The Multiplayer: Multi-perspective Social Video Navigation

Zihao Yu
Rutgers University
Dept. of Electrical and Computer Engineering
zihaoyu@eden.rutgers.edu

Nicholas Diakopoulos, Mor Naaman
Rutgers University
School of Communication and Information
diakop@rutgers.edu, mor@rutgers.edu

ABSTRACT

We present a multi-perspective video "multiplayer" designed to organize social video aggregated from online sites like YouTube. Our system automatically time-aligns videos using audio fingerprinting, thus bringing them into a unified temporal frame. The interface utilizes social metadata to visually aid navigation and cue users to more interesting portions of an event. We provide details about the visual and interaction design rationale of the multiplayer.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design**,** Human Factors

Keywords: Video, Multi-perspective, Social Media

INTRODUCTION

The linear, dynamic, and ephemeral nature of video has long led to interaction challenges associated with representing and navigating the medium across a range of tasks and contexts from annotation [10], to analysis [5] and surveillance [4], to consumption and playback [2]. To some extent, video editing has arisen as the dominant process for constructing narrative (or other) paths through a set of related video clips. It is through this ordering of video clips in time that the experience of meaning and emotion are constructed [8].

We now find ourselves in a world where video, and in particular video that has been recorded by consumers, can provide an abundance of raw material for consumption and presentation. In this demonstration we will showcase an interface that lets users create their own trajectory through a multi-perspective video event, such as a live music concert. Our multiplayer organizes social video aggregated from online sources like YouTube by automatically time-aligning videos using audio fingerprinting [6, 9] and incorporates social metadata to aid in navigation and in creating one's individual path or "edit" through the video material.

Multi-perspective video playback interfaces have been researched in a number of contexts including watching sports [2, 7] and monitoring or analyzing surveillance footage [4]. Commercial implementations of multi-perspective video such as the HBO video cube (http://www.hboimagine.com) have also explored the value of end-user interactivity in crafting the trajectory through a video space. Our interface

is novel in that it seeks not to organize *produced* multi-perspective video, but rather *social* video clips that have been shared by multiple people, each recording their own (potentially fragmented) view of an event. The nature of this content is different due to various factors: (1) *varied quality*: depending on the user, their capture device, and their position, (2) *limited capture control*: the many perspectives collected could be overlapping, unstable or moving, and are unknown to the application, (3) *varying time coverage*: the clips often do not extend across the entirety of the event in time, and (4) *available metadata*: the content provides an opportunity for incorporating social cues for navigation from video metadata; for instance, the number of views of a clip can be an indicator of quality.

Our system works by aggregating video clips from individual events and then uses audio fingerprinting to time-align the videos, as described in [6]. This pre-processing step occurs on the server before the multiplayer renders the interface. The drawbacks of this approach are that videos must come from the same aural space and non-unique audio characteristics can sometimes lead to difficulty in matching [6].

VISUAL AND INTERACTION DESIGN

In general, video navigation can be meaningful across a hierarchy of scales depending on the nature of the task. For instance some navigation tasks, such as those involved in video editing, logging, or annotation may benefit from fine-grained [3] or even object-based [1] navigation of video time. Other video navigation tasks do not require such precision and can benefit from a more coarse level of navigation such as to chapters in a DVD. Our design goal in this work was primarily to provide support for navigation *between* different time-aligned perspectives, rather than to enable more precise navigation within any one perspective. We designed the interface to give cues to help the user assess which perspective could be most interesting to switch to as the timeline advances and new perspectives become available. We also provide an *overview* of the multi-perspective space that signals how many videos are available and, more importantly, shows their distribution over time - an honest social signal of level of interest in different portions of the event.

Figure 1 shows our multiplayer, with one main video in the upper left (Figure 1a.), and a list of alternate video clips simultaneously playing in the panel to the right (Figure 1c). Below the video playback area is the multi-perspective

