Supporting Personal Media Authoring

Nicholas Diakopoulos and Irfan Essa

Georgia Institute of Technology / GVU Center

{nad, irfan}@cc.gatech.edu

ABSTRACT

Authoring media is a difficult task which is confounded by a huge range of possibilities for expressing any given message. In order to mitigate the task we argue for a tighter coupling between computer and author with human agency maintained through the use of suggestive user interfaces. The general authoring task is described as choosing a message and editing the selection and arrangement of media elements. For personal media in particular, story-telling and experience sharing are highly important and influence the general task profile. We discuss the context of personal media authoring and the degree of automated support possible in light of complexity and annotation.

Keywords

media authoring, personal media, suggestive user-interfaces

1. INTRODUCTION

A great majority of media can be composed or authored in a huge variety of ways. That is, messages can be communicated whether linguistically or more generally with multimedia in many different forms, each of which might give rise to a varied meaning in the recipient of the message. Artists may embrace this variability as it affords them *spielraum*, but many other modes of expression require more direct and structured communication of ideas. Human communication is an inexact process, which involves solutions that are neither right nor wrong, only better or worse, with relatively imprecise goals and evaluation criteria [13]. The large number of ways of expressing a message, combined with fuzzy goals and evaluation criteria can make practical media authoring difficult to get right, time-consuming, and in some cases tedious.

Due to the complexity of the authoring process it is desirable to offer the media author some degree of automation. We would argue in fact for the full range of task-migratability in authoring since this allows for fully automatic, partially automatic, or fully manual control of the process. A human in-the-loop approach toward authoring is beneficial since in many situations the ultimate decision should rest with the author whose message is being communicated.

Suggestive user interfaces are a promising avenue for addressing the above identified issues with media authoring [9]. The primary goal of these interfaces applied to media authoring is to support

the author by supplying intelligent (or at least informed) suggestions, which help in developing a communicational message or in designing a media presentation. Our position here is that such interfaces are valuable for authoring media since users will make better and faster choices from the range of authoring options available to them. In the following sections we will develop the media authoring tasks in more detail and then revisit issues of task-migratability and suggestive UIs as they apply to those concrete tasks.

2. AUTHORING

2.1 Task Definition

Consider for a moment the process of writing this position paper, which many of the readers can likely relate to. Though text based this task is representative of the general multimedia authoring process, which is similarly described in [8]. We first downloaded the layout template for this publication, which defines a visual *style* and impacts the perceptual presentation of the document. After reading the workshop description to elucidate expectations for the paper, we then decided on a rough idea for the *message* that we wanted to convey and began to outline the various sections of the paper. We then selected references to support points that we would make as well as to provide suggestions for important points that should be covered. We subsequently began to flesh out individual sections by writing text and in the process made decisions concerning the *structure* and the *content* of the document as it evolved.

Task analysis can roughly be described as identifying: goals, preconditions (i.e. what you need a priori in order to achieve goals), steps to perform, and interdependencies of those steps [11]. At the center of the process of writing is the initial goal or desire to communicate some chosen message. Preconditions include having the necessary references and sufficient understanding of the domain and recipients. The goal is implemented by making decisions about structure and content, which define the presentation of that message. In semiotic-speak this corresponds to choosing the syntagm and the paradigm for the media. Finally, there is a high level of interdependency of these steps since a choice in syntagm or paradigm may impact the meaning of future choices along either of these dimensions.

2.2 Personal Media

The validity of the choices in syntagm and paradigm are in large part affected by the communicational context and purpose of the media. For instance, the different purposes of advertising, scientific communication, and story-telling necessitate different messages, presentations, and selections of information.

As one example we consider the context of authoring a presentation from a personal photography collection, which has its own set of communicational purposes that have been explored in a

number of ethnographic and user studies [1][2][4][5][14]. For instance, in a study conducted by Crabtree it was found that the need to support the production of accounts and the telling of stories and thus the conveyance of experience were considered paramount to the process of photo sharing. Crabtree also notes that digital photo sharing should allow users to point out or highlight similarities within or between photos as this helps in storytelling or reminiscing [4].

The emphasis on storytelling, especially for authored presentations meant for remote sharing, resurfaces in other studies such as that of Balabanovic [2]. Frohlich notes that storytelling aspects of photo sharing are most important when sharing with people who were not present when the photo was taken. If the recipient was present when the photo was taken, the communication was more akin to reminiscing. He also indicates that recipient design, or an orientation and sensitivity to the recipients of the story, is a key factor in structuring the word and topic selection as well as sequencing in the story [5].

