correlated in some way (see Table 3). The phi coefficient can be interpreted like a Pearson correlation coefficient if there is a symmetrical distribution to be found in the contingency table. With asymmetrical distributions (i.e. one group is significantly smaller) upper and lower border that can be reached with this cross tabulation are defined by $\varphi_{max}(-) < \varphi < \varphi_{max}(+)$.

Table 3. Correlation matrix of instance to group assignments to the main 4 groups based on the phi coefficient φ. Numbers in the cells give the actual value of φ reading top to bottom: $\varphi_{max}(-)$, φ, $\varphi_{max}(+)$.

	Sharing	Affection	Function
Preservation	-0.59 -0.05 0.39	-0.78 -0.26 0.84	-0.36 -0.36 0.55
Sharing		-0.50 0.25 0.46	-0.93 -0.07 0.21
Affection			-0.43 -0.21 0.47

5. DISCUSSION

We set out with the goal to investigate whether the classes of Kindberg's taxonomy are disjoint and whether the two dimensions sufficiently explain users' intentions for video production. As reported above, 81% of the instances were assigned to more than one category. This implies that a discrete mapping to exactly one category is insufficient and would therefore constrain possible reasoning about users' intentions. In particular, the classes are not disjoint. This is in line with our qualitative findings from the interviews: P4 mentioned a

video he took on a mountain while snowboarding. He recorded the video because he "took it because [he loves] snowboard video tricks and [he thinks] that it is very important to reconsider them to improve [his own] technique". While he appreciates snowboard video tricks and we therefore can infer an affective intention, he uses videos like these to work on his snowboarding skills. So there is also a functional use.

P10 reported on a video she took in a karaoke bar, where she went with a friend. There were two reasons for her recording that very video, as she told us: "First my friend is so good at singing and also charming and second he was about to leave the city and that was our last meeting. So I took the video to remember the night". She clearly distinguished between affection, ad-hoc intention and the intent to preserve the moment to remember it later. While the intent for preservation is very much related to the affection there is still a difference noted in the interview.

In Table 2 it is assumed that "Preservation" is the opposite of "Sharing". Such an effect could be easily observed by a high negative correlation coefficient. However, our analysis shows that the group assignments to "Preservation" and "Share" do not correlate indicated by φ being nearly zero (see Table 2). We can therefore assume that the group "Preservation" is independent from the group "Sharing". This leads to the conclusion that the two dimensional model does not apply as-is to user intentions for recording videos. Moreover, "Function" correlates with "Preservation" and for this case the phi coefficient reaches the minimum possible with this cross tabulation ($\varphi_{max}(-) = \varphi = 0.36$). The negative correlations can be explained intuitively. Videos recorded driven by a functional intention are implicitly preserved until the task is completed. So there's no need for an explicit preservation intent. Together with the findings from our clustering, Kindberg's taxonomy is too coarse-grained. Again, this is in line with the qualitative findings from our interviews: P12 described a situation where he recorded a video of his University at different weather conditions which was meant to be sent to his friends back home. He mentioned that keeping his friends updated with videos is routine for him. While in Table 2 such an instance would be classified affective and social, his actual intention was two-fold: affection/sharing (to be in contact with his friends) and functional (due to being a routine considered as regular task).

P11 reported a situation of taking a video: "My children and me were hiking, visiting relatives in Tyrol. I took the video to remember that day later and what happened on the trip. We make a lot of trips and I like to record them." His intention was to preserve the hiking trip, which he also does on a regular basis. Contrary to preservation based on affection (e.g. to preserve a precious moment) there was a recognizable routine in this situation and a functional use of the camera for documentation purposes.

6. CONCLUSION

We contributed the results from an exploratory user study examining user intentions for video production. These show that there cannot be a clear distinction between different intention classes: functional as well as affective intentions trigger video capturing at the same time. Furthermore, current taxonomies for photo creation assume that people either share or preserve pictures, which therefore cannot be applied as-is for the use case of video creation. Our results also provide evidence that the two intention classes "Preservation" and "Sharing" are independent. Thence, the intention clusters proposed in Section 4 can be applied to annotate videos based on their creators' intentions allowing for new approaches in video retrieval. By assigning the creator's intentions to the relevant groups, one can for instance deliver videos recorded out of technical interest to users that are interested in the capabilities of a particular device. Future work will investigate how user intentions can support multimedia information systems and multimedia retrieval and how intentions can be mined from data.

7. REFERENCES

[1] Nis Bornoe and Louise Barkhuus, "Video microblogging: your 12 seconds of fame," in *Proc. CHI '10*. 2010, pp. 3325–3330, ACM.

[2] Andrei Broder, "A taxonomy of web search," *ACM SIGIR Forum*, vol. 36, pp. 3–10, 2002.

[3] Danny Levinson and Daniel E. Rose, "Understanding user goals in web search," in *Proc. WWW '04*. 2004, pp. 13–19, ACM.

[4] Tim Kindberg, Mirjana Spasojevic, Rowanne Fleck, and Abigail Sellen, "The ubiquitous camera: An in-depth study of camera phone use," *IEEE Pervasive Computing*, vol. 4, pp. 42–50, 2005.

[5] Mathias Lux, Marian Kogler, and Manfred del Fabro, "Why did you take this photo: a study on user intentions in digital photo productions," in *Proc. SAPMIA '10*. 2010, pp. 41–44, ACM.

[6] Ann Makela, Verena Giller, Manfred Tscheligi, and Reinhard Sefelin, "Joking, storytelling, artsharing, expressing affection: a field trial of how children and their social network communicate with digital images in leisure time," in *Proc. CHI '00*. 2000, pp. 548–555, ACM.

[7] Arto Puikkonen, Jonna Hakkila, Rafael Ballagas, and Jani Mantyjarvi, "Practices in creating videos with mobile phones," in *Proc. MobileHCI '09*. 2009, pp. 3:1–3:10, ACM.

[8] Anselm Strauss and Juliet Corbin, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, vol. 2nd, Sage Publications, 2008.