overview showing the layout of the different available clips and their temporal alignment to each other (Figure 1b.). Clicking a clip in the overview, or in the alternate video list makes **(a)** that clip the current main video. To aid navigation, we visually link the temporal representation of clips to the video thumbnails above using the outline color and style; a solid black outline indicates the main video and a dashed gray outline indicates the clips that are currently available to switch to. The vertical ordering of **(b)** the clips in the overview (Figure 1b) corresponds to the vertical ordering of the clips in the alternate video panel, and hovering over a clip brushes the corresponding video frame in the alternate video list. A
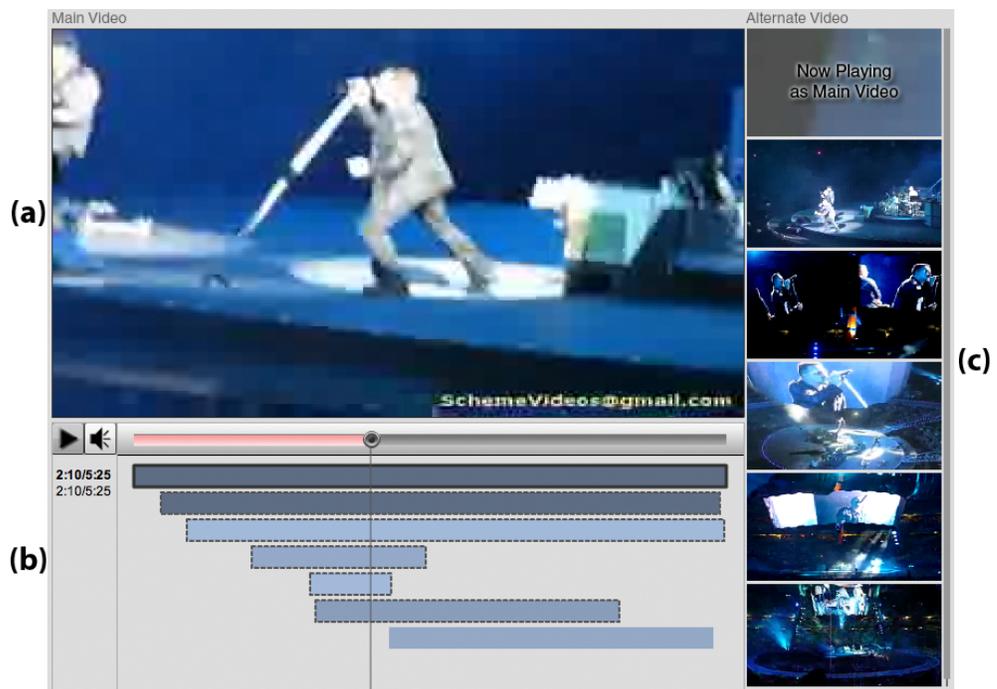

Figure 1. The multi-player showing seven videos from a U2 concert. (a) shows the main video, (b) shows the overview of the clips and their temporal alignment, (c) shows a list of alternative clips that can be switched to at that moment in time.

freeze-frame surrogate is inserted to indicate where the main video fits in the alternate video stack (Figure 1c.). We expect the consistency of these mappings to aid in navigation between these two spaces. Furthermore, to help maintain navigational awareness as the user switches perspectives, we animate transitions between the alternate video list and the main video. Traditional methods of video navigation such as start, stop, and non-linear seeking are all supported. The presentation scales by using a scroll bar when the number of clips exceeds the available space.

The multiplayer is designed to support *social* video to aid in navigation or selection of different views on an event. The player can map a clip's metadata to its brightness or hue on the overview (Figure 1b.), adding an additional social signal. For instance, YouTube currently provides access to such metadata as average rating, view count, number of comments, and comment text for each video. In future work we are interested in incorporating metadata from content analysis such as audio quality or camera stability which would further help users decide which perspective to switch to.

CONCLUSIONS AND FUTURE WORK

We have presented a system and interface for aggregating, time-aligning, and navigating multi-perspective social video aggregated from YouTube. We are currently exploring different visual mappings for social metadata to enhance the player and are planning a laboratory study to assess users' experience and satisfaction navigating using the interface. We are interested in studying the utility of the

multiplayer using a variety of event content from news and speeches to concerts and other performances.

REFERENCES

[1] Karrer, T., Weiss, M., Lee, E. and Borchers, J., DRAGON: A Direct Manipulation Interface for Frame-Accurate In-Scene Video Navigation. in *Proceedings of CHI*, (2008).

[2] Bentley, F. and Groble, M., TuVista: Meeting the Multimedia Needs of Mobile Sports Fans. in *ACM Multimedia*, (2009).

[3] Diakopoulos, N. and Essa, I., Videotater: an approach for pen-based digital video segmentation and tagging. in *Proceedings of UIST*, (2006), 221-224.

[4] Girgensohn, A., Shipman, F., Turner, T. and Wilcox, L., Effects of presenting geographic context on tracking activity between cameras. in *Proceedings of CHI*, (2007).

[5] Hagedorn, J., Hailpern, J. and Karahalios, K.G., VCode and VData: Illustrating a New Framework for Supporting the Video Annotation Workflow. in *Advanced Visual Interfaces (AVI)*, (2008).

[6] Kennedy, L. and Naaman, M., Less Talk More Rock. in *World Wide Web Conference (WWW)*, (2009).

[7] Lynn, S., Olsen, D. and Partridge, B., Time Warp Football. in *EuroITV*, (2009).

[8] Murch, W. *In the Blink of an Eye: A Perspective on Film Editing*. Silman-James Press, 2001.

[9] Prarthana Shrestha, H.W. and Barbieri, M., Synchronization of multi-camera video recordings based on audio. in *ACM Multimedia*, (2007).

[10] Ramos, G. and Balakrishnan, R., Fluid interaction techniques for the control and annotation of digital video. in *Proceedings of UIST*, (2003), 105-114.