MediaDiver: Viewing and Annotating Multi-View Video

Gregor Miller, Sidney Fels,
Abir Al Hajri, Michael Ilich,
Zoltan Foley-Fisher, Manuel Fernandez
Human Communication Technologies Laboratory
University of British Columbia
Vancouver, BC, Canada
{gregor,ssfels,abira,michaeli, zoltan}@ece.ubc.ca
manuel.fernandez@web.de

Daesik Jang
Kunsan National University
Gunsan, South Korea
dsjang2@gmail.com

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
We propose to bring our novel rich media interface called *MediaDiver* demonstrating our new interaction techniques for viewing and annotating multiple view video. The demonstration allows attendees to experience novel moving target selection methods (called *Hold* and *Chase*), new multi-view selection techniques, automated quality of view analysis to switch viewpoints to follow targets, integrated annotation methods for viewing or authoring meta-content and advanced context sensitive transport and timeline functions. As users have become increasingly sophisticated when managing navigation and viewing of hyper-documents, they transfer their expectations to new media. Our proposal is a demonstration of the technology required to meet these expectations for video. Thus users will be able to directly click on objects in the video to link to more information or other video, easily change camera views and mark-up the video with their own content. The applications of this technology stretch from home video management to broadcast quality media production, which may be consumed on both desktop and mobile platforms.

Keywords
Rich media viewing, video annotation, multi-view interaction

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

General Terms

Design, Human Factors

Introduction

The massive growth of semi-organized video content available on the web has stretched the metaphor of linear video browsing to its limit. The historical view of videos as a linear sequence of images that can be played, rewound or fast-forward no longer supports navigation through such a complex video space. This is apparent from the lack of a natural progression through videos at sites such as YouTube™or from the inability to share specific moments of a particular video source easily.

There are currently no convenient mechanisms or models to effectively support a user's viewing and authoring of video content in this complex environment. Simple actions such as changing camera viewpoints, selecting hypervideo linked objects or marking up content are virtually non-existent for sharing and experiencing multiple view video. We focus on viewing and annotation of multi-camera sports video as this provides a rich data set and generalizes to many other single- and multi-camera contexts. In this domain, the use of many cameras is common, with the players moving quickly each with different actions happening at different times appearing in each of the different views. We illustrate how the multi-camera data can be viewed effectively for taking advantage of the interactive, annotated video to couple annotation and viewing in the same environment. Further these mechanisms support finding the same object in different views. The MediaDiver that we propose to show supports all of these features.

We imagine that our viewing and annotation environment, MediaDiver, will be critical to support interactive video being a primary media of web applications in the future such as video sharing and social networking sites. Interactive annotated video supports home video organization and production for personal use and sharing over the network.

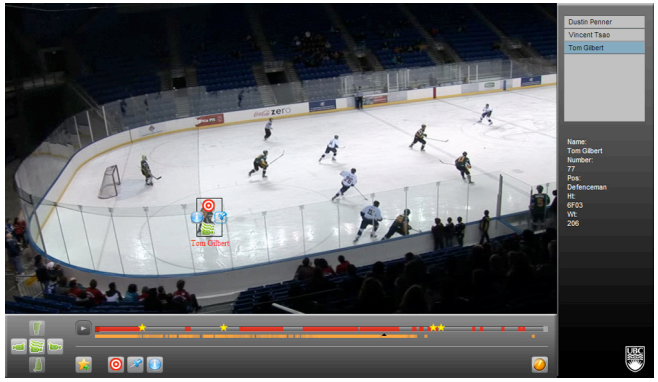


Figure 1: Selection of moving targets with MediaDiver

Moreover, the MediaDiver is helpful for broadcasters looking to create meta content. For example, in sports broadcasts, the MediaDiver supports the following example functions:

1. Users can specify that a player should remain in view so that cameras are automatically adjusted to keep him in view
2. Links to hypertext data such as player information or statistics associated with the player, or other video (e.g. interviews)
3. Users can pin player objects for future easy access
4. Users can easily adjust the camera view
5. Users can edit the current video annotation and easily view these changes

Related Work

Our research encompasses the application of computer vision processing techniques for integrating multi-video

spaces, hyperlinked video and semantic indexing by metadata creation. In our research, we define multi-video spaces as a collection of related videos, such as multiple camera angles covering a shared physical space. An example of a multi-video space is a broadcast sporting events, where a broadcast manager can select their own views based on personal goals. One of the approaches of the interface concept is to extend this role to the individual viewer for multi-camera broadcast events.