Shen et. al developed the Personal Digital Historian (PDH), a system for face-to-face sharing of photographs around a digital table, to explicitly support elements of flexible narrative generation [12]. They did this by enabling fast and simple queries and visualizations for the who, what, where, and when of the photos in the collection. While this is a step in the right direction for co-located generation of photo presentations, it does not address a very relevant scenario in light of the digital communication boom: remote presentation sharing.

Authoring for personal media thus has some specific features which influence the nature of the message chosen and the syntagmatic and paradigmatic choices made. The message is typically story or experience based and choices in syntagm and paradigm primarily center around basic aspects of good story-telling: who, what, where, and when. Furthermore, personalization from the author's side (as well as perhaps on the user's [3]) is important for augmenting the context of the experience and increasing relevance to the recipient. This is also in keeping with the general idea of placing more emphasis on the meaning-making process when producing media [10].

3. TASK MIGRATABILITY

As authoring can be a difficult or time-consuming task, off-loading certain aspects to an automated computer process would be beneficial. This is a flexibility principle in HCI known as task-migratability, and implies shifting part or all of the task to the computer.

The location of the brunt of the work in the media authoring process in some sense defines the nature of the application. A fully automatically authored piece of media may be useful for browsing or summarization since it can quickly be generated to portray something without any user intervention. A fully user centric application would allow the user creative flexibility in defining exactly how the pieces of media come together, and might be considered more artistic.

In between the two extremes is a sweet spot, which makes media authoring easier or faster for the user, but which doesn't sacrifice too much of the creative potential in the output. A human in-the-loop approach is important since it gives the user agency in the message produced by the media. This leads to the idea of incorporating suggestive UIs into media authoring, since a

number of automatically generated suggestions can be screened by the author and an appropriate choice made. Additionally, this may reduce the cognitive and memory load on the user by leveraging recognition versus recall of authoring choices. These basic concepts have already been applied to an image editing task in [7], and to a photo collage authoring task in [6], both of which are specific types of personal media authoring.

Generally speaking automation could be applied to any of the parameters of the task analysis brought up in the last section (i.e. message choice, syntagm and paradigm choice, precondition satisfaction). If the choice of initial message to portray is automated, this could be considered a summarization or browsing system in which the communication is occurring from the computer to the human. An automatically generated syntagm corresponds to a template driven approach to authoring whereas manual syntagm selection is free-form arrangement or sequencing of media elements. Automated paradigm selection again corresponds to a summarization system and could serve to generate different themes by editing which media elements appear in a presentation. Additionally, preconditions for authoring such as domain information of the context could be encoded into a knowledge-base which interactively helps train the user.

For personal media, we can use the ethnographic record to be more specific about what aspects of the tasks may be suitable to automation. Since a primary goal of personal media presentation is to share individual experience and stories, automation should focus on supporting good storytelling. As in the PDH project mentioned earlier, automated paradigmatic choices could be based on the who, what, where, and when of the story to be told [12]. Additionally, clustering of visually similar photos would allow paradigmatic choices which avoided repetitious content. Syntagmatically we could consider rhetorical templates or dramatic structures, though some user research with communicating everyday experiences indicates that this may not be satisfying for the user due to the inherently unstructured nature of the media [1]. A more interactive and iterative approach to syntagm generation may be more appropriate, with the system providing suggestions for local structure in the presentation and the user making the final choice. Spatial or temporal juxtaposition of related photos combined with automated within-frame highlighting could aid in the need to draw similarities between related photos [4].

3.1 Production and Consumption Context

The success of task migratability for media authoring is in some sense tied to the level of AI available for making valid syntagmatic and paradigmatic suggestions. The difficulty in identifying *valid* suggestions is of course context bound, since differing contexts require differing degrees of semantic understanding by the computer. Having poor suggestions wouldn't help the user and could even lead to frustration with the interface. Whether the context of the media be entertainment, advertising, or personal storytelling, awareness of the context of the communication is necessary to make good suggestions.

As noted by both Frohlich and Crabtree, recipient design is an important aspect of photo sharing [5][4]. The vacation photos that you show to your mother may be a somewhat different subset of what you might show to your best friend. The feedback of the recipient is also seen as a guiding factor in the dynamic construction of the photo story. Part of the authoring task could be

considered to migrate to the recipient of a photo presentation if we allow for interactivity to change the syntagm and/or paradigm of the presentation (thus increasing recipient relevance). The issue of "Task Distributability" begs for more attention: how can the authoring task be distributed among multiple humans and/or computers? Certainly more ethnography is needed in this area in order to see what would be useful or practical.

