Appeared in CHI'2001, Extended Abstract, ACM, pp.113-114, Seattle, WA., April 2001.

Time-ART: A Tool for Segmenting and Annotating Multimedia Data in Early Stages of Exploratory Analysis

Yasuhiro Yamamoto[1,2], Atsushi Aoki[3], Kumiyo Nakakoji[1,2,3],

[1]Grad. School of Information Science, NAIST
8916-5, Takayama,
Ikoma, 630-0101, Nara, Japan
+81 743 72 5381

[2]Intelligent Cooperation and Control Group, TOREST, JST
2-2-11, Taruoka, Miyagino,
Sendai, 983-0852, Miyagi, Japan
+81 22 295 8680

[3]Open-Source Business Department, SRA Inc.,
3-12 Yotsuya, Shinjyuku,
160-0004, Tokyo, Japan
+81 3 3357 9361

yasuhi-y@is.aist-nara.ac.jp, aoki@sra.co.jp, kumiyo@is.aist-nara.ac.jp

ABSTRACT

Time-ART is a tool that helps a user in conducting empirical multimedia (video/sound) data analysis as an exploratory iterative process. Time-ART helps a user in (1) identifying seemingly interesting parts, (2) annotating them both textually and visually by positioning them in a 2D space, and (3) producing a summary report. The system consists of Movie/SoundEditor to segment a part of a movie/sound, ElementSpace, which is a free 2D space where a user can position segmented parts as objects, a TrackListController that synchronously plays multiple sound/video data, AnnotationEditor with which a user can textually annotate each positioned object, DocumentViewer that automatically compiles positioned parts and their annotations in the space, ViewFinder that provides a 3D view of ElementSpace allowing a user to use different "depth" as layers to classify positioned objects, and TimeChart that is another 3D view of ElementSpace helping a user understand the location of each segmented part in terms of the original movie/sound.

Keywords

cognitive tools, exploratory data analysis, spatial positioning as representations, multimedia, 3D visualizations

EMPIRICAL VIDEO ANALYSIS TASKS

Usability testing, empirical software engineering, requirements analysis and many other types of empirical and field studies produce a multiple movie and sound data by simultaneously recording a variety of aspects of the study, such as subjects' behavior, computer displays that subjects use and experimenters use, subjects' gaze-trace, subjects' think-aloud protocols, or experimenters' voice annotations. The multiple data sources allow us to analyze behavior, thoughts, or mental situations of a subject during the study. This type of analysis task is best supported by viewing it as an exploratory task rather than by applying standard statistical analysis techniques [2].

Little studies have been done in supporting early stages of exploratory multimedia data analysis, where a user plays many what-if games to discover interesting aspects of the data, gradually understands what phenomena were taking place during the study, continuously segments data and refines the segmentation so that each segment illustrates a "meaningful" piece, and eventually develops coding schemes to classify segments. During the process, a variety of mental situations emerge such as "this part and that part (of a movie) may be related," "this might be interesting if it also happens in the other part," or "I do not know why but this part has drawn my attention."

Most of existing tools that help users understand video data are automated video summarization tools that help users more quickly browse video clips, or video annotations tools that aim at supporting a student to learn about the content of the videos. DIVA [2] is a tool that supports exploratory multimedia data analysis by providing operations on stream data using stream algebra. However, the approach presupposes that data are initially segmented and coding schemes exist to assign codes to each segmented stream. Our approach complements DIVA in the very beginning of exploratory data analysis by helping users segmenting data and developing potentially useful coding schemes that can then be used in DIVA.

THE TIME-ART SYSTEM

An approach: Spatial Positioning as a Representation

We view early stages of exploratory multimedia data analysis as early stages of a design task, and have applied spatial positioning as a representation that a user uses to explore the data. In early design tasks, sketching has been used as a powerful representation that helps a designer capture emerging concepts and have clearer understanding what the problem is [1]. We have studied that spatial positioning of objects serves as a representation similar to sketching, and supports exploratory, early-stages of design tasks [6]. We have applied the spatial-positioning in writing [3] and in programming [5] and conducted user studies. During the studies, we observed that subjects used a variety of positioning patterns to represent how this object is related to other objects, which are more important than others, or which objects need more attention, while gradually constructing structures among the objects.

A System Overview

Time-ART helps a user in (1) identifying seemingly interesting parts, (2) annotating them both textually and visually by positioning them in a 2D space (see Figure 1), and (3) producing a summary report. The system is built based