One method for handling multi-video spaces is to provide a summarization of the available video sources, the spatial area that they cover or a timeline indicating their temporal boundaries or overlap. Shipman et al [15] developed a player and editor for detail-on-demand video that provides hypervideo summarizations of video clips, organized in a hierarchical structure that enabled linear video navigation. Lou et al [8] developed a server application to multiplex several video streams from network-controlled cameras that provided a spatial view sweep of adjacent cameras, the ability to switch between cameras and the option to freeze the current view to clients over broadband connection. Girgensohn et al [3] introduced a temporal form of summarization with a hypervideo player that incorporates hyperlinks as partitions and annotations as peripheral screenshots in a timeline visualization. In our approach, all interactive affordances are incorporated into the video window, establishing semantic, temporal and spatial relationships between multiple video sources.

Summarization is used for displaying the results of a visual search engine that extends browsing capabilities to objects or events beyond the current window.

Within our video space, we consider that identified spatio-temporal regions or objects can be hyperlinked to other spatio-temporal content in the form of video or other online resources. Sans et al [12] developed an XML-based language for multimedia presentations in which SMIL[4] is used to describe the spatial and temporal relations between hyperlinked video. Girgensohn et al [3] demonstrated the analogy of a timeline slider mechanism, organizing video sources and their hyperlinked content into a linear, temporal sequence. Nitta et al [10] opted for a dynamic model of organizing broadcast sport video streams by temporal sequence. Our approach is based on providing hyperlink navigation through semantic relationships without interrupting the apparent spatial or temporal coherence to the viewer during navigation.

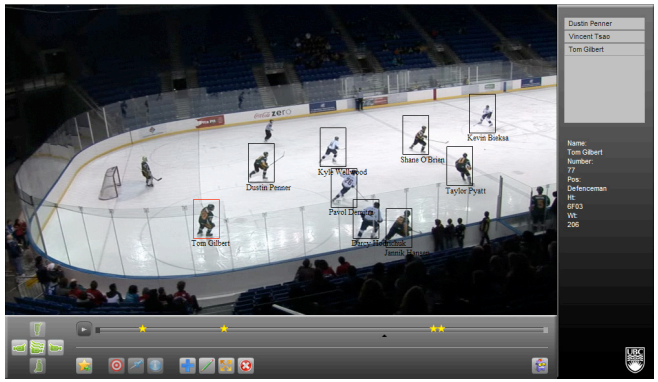


Figure 2: Annotation of hockey players with MediaDiver

An annotation is defined by emphasizing parts of a given information space[5]. We can state that an annotation in general fulfils two major tasks: first, it visually emphasizes parts of an information space and second it serves as an anchor between a point of interest and additional information, which eventually results in a link-structure. Annotations are used because they help us navigate in an information space and retrieve data. Annotations in video follow the same definition given above by visually emphasizing objects in a video stream and by serving as an anchor between objects and additional information.

Annotated videos are also often referred to as hypervideo[14], which resolves the linearity of video structures and so creates a non-linear information space. Among others, HyperCafe[13], HyperSoap[1], VisualShock MOVIE[6] and HyperFilm[11] are considered to be the first research projects that integrate video annotation as a core concept. IBM developed a powerful video annotation editor called VideoAnnEx[7]. The stand-alone application stored the annotation data in XML files using the MPEG-7[9] description format. Finke et al. introduced the first Web-based video player[2] that allowed viewers not only to interact with annotated videos but also to create own annotations and share them with others over the Internet in real-time.
The MediaDiver contains some proven existing technologies and builds on these to provide a new video viewing and annotation experience.