Otherwise we could consider having recipient user models which would influence the choice of photos or their presentation as the author was constructing the presentation. Thus the author could specify an intended recipient and any automated authoring suggestions would then be geared toward that recipient based on the user model. In the end, however, it may be that interactivity between author and recipient is needed in order to provide a satisfactory level of personalization for this type of media.

3.2 Annotation

The successfulness or helpfulness of authoring suggestions is subject to the annotations which support the desired context. As the particular focus here is on the context of personal photo presentations and storytelling, annotations should support aspects of the photos such as *who* is in the picture, *what* is being depicted, *when* the photo was taken and *where* the photo was taken. Some of this information (e.g. when and where) can be extracted automatically from EXIF and GPS metadata, whereas who and what information must be added manually to the IPTC fields of the image. Enrichment of the image through user added audio or textual captions could also augment the computer's understanding of the image and ability to suggest a suitable photo. Not only that, but captions can also help a lot in recipient understanding of the meaning of the photo.

Aesthetically, a certain amount of content analysis of the photos may also be helpful in generating good suggestions. As mentioned above, visual clustering could improve paradigmatic suggestions by reducing redundant content. Color and texture information could also be used to aid syntagmatic suggestions by increasing visual variability through juxtaposition. Face detection is another aspect of content analysis which is important since faces and people typically represent subjects of interest within photos.

Annotations coming from EXIF, IPTC, or even face detection can be managed and processed most easily using textual descriptions. Content analysis based annotations aren't as amenable to such descriptions, therefore an internal and lower-level description of color and/or texture features would be needed. The user may not need to know about the nature of that lower-level description as long as the suggestions provided by the system are distinctive and palpable.

4. CONCLUSIONS

Media authoring is a complex task which can benefit from intelligently automating aspects of it. We argued in favor of suggestive user interfaces which maintain the agency of the user in the media authoring task. We outlined some of the key components of the general authoring task and in particular the task of constructing a presentation from personal media, which was based on user studies and ethnographies. Some ideas for migrating the tasks and the annotation process to an automated process were then discussed.

5. REFERENCES

[1] Appan, P., Sundaram, H. and Birchfield, D. Communicating everyday experiences *Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, ACM Press, New York, NY, USA, 2004.

[2] Balabanovic, M., Chu, L.L. and Wolff, G.J. Storytelling with digital photographs *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, The Hague, The Netherlands, 2000.

[3] Bulterman, D.C.A. User-centered abstractions for adaptive hypermedia presentations *Proceedings of the sixth ACM international conference on Multimedia*, ACM Press, Bristol, United Kingdom, 1998.

[4] Crabtree, A., Rodden, T. and Mariani, J. Collaborating around collections: informing the continued development of photoware *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM Press, Chicago, Illinois, USA, 2004.

[5] David Frohlich, A.K., Celine Pering, Abbe Don, Steven Ariss, Requirements for Photoware. in *CSCW*, (New Orleans, 2002), 166-175.

[6] Diakopoulos, N. and Essa, I., Mediating Photo Collage Authoring. in *Symposium on User Interface Software and Technology (UIST)*, (Seattle, 2005).

[7] Diakopoulos, N., Essa, I. and Jain, R., Content Based Image Synthesis. in *Conference on Image and Video Retrieval (CIVR)*, (Dublin, 2004), 299-307.

[8] Falkovych, K. and Nack, F. Context Aware Guidance for Multimedia Authoring: Harmonizing domain and discourse knowledge. *In: Multimedia Systems Journal (MMSJ) (to appear)*, 2006.

[9] Igarashi, T. and Hughes, J.F. A suggestive interface for 3D drawing *Proceedings of the 14th annual ACM symposium on User interface software and technology*, ACM Press, Orlando, Florida, 2001.

[10] Nack, F. The Future in Digital Media Computing is Meta. *IEEE Multimedia*, *11* (2). 10-13.

[11] Nielsen, J. *Usability Engineering*. Academic Press, 1993.

[12] Shen, C., Lesh, N. and Vernier, F. Personal digital historian: story sharing around the table. *interactions*, *10* (2). 15-22.

[13] Terry, M., Mynatt, E.D., Nakakoji, K. and Yamamoto, Y. Variation in element and action: supporting simultaneous development of alternative solutions *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, Vienna, Austria, 2004.

[14] Rodden, K. and Wood, K.R. How do people manage their digital photographs? *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, Ft. Lauderdale, Florida, USA, 2003