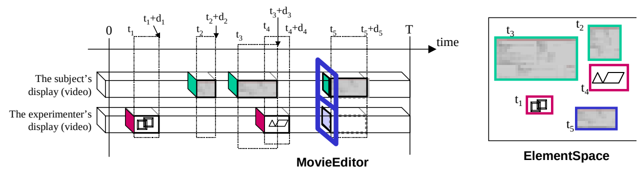

Figure 1: The use of Spatial Positioning in Exploratory Data Analysis

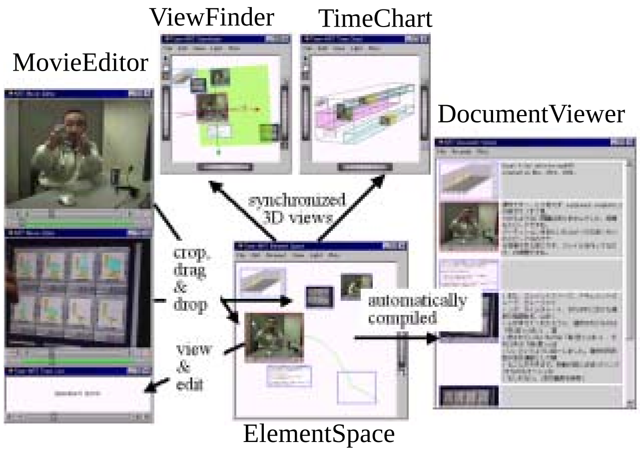

Figure 2: Components of Time-ART

on ARTWare, a suite of open-source tools that supports multimedia and 3D objects [4]. The system is written in Visualworks Smalltalk 3 and 5i and runs on Windows, Macintosh and Linux. Figure 2 shows components of Time-ART

Movie/SoundEditor is an extension of QuickTime movie player with an interface to more finely crop a part of a movie/sound data (Figure 3). When cropped, a user can drag-and-drop from the Editor to ElementSpace, where the system displays the first frame of the segmented movie as a thumbnail icon, which the user can then position and resize. Start/end points of this segmented part can be changed in MovieEditor at any time. In addition to segmented movies and sound data, text and image objects can also be positioned in ElementSpace. Each object (movie, sound, image and text) can be textually annotated by using AnnotationEditor (not shown in Figure 2).

Once positioned, when the user moves an object(s), a trajectory line (a green dotted line) is displayed to help the user better understand how positioning and re-positioning is done. A positioned segmented movie/sound can be played at any time in MovieEditor. TrackListController synchronizes MovieEditors to play all the multimedia data currently available in the analysis session.

Two 3D views are provided for ElementSpace. Both views dynamically display objects at the X-Y coordinates in ElementSpace but uses the Z-axis differently. ViewFinder provides a 3D view of ElementSpace allowing a user to “pop” and “dive” each object in ElementSpace. The Z-axis of the space serve as layers and helps the user classify positioned objects. A filter (a green plane in Figure 1) is provided to make objects at certain depth invisible in ElementSpace. TimeChart provides the other 3D view of segmented multimedia data. The Z-axis is used to represent the start/end

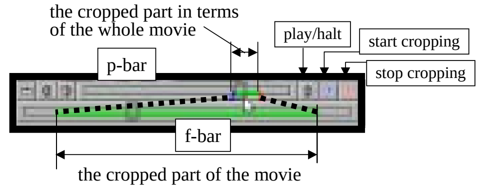

Figure 3: Control Bars of MovieEditor

time of each segmented movie/sound. Each segmented multimedia data is represented as a square tube with the length corresponding to the length of the original data, and the thumbnail is displayed at the Z coordinate of the tube at the starting time; duration of the segment is represented as colored sides of the tube.

Finally, DocumentViewer automatically compiles positioned parts in ElementSpace. The left column shows a list of positioned objects in the order specified by the user (e.g., top-to-bottom, left-to-right, then closer-to-farther in ViewFinder). The right column shows corresponding annotations added with AnnotationEditor. The content can be converted into a html format and viewed in a Web browser.

Current Status

We are currently conducting Time-ART user studies and self-applying Time-ART to analyze the collected data. We are using the eye-tracking system to analyze how representations provided by Time-ART are used by subjects. Although results are still preliminary, subjects’ reactions are very positive. We are further improving interactions and refining the system.

ACKNOWLEDGEMENTS

This research is partially supported by NEDO (New Energy and Industrial Technology Development Organization), MITI, Japan.

REFERENCES

1. Arnheim, R., Visual Thinking, University of California Press, CA, 1969.
2. Mackay, W.E. and Beaudouin-Lafon, M., DIVA: Exploratory Data Analysis with Multimedia Streams, Proc. of CHI'98, ACM Press, pp.416-423
3. Nakakoji, K., Yamamoto, Y., Reeves, B.N., Takada, S., Two-Dimensional Positioning as a Means for Reflection in Design, Proc. of DIS'2000, ACM Press, pp. 145-154, August, 2000.
4. Nakakoji, K., Yamamoto, Y., Takashima, A., Aoki, A., A Tool Suite for Building Interactive Systems using Multimedia Objects, Symposium on Smart Graphics, New York, March, 2001 (submitted).
5. Takada, S., Yamamoto, Y., Nakakoji, K., Two-Dimensional Positioning as Visual Thinking, Theory and Application of Diagrams (Diagrams2000, Edinburgh, UK), M. Anderson, P. Cheng, V. Haarslev (Eds.), Springer-Verlag, Berlin, pp.437-452, September, 2000.
6. Yamamoto, Y., Nakakoji, K., Takada, S., Hands-on Representations in a Two-Dimensional Space for Early Stages of Design, Knowledge-Based Systems Journal, Elsevier Science, Amsterdam, The Netherlands, vol. 13, No.6, pp.375-384, 2000.