Demonstration Overview

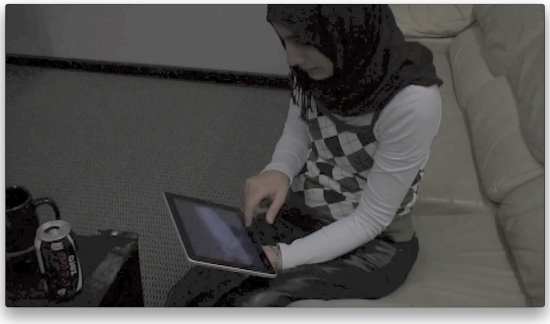


Figure 3: User playing with mobile version of MediaDiver

We are proposing a demonstration of new methods for viewing and annotating rich media with multiple views. One of the major challenges of interaction with rich media is the selection of a moving target. Using our novel method called Hold, a target may be selected by holding down the mouse button to pause the video, moving to the target and releasing to select. Selection serves as a context switch which enables functions such as highlighting, information retrieval and pinning. After target selection, the timeline displays the presence of the target in this view and other views. Selection persists if a target leaves the current field of view, after which the interface switches to a view in which the target is visible. Our interface provides several additional options after selection. With the mouse button held down over a target, a pie menu appears with options for following the target, information retrieval, pinning and annotation. Target specific information is displayed in the panel on the right hand side of the interface if requested, below a target shortcut menu. Users can pin targets to the shortcut menu for more efficient re-selection. Annotation is integrated into our interface and enables enthusiasts to create their own rich media experiences. Our delayed annotation method allows users to add bookmarks while viewing so that subsequently they may return to add their own annotations. Alternatively users may annotate at any time by enabling the editor interface and can add new or edit existing content. We have implemented our multiple view rich media interface on mobile platforms, which includes manual view selection methods such as gestures to switch directly in the viewer. Absolute view selection is supported with our grid interface, and relative view selection can be accomplished with our video flow visualisation.

Conclusion

For CHI Interactivity 2011 we propose to bring our novel viewing and annotation interface, the MediaDiver, to allow attendees to experience state-of-the-art interaction techniques. We hope that as people play with our demonstration they will be inspired to imagine the interaction complexities resulting from the emergence of

multi-view rich media. We will bring both desktop and mobile devices for attendees to play with. We would also like to work with the CHI organisers so that we have permission to allow participants to upload their personal recorded content of the conference so that they can view and annotate CHI happenings for everyone to share and enjoy. Thus, together, attendees will experience the future.

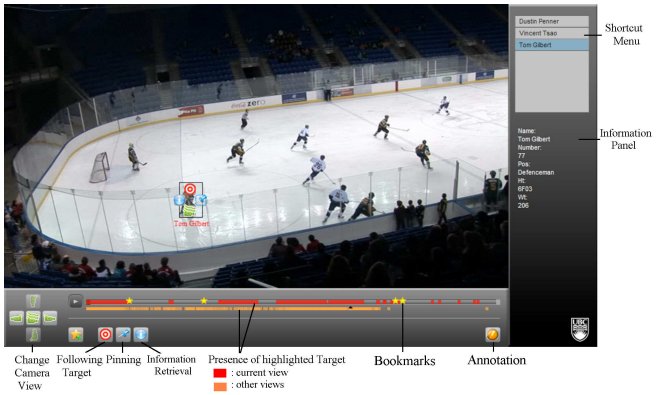


Figure 4: Overview of MediaDiver in view mode

References

[1] M. Bove, J. Dakss, S. Agamanolis, and E. Chalom. Hyperlinked television research at the mit media laboratory. *IBM System*, 39, 2000.

[2] M. Finke and C. Zahn. Collaborative knowledge building based on hypervideo. In *Computer Support for Collaborative Learning*, Bergen, Norway, June 2003.

[3] A. Girgensohn, L. Wilcox, F. Shipman, and S. Bly. Designing affordances for the navigation of detail-on-demand hypervideo. In *Working Conference on Advanced Visual interfaces*, Gallipoli, Italy, May 2004.

[4] P. Hoschka. Smil: an introduction. In *ACM SIGGRAPH 2002 Conference Abstracts and Applications*, San Antonia, Texas, July 2002.

[5] J. Hummes, A. Karesently, and B. Merialdo. Active annotations of web pages, voting, rating, annotation - web4groups and other projects. *Approaches and First Experiences*, 104, 1997.

[6] M. A. Inc. Authoring with hyperlinked video. Technical report, Mitsubishi Electric America Inc., 2000.

[7] C.-Y. Lin, B. L. Tseng, and J. R. Smith. Universal mpeg content access using compressed-domain system stream editing techniques. In *IEEE International Conference on Multimedia and Expos*, Switzerland, August 2002.

[8] J. Lou, H. Cai, and J. Li. A real-time interactive multi-view video system. In *ACM International Conference on Multimedia*, Hilton, Singapore, November 2005.

[9] B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7*. Wiley, 2002.

[10] N. Nitta, Y. Takahashi, and N. Babaguchi. Automatic personalized video abstraction for sports videos using metadata. In *Multimedia Tools Appl.*, January 2009.

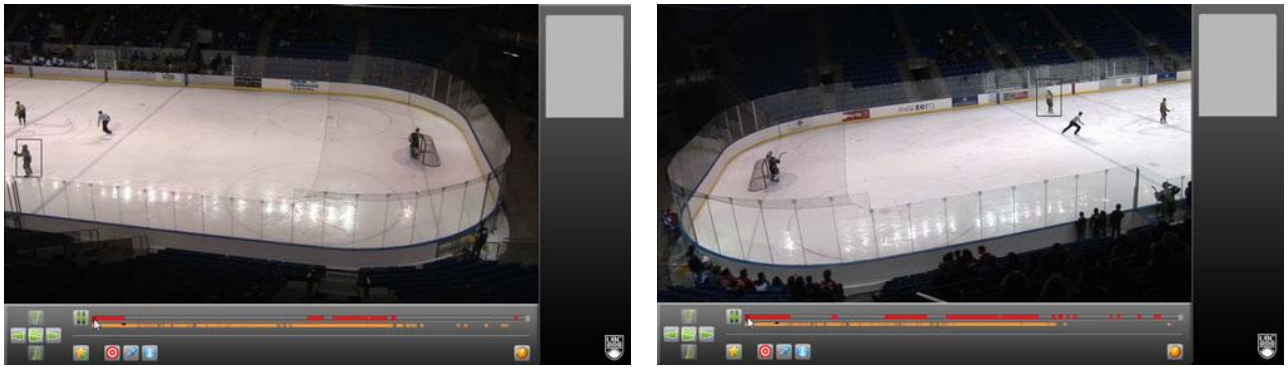
[11] M. Pollone, M. Rusconi, and R. Tua. From hyper-film to hyper-web: The challenging continuation of a european project. In *Electronic Imaging and the Visual Arts Conference*, Florence, Italy, 2002.

[12] V. Sans and D. Laurent. Navigating with inheritance in hypermedia presentations. In *ACM Symposium on Applied Computing*, Dijon, France, April 2006.

[13] N. Sawhney, D. Balcom, and I. Smith. Hypercafe: Narrative and aethetic properties of hypervideo. In *ACM Conference on Hypertext*. ACM, 1996.

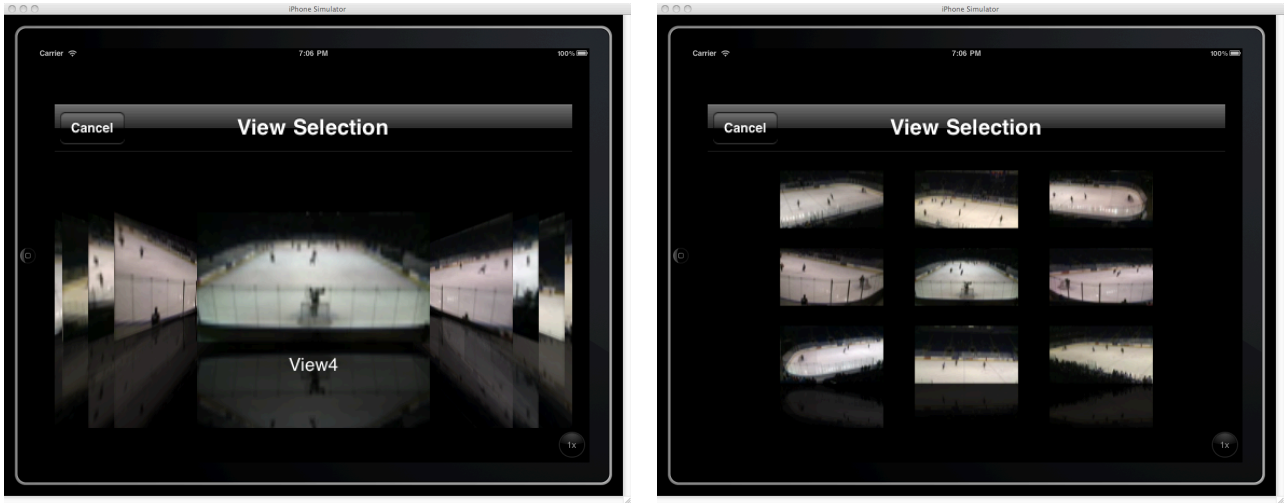
[14] N. Sawhney, D. Balcom, and I. Smith. Authoring and navigating video in space and time: An approach towards hypervideo. *IEEE Multimedia*, October-December 1997.

[15] F. Shipman, A. Girgensohn, and L. Wilcox. Authoring, viewing, and generating hypervideo: An overview of hyper-hitchcock. In *ACM Trans. Multimedia Comput. Commun.*, November 2008.



(i) (ii)

Figure 5: View automatically swithches from (i) to (ii) when player X leaves the current field of view (i). Note that the player must be selected for ”target following”.



(i) (ii)

Figure 6: View Selection in (i) Video Flow and (ii) Video Grid on a mobile device with MediaDiver